Centre for Maintenance Optimization & Reliability Engineering

Director

Chi-Guhn Lee

Semi-annual report

June 2021

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

cmore.mie.utoronto.ca | 416·946·7867

Table of Contents

C-MORE progress meeting agenda

Virtual meeting by Microsoft Teams
Thursday June 10, 2021

Opening remarks, executive summary 9:00-9:40		Chi-Guhn Lee
Feature presentation: Asset management competency development 9:40-10:00		Ali Zuashkiani
10:00-10:15	15-minute break	
Collaborations with industry partners		
10:15-10:35	Results on resource extraction working group	Janet Lam
10:35-10:55	Kinross: A business case for machine learning methods	Zoha Sherkat-Masoumi
10:55-11:10	15-minute break	
Collaborations with industry partners		
11:10-11:30	UKMOD: Pilot implementation of MOD Aircraft carrier optimisation project	Varun C Ananda Rao
11:30-11:50	TTC: Track linetest and re-inspection update	Janet Lam
11:50-12:35	45-minute break	
Collaborations with industry partners		
12:35-12:55	CEA: An audit of data encoding for multi-site maintenance records	Janet Lam
12:55-1:15	Criticality Analysis and Asset Management of a Power Generation Facility	Pooyan Sharifi
1:15-1:30	15-minute break	
Student research projects		
1:30-1:50	A reinforcement learning approach to maintenance optimization of multi-unit systems	Vahid Najafi
1:50-2:10	A novel pattern-driven stochastic process for degradation forecasting with applications to rechargeable batteries	Zihan Zhang
2:10-2:25	15-minute break	
Student research projects		
2:25-2:45	Reinforcement Learning and Inventory Control	Avi Sokol
2:45-3:05	Learning from a few samples	Kuilin Chen
3:05-3:25	Towards Automatic Airport Security Imaging via Deep Domain Adaptation	Lucrece Shin
Closing remarks 3:25-3:30		

Executive summary

Chi-Guhn Lee, C-MORE Director

Introduction

Throughout the past year, the C-MORE team, including staff, students, academics, and industry collaborators, has adapted well to pandemic conditions. We have continued to operate with work-from-home modifications, and all meetings have happened virtually. Like many other organizations, we attended our first virtual conferences, and we hosted our progress meetings online. I am proud to say there has been no loss in quality – we have maintained steady and consistent work with our collaborating companies and engaged with new ones, including Canadian Electricity Association and Capital Power. In fact, it will be interesting to see what the next six months bring, as we may well decide some of our adaptations work too well to return to "normal."

The following report summarizes work undertaken since the meeting in December 2020.

The C-MORE team

Chi-Guhn Lee, Director

Chi-Guhn has leveraged the all-virtual environment to give invited talks and seminars at conferences around the world. While a detailed list of talks is reserved for the Activities section of the report, one initiative of note is the resource extraction working group, spearheaded by Chi-Guhn. This group met three times throughout the first part of 2021 to discuss ways C-MORE could better meet the needs of the mining sector. Outside this particular sector, Chi-Guhn is increasing the presence of C-MORE in asset management and maintenance optimization. On June 17, he will be hosting a workshop through the PEMAC GTA chapter, titled "Demystifying machine learning – hands on workshop using Excel."

Janet Lam, Assistant Director

Janet has been working with graduate and undergraduate students to push on with various industry-sponsored research projects. In particular, she worked with CEA on a data audit project, Capital Power on a criticality project, Kinross on a machine learning case studies project, and Titan Technologies on two different wind-turbine projects. She also served as the lead writer in several grant proposals, including Gates Foundation Grand Challenges and LG Chem. She met virtually with several potential collaborators to discuss ways to combine forces and maintained as contact with current members to keep projects fresh. In other work, she taught three

undergraduate courses in Reliability and Maintainability, Operations Research, and Engineering Strategies.

Andrew K. S. Jardine, Professor Emeritus
Andrew has been active throughout the first half of 2021, serving on various committees as a reviewer, board member and more. He was a reviewer for papers submitted to the 15th World Congress on Engineering Asset Management (15th WCEAM), hosted online, but based out of Bonito Brazil, August 15-18, 2021. He is serving on the advisory board of Mechanical Systems and Signal Processing (MSSP) as an appointed member as of 2021. On February 8, Andrew was invited as a member of the editorial board for the *Journal of Maintenance, Reliability and Condition Monitoring* (jMARC), a new open access journal with Professor Jyoti K. Sinha from the University of Manchester as Editor in Chief. Through March 6, 7, 27, and 28, 2021, Andrew taught a graduate class in Maintenance Analysis and Optimization at the University of the West Indies in Trinidad. In April, he submitted the manuscript for the 3rd Edition of *Maintenance, Replacement and Reliability: Theory and Application* to CRC Press. The official publication date is 2022. The University of Western Australia appointed Andrew as an assessor for promotion of an academic to the rank of Professor, Level E, in April. He continued his work on the PEMAC awards committee, with meetings on April 16 and 27th. He was on the board of examiners to assess a PhD thesis titled *Development of Strategies for Effective Maintenance of Induction Motors*, Indian Institute of Technology (IIT) Kharagpur, on May 10, 2021. For the upcoming 11th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR) taking place June 29-July 1, he is a member of the program committee.

Dragan Banjevic, C-MORE Consultant
Dragan continued to collaborate with C-MORE on projects with consortium members, mostly with Kinross Gold, TTC, and MOD. He is currently busy with the implementation of the MOD spares management project. He has also provided help in other projects with C-MORE students, as well as in their research.

Sharareh Taghipour, Ryerson, External Collaborator
From January to June 2021, Sharareh lectured one session on the role of emerging technologies in physical asset management (University of West Indies, graduate course, with Andrew Jardine). She is working on a project entitled "Decentralized data analytics and optimization methods for Physical Asset Management" (NSERC Discovery grant), as well as two collaborative projects with industry: "Real-time optimization of production scheduling" and "Developing a decision support tool to optimize manufacturing for productivity and safeguarding the workforce against COVID-19," both with Axiom Group (NSERC Alliance). In addition, she is using the "Industry 4.0 Smart Factory System" to develop predictive maintenance models and real-time optimization of production scheduling (funding from the Ministry of Economic Development, Job Creation and Trade and John R. Evans Leaders Fund). She recently finished her collaborative project entitled "Developing methods for measuring social, economic, and environmental impacts of maintenance activities for physical assets," with Fiix Inc. (NSERC CRD). Sharareh submitted a book chapter entitled "The role of emerging technologies in physical asset management" for the third edition of *Maintenance, Replacement and Reliability* by Andrew Jardine and Albert Tsang. Sharareh also presented "The role of emerging technologies in physical asset management" at the 15th International Physical Asset Management Conference, online, March 9-10, 2021, Tehran, Iran

Scott Sanner, University of Toronto

Scott's group continued to work on a range of applied projects covering predictive modelling for residential HVAC (article published in *Journal of Building Performance Simulation*) and text analytics for urban studies (article published in *Cities*). Scott's group has also continued fundamental AI research on the topic of continual learning for deep learning (with one conference paper published at AAAI-21), as well as conversational recommender systems (a conference paper published at WWW-21).

Fae Azhari, University of Toronto
Fae's research group now consists of 5 doctoral students, 4 MASc students, and 4 undergraduate students. Her projects include: complex naval asset management using sensor data, optimizing the fabrication and performance of multifunctional cementitious composites, fibre optic sensors for vibration monitoring, sensing systems for prostheses and gait analysis, bridge scour monitoring, condition-based maintenance of bridges, and swimming pool freshwater usage optimization. Her group members submitted a number of journal articles and presented at virtual conferences this past year. Fae has initiated a number of new collaborations with colleagues in psychology, physiotherapy, medicine, and electrical, chemical, and biomedical fields.

Jue Wang, Affiliate Professor
Jue Wang is an Assistant Professor at Smith School of Business Queen's University. He is currently supervising/co-supervising two PhD students on sequential decision making. He also started working on a new project on optimal learning and optimization when the system can stop due to random failure. He and his collaborator meet online several times a week to discuss this project.

Ali Zuashkiani, Director of Educational Programs
Ali has been busy providing consulting services to various industries, such as oil and gas, power generation and distribution, mining, and petrochemical. He has been particularly active working with JESCO steel company in Saudi Arabia and DP World in Dubai on the topic of spare parts management practice and with SABIC on the topic of RCM. He has been also working with Aramco in Saudi to train their engineers to get CMRP designation. Ali has been assisting the Institute of Asset Management in the development of a Subject Specific Guideline (SSG) on Management of Change and is working with General Forum on Maintenance and Asset Management (GFMAM) on the third edition of *Asset Management Landscape*. Ali has started working with Society for Maintenance and Reliability Professionals (SMRP); he will represent SMRP at GFMAM meetings and is working on a new body of knowledge of SMRP. Most training programs were cancelled or conducted online due to COVID-19, but a few managed to go through. These appear later in the Executive Summary under "Educational activities."

C-MORE graduate students

Doctoral students

Kuilin Chen, a third-year PhD student, is working on few-shot learning. One paper on incremental few-shot learning will be published in ICLR 2021; he also submitted a paper on few-shot learning with stochastic weight averaging to NeurIPS 2021.

Michael Gimelfarb has continued his doctoral work on knowledge transfer in reinforcement learning and has been a postgraduate affiliate of the Vector Institute since April 2020. He is currently preparing two papers for submission to the NeurIPS Conference in 2021.

Scott Koshman continues research on equipment health monitoring (EHM) for Halifax Class Frigates under the supervision of Fae Azhari. As a member of the International Council of Systems

Engineering (INCOSE), he recently participated in the 1st Virtual INCOSE International Workshop (IW2021). In his workplace, he is the lead in the development of graduate courses and the transition to a virtual delivery model in global strategic security studies.

Seyedvahid Najafi is a full-time doctoral student working on the maintenance modelling and optimization of multi-unit series systems. A recent paper is under review by *Reliability Engineering and Safety Systems* and another has been accepted by the 11th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR).

Avi Sokol is a flex-time PhD student and full-time Business Data Scientist and Inventory Specialist who is researching the integration of Reinforcement Learning and Inventory Control to reduce waste in supply chains. He is currently working towards developing a novel algorithm with discrete-continuous action space and reward decomposition.

Songci Xu is applying transfer learning to Intelligent Fault Diagnosis (IFD), one of the projects of LG Sciencepark. He is also working on the explanatory ability of deep domain adaptation via Optimal Transport Theory.

Master's students

Tushar Aggarwal (MEng) is working on artificial intelligence and robotics. He completed a project on anomaly detection on train tracks using thermal imaging.

Varun C Ananda Rao (MEng) is working on a carrier optimization project; the goal is to build a model that provides an optimized solution for maintaining inventory on the carrier.

Ozkan Elmali (MASc) is working on a reinforcement learning approach to the vehicle routing and safe routing problem. As a newly registered student, he is currently taking several courses.

Sahil Nagpal (MEng) is a part-time student who is currently working at Magna International. The main goal of his project is to implement a Predictive Maintenance protocol for critical components of a manufacturing equipment presently operating at one of Magna's Manufacturing facilities.

Jeong Cheol Seok (MEng) is working on project to develop a deep learning model to predict the remaining useful life (RUL) of engines used in a mining site. He is also working on a project to develop a machine learning model that measures actual wind parameters such as speed and direction using a LiDAR measurement floating on sea surface. The preliminary model has been completed and is awaiting instructions from the requested company.

Pooyan Sharifi (MEng) is working on a project on asset management and criticality analysis for a medium sized power generation company in the GTA.

Zoha Sherkat Masoumi (MEng) is currently working on sensor-based maintenance for Kinross Gold Corporation. She is aiming to predict engine failures for trucks in real mine operating environment with the help of sensor readings. Her goal is to reduce the number of unscheduled truck failures to reduce downtime and save money.

Jahyun Shin (MEng) is primarily interested in analytics and machine learning. Under Chi-Guhn Lee, she is taking charge of an early-stage deep learning research project to develop a model that can automatically detect and classify restricted items for an airport's X-ray baggage scanner. She is also applying adversarial domain adaptation techniques to compensate for the small number of labelled X-ray images provided by the stakeholder

Sophie Tian (MASc) is in the first year of the program. She has been taking coursework and is conducting a literature review on time series classification methods.

Katie Xu (MASc) has been working with collaborators in the Department of Materials Science and Engineering on the use of machine learning techniques for in-situ process monitoring and control in 3D printing systems. She has also been completing coursework.

Zihan Zhang (MASc) has been completing coursework and working on a literature review on PHM application. In the summer session, she will work on the battery project with undergraduates.

Independent thesis project
Kimia Taghvaei Ganjali (IndE) graduated this year, and is now working as a logistics continuous improvement analyst at Walmart. She completed the CEA data audit project as her thesis. The objective was to build a framework for CEA to help assess their records and build a consistency throughout all sites working with CEA.

C-MORE activities with consortium members

Defence Science and Technology Laboratory (DSTL)
The aircraft carrier spares optimization project continued under Dragan's supervision with MEng student Varun C. Ananda Rao. They are developing a software tool to optimally stock spare parts. This project will be presented today. A second project in interpreting and fusing data from disparate environmental conditions began during this time with Siddharth Patel, an MEng student. This work is in its early stages.

Department of National Defence (DND)
C-MORE wrapped up the propulsion diesel engine project with recommendations on data collection and the implications of the oil analysis results.

Kinross Gold Corporation
We continued work on the machine learning business case project, with Zoha Sherkat-Masoumi and Theresa Taylor leading the project. The interim objective is to predict the remaining useful life (RUL) of the haul truck engines using sensor data. The current status of this project will be presented today.

Toronto Transit Commission (TTC)
TTC set a distinct timeline of interest for the linetest and re-inspection projects for decision making. The updated results of these projects will be presented today.

C-MORE educational programs
In December 2020, Ali Zuashkiani gave a five-day Physical Asset Management online course to a group of asset managers from different industries. In April 2021, Ali delivered a 5-day CMRP preparation training program to group of reliability engineers from Aramco in Saudi. Ali has been working with LEORON and PAMCo to deliver more C-MORE certified training programs including an upcoming five-day Asset Management 4.0 program that will be given online during August 2021 and a course on Spare Parts Management and Life Cycle Costing Management in September 2021.

Janet and Chi-Guhn's new, complementary five-day course, Machine Learning and AI Applications in Physical Asset Management, or "PAM2," will be delivered virtually in late June 2021; it is offered through the School of Continuing Studies. We are very excited about this new course and see it as an important addition to the C-MORE repertoire.

In November 2021, as in previous years, Andrew Jardine will be leading a five-day course in Physical Asset Management through the School of Continuing Studies. Other lead instructors are Don Barry and Sharareh Taghipour. Chi-Guhn Lee will be presenting as well. The course is consistently excellent, with high ratings from attendees. As in November 2020, it will be delivered virtually.

Conclusion

I generally conclude the Executive Summary with a retrospective look at the past six months, but in this case, I want to say a special thank you to everyone at C-MORE for their hard work, their dedication, and, above all, their creativity over the past year. Collectively, we have used the pandemic as an opportunity to change and grow. Thanks to all!

Chi-Guhn Lee
June 2021

C-MORE leadership activities

Chi-Guhn Lee, Director

Chi-Guhn has leveraged the all-virtual environment to give invited talks and seminars at many conferences around the world. While the detailed list of talks is reserved for the Activities section of the report, one initiative of note is spearheading the resource extraction working group. This group met three times throughout the first part of 2021 to discuss ways C-MORE could better meet the needs of the mining sector. Outside of this particular sector, Chi-Guhn is increasing the presence of C-MORE in the asset management and maintenance optimization. On June 17, Chi-Guhn will be hosting a workshop through the PEMAC GTA chapter, titled "Demystifying machine learning – hands on workshop using Excel." In this term, we submitted three project proposals, and signed one project contract: The Canadian Institute for Advanced Research (CIFAR) Solution Network Proposal on Digital Supply Chain for Food Products in collaboration with IIT-Mumbai, Gates Foundation Proposal on Digital Farming in collaboration with IIT-Mumbai, LG Chemistry Proposal on Degradation Model, and a contract with Titan Technologies for a LiDAR data compensation model.

Janet Lam, Assistant Director

Janet has been working with graduate and undergraduate students to push on with various industry-sponsored research projects. In particular, she worked with CEA on a data audit project, Capital Power on a criticality project, Kinross on a machine learning case studies project, Titan Technologies on two different wind-turbine projects. She also served as the lead writer in several grant proposals, including Gates Foundation Grand Challenges and LG Chem.

She met virtually with several potential collaborators to discuss ways to combine forces, as well as renewed contact with current members to keep projects fresh.

In other work, she taught three undergraduate courses in Reliability and Maintainability, Operations Research, and Engineering Strategies.

Andrew K. S. Jardine, Professor Emeritus

Andrew has been active throughout the first half of 2021, serving on various committees as a reviewer, board member and more. He was a reviewer for papers submitted to the 15th World Congress on Engineering Asset Management (15th WCEAM), hosted online, but based out of Bonito Brazil in August 15-18. He is serving on the advisory board of Mechanical Systems and Signal Processing (MSSP) as an appointed member as of 2021. On February 8, Andrew was invited

as a member of the editorial board for the Journal of Maintenance, Reliability and Condition Monitoring (jMARC), a new open access journal with Professor Jyoti K. Sinha from the University of Manchester as Editor in Chief. Through March 6, 7, 27 and 28, he taught a graduate class in Maintenance Analysis and Optimization, at the University of the West Indies in Trinidad. In April , he submitted the manuscript for the 3rd Edition of Maintenance, Replacement and Reliability: Theory and Application to CRC Press. The official publication date is 2022. The University of Wester Australia appointed Andrew as an assessor for promotion of an academic to the rank of Professor, Level E in April. For the upcoming 11th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR) taking place June 29 – July 1, he is a member of the programme committee, and on the board of examiners to assess the PhD thesis titled Development of Strategies for Effective Maintenance of Induction Motors, Indian Institute of Technology (IIT) Kharagpur on May 10, 2021. He continues his work on the PEMAC awards committee, with meetings on April 16 and 27th.

Dragan Banjevic, C-MORE Consultant

Dragan continued to collaborate with C-MORE on projects with consortium members, mostly with Kinross Gold, TTC, and MOD. He is currently busy with implementation of MOD Spares management project. He also provided help in other projects with C-MORE students, as well as in their research.

Sharareh Taghipour, Ryerson University, External Collaborator

From January to June 2021, Sharareh lectured one session about the role of emerging technologies in physical asset management (University of West Indies, graduate course, with Prof. Andrew Jardine). She is working on a project entitled “Decentralized data analytics and optimization methods for Physical Asset Management” (NSERC Discovery grant), as well as two collaborative projects with industry: “Real-time optimization of production scheduling” and "Developing a decision support tool to optimize manufacturing for productivity and safeguarding the workforce against COVID-19", both with Axiom Group (NSERC Alliance). In addition, she is using the “Industry 4.0 Smart Factory System” to develop predictive maintenance models and real-time optimization of production scheduling (funding from the Ministry of Economic Development, Job Creation and Trade and John R. Evans Leaders Fund). She recently finished her collaborative project entitled “Developing methods for measuring social, economic, and environmental impacts of maintenance activities for physical assets,” with Fiix Inc. (NSERC CRD). Dr. Taghipour also submitted a book chapter entitled "The role of emerging technologies in physical asset management" for the third edition of the book "Maintenance, Replacement and Reliability" by Andew Jardine and Albert Tsang. Sharareh also presented "The Role of emerging technologies in physical asset management" at the 15th International Physical Asset Management Conference. Online, March 9 and 10, 2021, Tehran, Iran

Scott Sanner, University of Toronto

Scott's group continues work on a range of applied projects covering predictive modelling for residential HVAC (journal article published in Journal of Building Performance Simulation), and text analytics for urban studies (journal article published in Cities). Scott's group also continues fundamental AI research on the topic of continual learning for deep learning (with one conference paper published at AAAI-21) as well as conversational recommender systems (with a conference paper published at WWW-21).

Fae Azhari, University of Toronto

Fae's research group now consists of 5 doctoral students, 4 MASc students, and 4 undergraduate students. Her projects include: complex naval asset management using sensor data, optimizing the fabrication and performance of multifunctional cementitious composites, fibre optic sensors for vibration monitoring, sensing systems for prostheses and gait analysis, bridge scour monitoring, condition-based maintenance of bridges, and swimming pool freshwater usage optimization. Her group members submitted a number of journal articles and presented at virtual conferences this past year. Fae has initiated a number of new collaborations with colleagues in psychology, physiotherapy, medicine and electrical, chemical and biomedical

Jue Wang, Affiliate Professor

Jue Wang is an Assistant Professor at Smith School of Business Queen's University. He is currently supervising/co-supervising two PhD students on sequential decision making. He also started working on a new project on optimal learning and optimization when the system can stop due to random failure. He and his collaborator discuss several times a week online for this project.

Ali Zuashkiani, Director of Educational Programs

Ali has been active in providing consulting services to various industries such as oil and gas, power generation and distribution, mining, and petrochemical. He has been in particular active in working with several companies including JESCO steel company in Saudi Arabia and DP World in Dubai to improve their spare parts management practice and with SABIC on the topic of RCM. He has been also working with Aramco in Saudi to train their engineers to get CMRP designation.

Ali has been also active with Institute of Asset Management on developing a Subject Specific Guideline (SSG) on Management of Change and is working with General Forum on Maintenance and Asset Management (GFMAM) on the third edition of Asset Management Landscape. Ali has started working with Society for Maintenance and Reliability Professionals (SMRP) to represent them at GFMAM meetings and to work on new body of knowledge of SMRP.

Majority of training programs were cancelled or conducted online due to COVID-19 however few managed to go thorough.

In Dec 2020, Ali conducted 5-day Physical Asset Management training online to a group of asset managers from different industries.

In April 2021 Ali delivered a 5-day CMRP preparation training program to group of reliability engineers from Saudi Aramco oil company.

Ali has been working with LEORON and PAMCo to deliver more CMORE certified training programs including an upcoming 5-day Asset Management 4.0 program that will be delivered online during Aug 2021 and a course on Spare Parts Management and Life Cycle Costing Management in September 2021.

Overall project direction

Janet Lam, Assistant director

Goals and retrospectives

This section highlights the some of the main achievements in C-MORE for the period January 2021 – June 2021. Throughout the months of the pandemic, the C-MORE team continued to operate with work-from-home modifications, with all meetings happening virtually.

As with many other organizations, we attended our first virtual conferences, including hosting our own progress meetings online. Our PAM2 course is scheduled for late June of 2021; it is being planned for virtual delivery.

We continue to seek opportunities to leverage our work with external funding. Some proposals submitted include LG Chem Global Innovation Contest, Gates Foundation Grand Challenges for Smart Farming, and Canadian Institute for Advanced Research AI governance solutions.

Beyond our continuous work with consortium members, we have engaged with other industry partners on an ad-hoc basis, including Canadian Electricity Association and Capital Power. Projects from both of these partners will be presented today.

Activities

Collaboration with companies and site visits

This section gives details on progress in research conducted with consortium members

Member	Collaborations
Defence Science and Technology Laboratory	The aircraft carrier spares optimization project continued under Dragan's supervision with M.Eng. student Varun C. Ananda Rao. They are developing a software tool to optimally stock spare parts. This project will be presented today. A second project in interpreting and fusing data from disparate environmental conditions began during this time with Siddharth Patel, an M.Eng student. This work is in its early stages.

Member	Collaborations
Department of National Defence	C-MORE wrapped up the propulsion diesel engine project with recommendations on data collection, and the implications of the oil analysis results.
Kinross	We continued work on the machine learning business case project with Zoha Sherkat-Masoumi and Theresa Taylor leading the project. The interim objective is to predict the remaining useful life of the haul truck engines using sensor data. The current status of this project will be presented today.
Toronto Transit Commission	TTC set a distinct timeline of interest for the linetest and re-inspection projects for decision making. The updated results of these projects will be presented today.

Theoretical work

This section on theoretical work is oriented toward students' and postdoctoral fellows' research topics.

Name	Activity
Tushar Aggarwal, M.Eng. student	Tushar is currently in his first year of M.Eng in Mechanical Engineering program focused towards artificial intelligence and robotics. He completed his project on anomaly detection on train tracks using thermal imaging.
Varun C Ananda Rao, M.Eng. student	Varun is a first year MEng student in the Mechanical and Industrial engineering program focused on supply chain management and optimisation. He is working on carrier optimisation project focusing on building a model that provides an optimised solution for maintaining inventory on the carrier. His work is focused on the implementation of the optimisation model and thereby creating an interface for users to easily obtain the solution.
Kuilin Chen, Ph.D. candidate	Kuilin is a third-year Ph.D. student. His current research interest is few-shot learning. One paper on incremental few-shot learning is published in ICLR 2021. He submitted a paper on few-shot learning with stochastic weight averaging to NeurIPS 2021.
Ozkan Elmali, M.ASc. student	Ozkan is a new M.ASc student working on a reinforcement learning approach to the vehicle routing and safe routing problem. As vehicle routing is known to be a computationally hard problem, there is great potential in taking a RL approach. Safe routing is a special case of vehicle routing for transporting hazardous materials that have additional restrictions and a heavier weight to reducing crash risks. As a newly registered student, he is currently taking several courses.
Michael Gimelfarb, Ph.D. candidate	Michael has continued his doctoral work on knowledge transfer in reinforcement learning and has been a postgraduate affiliate of the Vector Institute since April 2020. Currently, his research focuses on

Name	Activity
	transferring skills robustly and safely in a risk-aware setting. His current work leverages robust and risk sensitive MDPs, representation learning and planning under uncertainty. He published three papers since 2020 and is currently preparing two papers for submission to the NeurIPS Conference in 2021
Scott Koshman, Ph.D. student	Scott continues his research on equipment health monitoring (EHM) for Halifax Class Frigates, under the supervision of Professor Fae Azhari. His recent focus has been data conditioning, the optimum application of parallel assets to the analysis of largish data sets (40+ billion transactions), and the fusion of data across databases. He works with data from diverse sources including EHM systems, an ERP, internal reporting, and external public environmental data. This research will inform approaches for the development of maintenance optimization models given certain types of imperfect data inputs. Scott is a senior member of American Society of Quality and recently recertified both his Certified Reliability Engineer and Certified Manager of Quality / Organizational Excellence credentials. As a member of the International Council of Systems Engineering (INCOSE), he recently participated in the 1st Virtual INCOSE International Workshop (IW2021). In his workplace, he has also been the lead for the development of graduate courses and the transition to a virtual delivery model in the subject area of global strategic security studies.
Sahil Nagpal, M.Eng. student	Sahil Nagpal is a part-time M.Eng student, who is currently working at Magna International. The main goal of his project is to implement a Predictive Maintenance protocol for critical components of a manufacturing equipment which is currently operating at one of Magna's Manufacturing facilities. Sahil is looking to utilize Weibull analysis, as well as investigating the use of Cox Proportional Hazard Models with various sensor readings as predictors for the remaining useful life of the components. The project has been assigned by Magna International.
Seyedvahid Najafi, Ph.D. student	Vahid is a full-time Ph.D. student and works on the maintenance modeling and optimization of multi-unit series systems. In his recent paper, which is under review by the Reliability Engineering and Safety Systems journal, an opportunistic maintenance policy with general repair is developed for a two-unit series system, in which the condition of one unit is monitored, and only age information is available for the other unit. The problem is formulated in the semi-Markov decision process framework, and an algorithm is developed to find the optimal policy that minimizes the long-run average cost per unit time. A paper entitled “A condition-based maintenance policy for a two-unit system subject to dependent soft and hard failures: A reinforcement learning approach” has been accepted by the 11th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR).

Name	Activity
	His research is now focused on developing condition-based maintenance policies for multi-unit systems, where all units are subject to condition monitoring, using deep reinforcement learning.
Jeong Cheol Seok, M.Eng. student	Jeong Cheol Seok is a 2nd year MEng student in Mechanical and Industrial program with emphasis in data analytics. He is currently working on project to develop a deep learning model that predicts the remaining useful life (RUL) of engines used in a mining site. He has been exploring neural networks such as CNN – LSTM, and GRU for prediction of the RUL. He is working on another project to develop a machine learning model that measures for the actual wind parameters such as speed and direction respect to ground using the measurement of a LiDAR measurement floating on sea surface. The preliminary model has been completed and is waiting for the further instructions from the requested company.
Pooyan Sharifi, M.Eng. student	Pooyan is a second year M.Eng student in the Mechanical and Industrial Engineering. His project is focused on asset management and criticality analysis for a medium sized power generation company in the GTA. Pooyan completed his undergraduate degree at the University of Waterloo in Chemical Engineering with a specialization in Process Modelling, Optimization and Control. He also has over two years of experience working full-time in the power generation industry as well as experience working in other industries including automotive manufacturing and the defence industry. He is excited to utilize his blend of academic and industrial experience to tackle his M. Eng project.
Zoha Sherkat Masoumi M.Eng. student	Zoha is a second year Meng student in Mechanical and Industrial Engineering. Her area of specialty is optimization and data analytics. She is currently working on sensor-based maintenance for Kinross Gold Corporation. She is aiming to predict engine failures for trucks in real mine operating environment with the help of sensor readings. She is implementing deep learning algorithm that predicts failures 3 hours before it occurs. Her goal is to reduce the number of unscheduled truck failures to reduce downtime and save money.
Jahyun Shin, M.Eng. student	Jahyun is a first year M.Eng student in Mechanical and Industrial Engineering program with emphasis in analytics and machine learning. Under Professor Chi-Guhn Lee, she is taking charge of an early-stage deep learning research project to develop a model that can automatically detect and classify restricted items for an airport's X-ray baggage scanner. She is also applying adversarial domain adaptation techniques to compensate for the small number of labeled X-ray images provided by the stakeholder
Avi Sokol, Ph.D. student	A flex-time PhD student and a full-time Business Data Scientist and Inventory Specialist, Avi is researching the integration of Reinforcement Learning and Inventory Control to reduce waste in

Name	Activity
	supply chains. Recently Avi has passed his first qualifying exam and currently working towards developing a novel algorithm with discrete-continuous action space and reward decomposition.
Kimia Taghvaei Ganjali, IndE thesis student	Kimia graduated this year, and is now working as a logistics continuous improvement analyst at Walmart. She completed the CEA data audit project as her thesis. The objective of the project was to build a framework for CEA to help assess their records and build a consistency throughout all sites working with CEA.
Sophie Tian, M.ASc. student	Sophie started her M.ASc. in September 2020. She has been taking the courses Neural Networks and Deep Learning and the Scientific Writing Course during the last few months. She is working on the combustion signature analysis project in collaboration with NRC and has been conducting literature review on time series classification methods.
Katie Xu, M.ASc. student	Katie started her M.ASc. in September 2020. She has been working with collaborators in the Department of Materials Science and Engineering on the use of machine learning techniques for in-situ process monitoring and control in 3D printing systems. To this end, she has been studying computer vision and establishing requirements for the physical system. Additionally, she took two courses during the previous semester: Neural Networks and Deep Learning, and Markov Decision Processes.
Songci Xu, Ph.D. student	Songci has started his first year Ph.D. program since the January of 2020. His current research focuses on applying transfer learning onto Intelligent Fault Diagnosis (IFD), which is also one of the projects of LG Sciencepark. He is also working on the explanatory ability of deep domain adaptation via Optimal Transport Theory, which counts towards his thesis.
Zihan Zhang, M.ASc. student	During the past 6 months, Zihan has finished 2 courses (Markov Decision Process, Scientific Writing Course), continued literature review in PHM application and proposed four project proposals: (a) Hierarchical maintenance optimization considering replacement impact; (b) Short-time-Fourier-transform-based fault diagnosis using Graph Convolution Network; (c) Health-oriented group maintenance; and (d) Prognosis management for rechargeable batteries integrating deep neural network and physical-stochastic processes. She will proceed battery project in summer session with undergraduates.

Visits and interactions with consortium members and others

December 2020 – June 2021

All meetings during this period were virtual, in accordance with safe COVID-19 protocols.

December, 2020
Ali conducted a 5-day Physical Asset Management training online to a group of asset managers from different industries.

Biweekly throughout **Kinross**
Zoha Sherkat-Masoumi, Seyedvahid Najafi and Janet met with Theresa Taylor of Kinross on a bi-weekly basis throughout, to discuss progress on the machine learning case studies project. This project will be presented today.

January 7, 2021 **Eaigle**
Fae Azhari met with Amir Hoss of Eaigle on the use of sensors and AI for swimming pool freshwater usage optimization.

January 8, 2021
Fae attended a meeting of the Transportation Research Board AKT40 committee on Structures Maintenance.

January 21, 2021 **CEA**
Janet and Kimia met with Dan Gent and Asuka Boehm of CEA to discuss progress on the data audit project, and plans for the upcoming term.

January 26, 2021 **Kroon**
Chi-Guhn, Janet, and Dragan met with Michael Kroon, Huthaifa Abderahman, and Jeff Gaudreau of Kroon to discuss potential collaboration opportunities.

January 27, 2021
Chi-Guhn, gave an invited talk titled “MDP and Reinforcement Learning” to undergraduate students in the Machine Intelligence Option of the Engineering Science Class

February 1, 2021 **Capital Power**
Janet met with Bill Mercer of Capital Power to discuss the administrative details of the criticality project.

February 1, 2021
Chi-Guhn and Katie Xu met with Professor Yu Zou and Jiahui Zhang to discuss their project direction and data requirements. Katie presented some basics of reinforcement learning.

February 3, 2021 **Xtivity**
Janet met with Kevin Gates and Andrew Jordan of Xtivity to share knowledge and explore collaboration opportunities.

February 8, 2021 **TTC**
Janet and Dragan met with Tim Southworth, Tauqeer Qurashi, Mo Ghaus, Jennifer Lu, Hossein Mohammadian, Ali Rezaie, and Rezwan Razzaque to report on progress of the re-inspection project and to discuss changes to the interpretation of the dataset.

February 8, 2021 **jMARC**
Andrew was invited to the Editorial board for a new open access journal – Journal of Maintenance, Reliability and Condition Monitoring (jMARC), with editor in Chief Professor Jyoti K. Sinha, The University of Manchester, UK

February 16, 2021 **DND**
Janet met Abaida Al-Azzawi and Dan Saulnier of DND to discuss results of the propulsion diesel engine CBM project, and recommend next steps.

February 16, 2021 **AI2HR**
Janet met with Richard Beer and Michael Rosenberg to consult on companies that provide AI implementation services in the context of human resource management.

February 17, 2021 **Jun-Ju University**
Chi-Guhn gave a seminar titled "Machine Learning: Challenges and Our Efforts to Overcome" to researchers at Jun-Ju University in South Korea.

February 18, 2021 **CEA**
Janet and Kimia met with Dan and Asuka of CEA for a monthly project update. An idea to compare states 24 and 25 of each plant was considered.

February 18, 2021 **Concordia University**
Chi-Guhn gave an invited talk titled "Reinforcement Learning on Financial Optimization" to graduate students and researchers at Concordia University.

Feburary 25, 2021 **STNG**
Chi-Guhn, Janet, Dragan and Jeong-Cheol Seok met with Pedro Cancino and Daniel Diaz from Asesorías y Servicios Tribológicos del Norte Grande S.A to discuss a mining haul truck predictive analytics project.

March 1, 2021 **Capital Power**
Janet and Pooyan Sharifi met with Bill Mercer, Robert Mozzoni, Adrian Martinez, Iain Ogilvie, Juri Baroskov, Brent Tarnowski and Terry Myers to kick off the Goreway plant criticality project. It was decided to start with the boiler feed pumps as the first equipment of interest.

March 3, 2021 **Titan**
Chi-Guhn, Janet, Dragan and Jeong-Cheol met with Joe Xu, Yinxing Ma, Luffy Zhou and Qiang Zhao of Titan to discuss the buoy-base LiDAR compensation project. The objective of the meeting was for data description and project kick-off.

March 6, 7, 27, 28, 2021 **UWI**
Andrew taught a graduate class titled Maintenance Analysis and Optimization, at The University of the West Indies, Trinidad.

March 9, 10, 2021 **IPAMC**
Sharareh also presented "The Role of emerging technologies in physical asset management" at the 15th International Physical Asset Management Conference.

March 12, 2021
Katie met with Jiahui to receive training on how to use the 3D printer for her research project.

March 18, 2021 **CEA**
Janet and Kimia met with Asuka to share the results of the data audit project. The final deliverables were agreed upon. The results of this project will be presented in today's meeting.

March 30, 2021 **DND**
Chi-Guhn and Janet met with Dan to wrap up the propulsion diesel engine project. A summary of the whole project was provided, along with a final report and recommendations.

Weekly starting from March 31, 2021 **Capital Power**
Janet and Pooyan, met with Bill, Robert, Adrian and Iain to report weekly on the progress of the criticality project. The current status of the project will be presented today.

March 26; April 23; May 28
The resource extraction special group, (Chi-Guhn, Janet, Dragan, Emilio Sarno, Jean-Pierre Pascoli, Alberto Van Oordt, Joe Ashun, Simone Smith, Marcelo Aliendre) met for the first time to discuss potential ways C-MORE can better serve the resource extraction industry. A main gap of knowledge and language between practioners, managers and researchers was identified.

April, 2021
Ali delivered a 5-day CMRP preparation training program to group of reliability engineers from Saudi Aramco oil company.

Andrew was appointed Assessor by University of Western Australia for promotion of an academic to the rank of Professor, Level E.

April 5, 2021
Andrew submitted to the publisher the manuscript for the 3rd Edition of Maintenance, Replacement and Reliability: Theory and Applications, CRC Press.

April 16, 27, 2021 **PEMAC**
Andrew met with the PEMAC awards committee as a member.

April 27, 2021 **TTC**
Chi-Guhn and Janet met Mo Ghaus from TTC to discuss ways that C-MORE could serve more departments of TTC. C-MORE committed to delivering a presentation that proposed potential projects with TTC.

April 28, 2021 **Intellistruct**
Fae met with Brian Westcott of Intellistruct on structural health monitoring for resilient and smart bridges.

April 29, 2021 **Titan**

Chi-Guhn met with Joe to discuss the objectives of the wind turbine AI-algorithm development project.

May 3-7, 2021
Kuilin Chen gave a presentation titled "incremental few-shot learning via vector quantization in deep embedded space" at The Ninth International Conference on Learning Representations, 2021.

May 10, 2021 **IIT**
Andrew was a member on the Board of Examiners to examine PhD thesis titled Development of Strategies for Effective Maintenance of Induction Motors, Indian Institute of Technology (IIT) Kharagpur.

May 24, 2021 **IISE**
Mike Gimelfarb gave an invited talk titled "\epsilon-BMC: A Bayesian Ensemble Approach to Epsilon-greedy Exploration in Model-free Reinforcement Learning" at the Annual Conference and Expo 2021 hosted by the Institute of Industrial and Systems Engineers.

May 26, 2021 **CSCE**
Fae attended a meeting with The Canadian Society for Civil Engineering (CSCE) Structures Division.

June 3, 2021 **Titan**
Janet, Shelina Terrance and Sylvia Chen met with Joe, Yinxing and Qiang of Titan to discuss details on the data for the wind turbine AI-algorithm development project.

June 7-10, 2021 **CORS**
The following presentations were given at the Canadian Operational Research Society (CORS) annual meeting 2021.

Kuilin gave a presentation titled "Attentive gaussian processes for probabilistic time-series generation".

Mike Gimelfarb gave a presentation titled "Improving the Sample Efficiency of Model-free Reinforcement Learning using Bayesian Mixtures of Experts", co-authored with Chi-Guhn.

Amine Aboussalah gave a presentation titled "Symmetry Augmentation for Time-series Reinforcement Learning", co-authored with Chi-Guhn.

Babatunde Giwa gave a presentation titled "High Dimensional Continuous Reinforcement Learning for Finance", co-authored with Chi-Guhn.

Wentao Liu gave a presentation titled "Risk-sensitive Portfolio Investment Strategy Selection with Distributional Reinforcement Learning", co-authored with Chi-Guhn.

Kinross: A business case for machine learning methods

Zoha Sherkat Masoumi

Introduction

Kinross Gold corporation is a gold mining company. Gold from the mines are carried with trucks from mines to facilities and distribution centers. Trucks are expensive assets at the same time, they often fail and cause excessive downtime and cost. As a result, the company aims to detect failures before they occur with the help of sensors. In this project sensor recordings from 2019 are utilize to train machine learning algorithm to predict failures that are expected to take place in 3 hours.

Data Preprocessing

Data

The data from the project is provided by Kinross Gold Corporation. The sensor recordings and unscheduled failure information is available on two different tables for 13 mining trucks for the year of 2019. There are total of 33 sensors however, only 25 of them have enough recordings to be considered for future analysis. At determined time intervals all 25 sensors have recorded information for each truck. There are no recordings during maintenance or when the truck is at rest. There are between 18,000 to 38,000 sensor recordings for each truck.

Remaining Useful Life (RUL)

The unscheduled failure table has been utilized to calculate the hours remaining to the next failure (Remaining Useful Life (RUL)) for each sensor recording. The overall RUL and RUL obtained for each failure mode for each recording is added as separate columns to sensor information table. Figure 1 displays the RUL for each sensor recording. There are some long and some short failure cycles. "DT79" has the longest cycle without failure (867.43 hours).

Table 1 Formatted sensor recordings table for "DT49"

	TimeStamp	Time	Asset	Atmospheric Pressure	Time_Format	Boost Pressure	Brake Air Pressure	Differential Lube Pressure	Differential Temperature (Oil)	Engine Coolant Temperature	...	Right Minus Left Exhaust Temperature
0	2019-01-03 16:21:41.000	1.546520e+09	DT49	100.5	2019-01-03 16:21:41	2.5	864.0	964.0	71.0	79.0	...	-17.0
1	2019-01-03 16:31:42.000	1.546521e+09	DT49	101.5	2019-01-03 16:31:42	140.5	783.0	835.0	71.0	82.0	...	-7.0
2	2019-01-03 17:01:42.000	1.546522e+09	DT49	102.0	2019-01-03 17:01:42	1.0	809.0	969.0	67.0	81.0	...	-11.0
3	2019-01-03 17:16:42.000	1.546523e+09	DT49	101.0	2019-01-03 17:16:42	2.5	828.0	896.0	69.0	85.0	...	-8.0
4	2019-01-03 18:06:42.000	1.546526e+09	DT49	100.5	2019-01-03 18:06:42	1.5	556.0	1094.0	64.0	69.0	...	13.0
...	...	...	...	...	...	...	...	...	...	...	...	...
26572	2019-12-31 08:06:59.000	1.577767e+09	DT49	100.5	2019-12-31 08:06:59	0.5	670.0	0.0	33.0	57.0	...	1.0
26573	2019-12-31 08:36:59.000	1.577769e+09	DT49	100.5	2019-12-31 08:36:59	1.5	729.0	1207.0	33.0	54.0	...	-13.0
26574	2019-12-31 08:41:59.000	1.577769e+09	DT49	100.5	2019-12-31 08:41:59	0.0	778.0	0.0	32.0	56.0	...	-16.0
26575	2019-12-31 09:31:59.000	1.577772e+09	DT49	100.5	2019-12-31 09:31:59	0.0	690.0	1215.0	31.0	52.0	...	-11.0
26576	2019-12-31 09:36:59.000	1.577772e+09	DT49	100.5	2019-12-31 09:36:59	1.0	773.0	0.0	31.0	55.0	...	-17.0

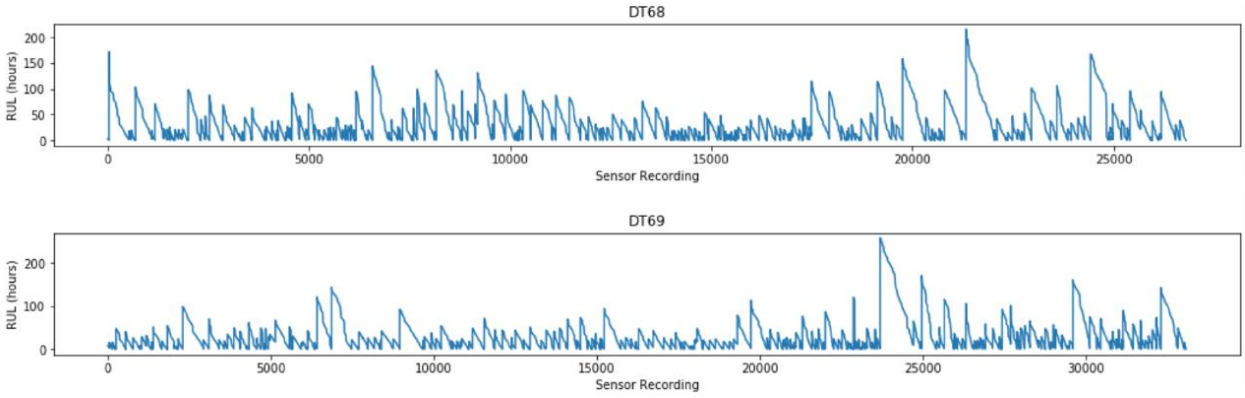


Figure 1 RUL vs. sensor recording for "DT68" and "DT69"

Failure Frequency and Duration

Unscheduled failure table includes eight failure modes: Dispatch hardware, Electrical, Engine, Hydraulics, Lubes/Oil , Operator damage, Structure, and Tire repair. In figure 2 displays jackknife scatter plot. y-axis of this figure represents the frequency of failure per truck and x-axis is the average downtime per truck in 2019. The horizontal and vertical red lines are the average failure frequency and average downtime. With the help of the red lines failures are divided into four different categories: chronic and acute, chronic, acute and neither chronic nor acute. Structure (frequency= 39.5, average downtime= 12.3 hours) and Engine (frequency= 33.5 , average downtime=11.7 hours) are both acute and chronic. Failures with high frequency and low average downtime are chronic and not acute such as Hydraulics (frequency= 34, average downtime= 2.4 hours) and Electrical (frequency= 33.5 , average downtime=3.0 hours).

The remaining failure modes are neither acute nor chronic. The investigation on duration and frequency failures confirms the following results:

- On average the number of daily shutdowns per vehicle is 0.48. Thus, unscheduled failures on average occur every other day.
- On average the duration of daily shutdowns due to unscheduled failure per vehicle is 3.3 hours
- On average for each unscheduled failure the duration of shutdowns per vehicle is 3.8 hours

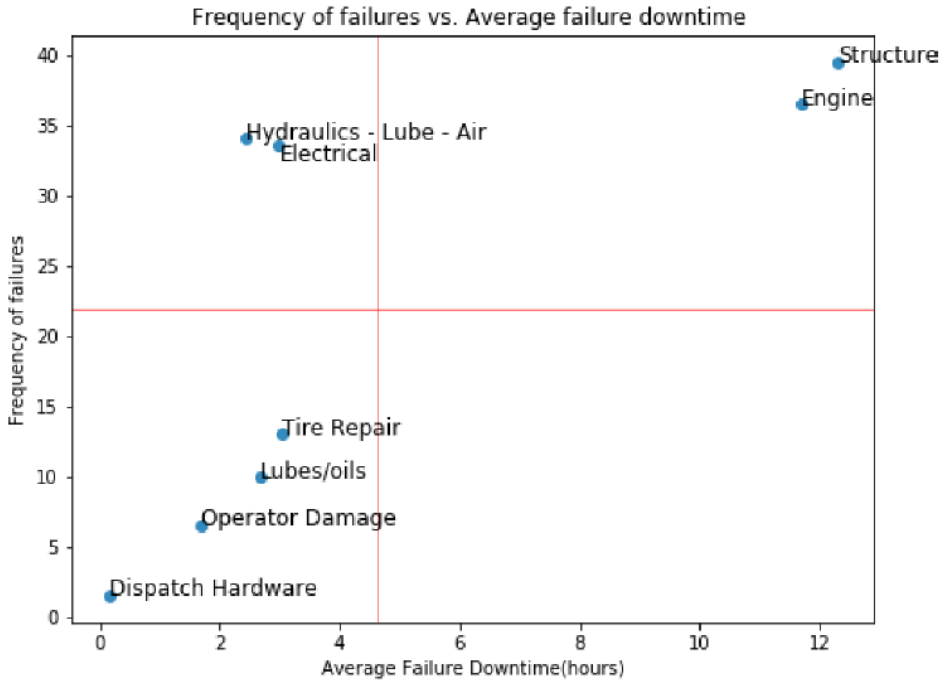


Figure 2 Jackknife diagram for "DT82" and "DT86"

Engine Failure

Based on our discussion with Kinross Gold Corporation representative the company is interested in Engine failure mode thus, for future analysis we our considering engine failures only. "DT66" has the largest number of unscheduled engine failures (46). However, for most of the trucks the number of engine failures is less than 20. The small number of engine failures cause severe imbalance in the dataset and increases the difficulty of this project. Figure 3 displays the number of engine failures for each truck during 2019. In addition, failure patterns are different for each truck, some of them did not fail in the first 6 months of the year and others failed evenly throughout the year. Thus, the suggested ML algorithms must be able to detect failures in the test set that has different failure pattern compared to the train set.

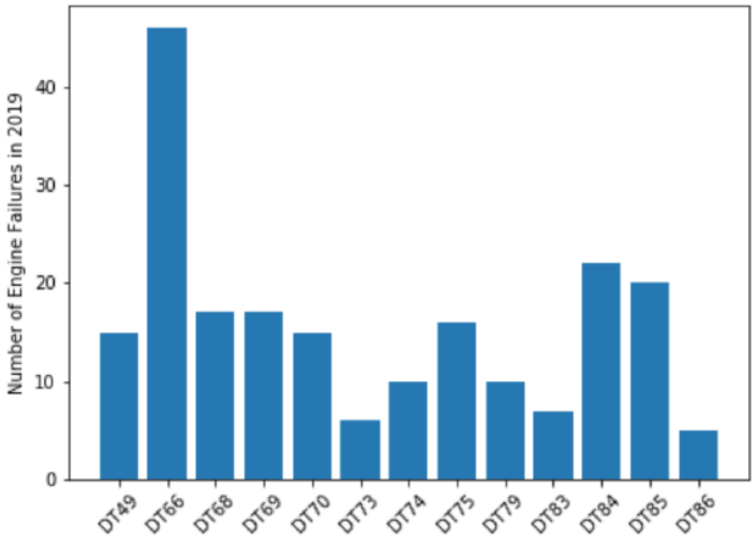


Figure 3 Number of engine failures for each truck in 2019

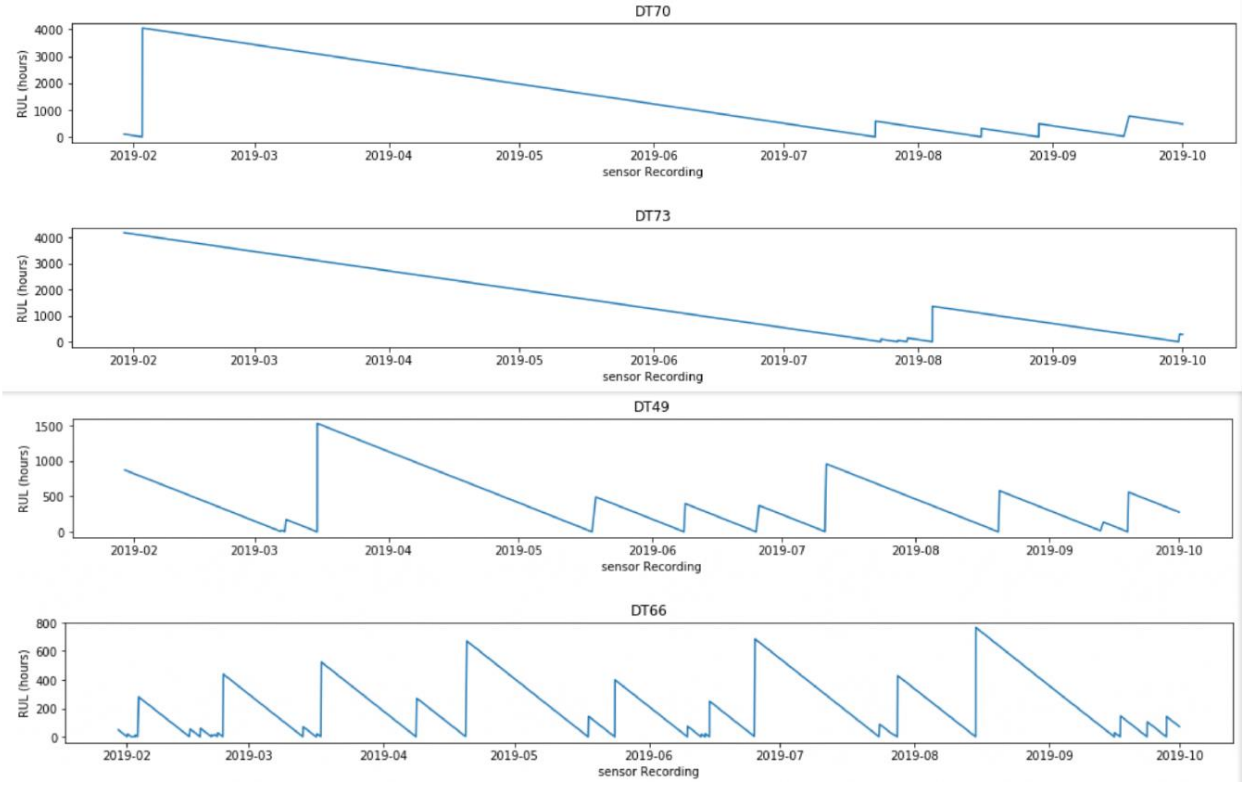


Figure 4 Sample RUL for engine failure

Data transformation

The following three data transformation methods are applied to the original table:

Firstly, In the original data the time intervals between sensor recordings are not equal which misleads ML algorithms. Many recordings that are very close to each other (5 minutes) will document similar results with low variance for many sensors. As a result, by interpolation method the time intervals between failures fixed to 30 minutes. Secondly, for each sensor, the

difference between two consecutive recordings is added as a separate column to the dataset. Lastly, In order to transform my skewed dataset to approximately normal, log transformation have been utilized.

Algorithm

Model preparation
For each truck the input to the algorithm is the recordings for 25 different sensors for 2019. The predicted output is a binary variable which is equal to 0 if there are less than 3 hours to the next failure and 1 otherwise. This binary output is confirmed with Kinross Gold representatives due to the following reasons:

- Company's main interest is to detect failures within 3 hours so that there are enough time available to replace the truck with new one and send the current truck for maintenance.
- Sensors are used to detect failures and they are expected to display normal recordings when the time remaining to the failure is more than 3 hours. Thus, the RUL which is the exact time remaining to the failure can not be predicted by sensor recordings with high precision.
- Binary classifiers are easier, faster to train and their results outperforms the prediction models for this case study

The aim of our model is to utilize sensor recordings in order to predict the upcoming failure closed to a pre-specified threshold (3 hours).

Performance matrix
Accuracy is the proportion of correctly predicted recordings to the total number of prediction and its equation is as follows:

$$Accuracy = \frac{T_P+T_N}{T_P+F_P+T_N+F_N}$$

Accuracy is the most common matrix used to evaluate the performance of the model. However, it is misleading when it comes to imbalanced data. Because, in case of imbalanced data many machine learning algorithms predict the majority class and this will result in good accuracy matrix even though non of the minority class was predicted correctly. Thus, we will focus on precision and recall for our model.

Precision indicates which proportion of the recordings that are predicted to be more than 3 hours to the failure were actually correct and its equation is as follows:

$$Precision = \frac{T_P}{T_P+F_P}$$

Recall indicates what proportion of the recordings that are more than 3 hours to the failure were actually correctly identified and its equation is as follows:

$$Recall = \frac{T_P}{T_P+F_N}$$

Sensor recordings for each truck is randomly split in training and test set such that 80% of the recordings are assigned to train set and 20% to test set. In addition, each truck is trained on the

recordings from other trucks. Moreover, since the objective is to correctly identify failures, four trucks without any failures in their test set have been excluded from analysis. For evaluation two consecutive sensor recordings will identify failure.

Results

As noted before the dataset is highly imbalanced and on average only 0.3% of the recordings are failure samples. To deal with imbalance Borderline SMOTE is applied to the train set. With this method only the samples near the borders are oversampled. By comparing different classification algorithms Naive Bays resulted in higher precision and recall value (table below displays the final results). In future, combination of sampling methods and feature engineering techniques to improve precision of my algorithm.

Table 2 Performance results for different machine learning algorithms for "DT82"

Truck Name	Accuracy	Average Precision	Average Recall
DT49	64	51	82
DT66	60	51	80
DT68	56	51	74
DT69	61	50	81
DT70	70	51	85
DT73	66	50	83
DT74	61	50	81
DT84	66	50	83
DT85	59	50	57
Average	52.5	50.4	78.4

An update on TTC subway infrastructure inspection projects

Janet Lam

Introduction and background

In this report, we provide an update on two projects with the TTC – the linetest inspection optimization project and the reinspection optimization project.

Linetest project

The TTC non-destructive testing (NDT) team is responsible for inspecting the subway track with ultrasound equipment, identifying small defects that may propagate into significant track failures. A linetest is an exploratory search from one station to the next, with the objective of discovering existing defects.

The NDT team is resourced to perform a linetest on the entire subway system once per year. However, there are certain parts of the track that are more susceptible to developing defects than other parts of the track.

The objective of our project with the TTC was to reallocate the NDT linetest resources to better identify defects by focussing on sections of the track that are more prone to defects, and borrowing resources from sections of track that are less likely to develop defects.

Summary of previous work

The last time we updated this project was in December 2018, where an optimal reallocation that takes into consideration the location and geometry of the track. The objective criterion was the total expected time between a defect occurring, and then being discovered via a linetest. Since some sections of track have fewer defects than others, re-allocating the annual linetests away from healthier sections of track and toward more problematic sections of track would reduce the unsupervised time of the defects.

We incorporated the different priorities of defects with different weights, as well as defects that occur on straight parts of track vs. mild curves of tight curves of track.

Extension of work

Since several years have passed, TTC requested an updated analysis with data from April 1 2016 to March 31 2021. All models remained the same, only the data changed. The recommended linetest schedule reduces the total expected unsupervised time by 30% over an equally distributed linetest schedule.

The resulting recommended linetest interval is given as follows. The relative frequency column indicates how much more often a certain section of the track should be linetested compared to the least frequent inspection. For example, the section from Sherbourne to Castlfrank should be inspection about 17 times more frequently than the section from St. Clair West to Dupont.

Line	Station	Relative Inspection Frequency	Annual Frequency
Bloor-Danforth	Sherbourne to Castlefrank	16.9	2.5
Bloor-Danforth	Dundas West to Lansdowne	16.5	2.4
YUS	North York Centre to Finch	15.9	2.3
YUS	Union to King	15.4	2.3
YUS	Lawrence to York Mills	14.2	2.1
Bloor-Danforth	Dufferin to Ossington	13.6	2
Bloor-Danforth	Donlands to Greenwood	13.2	1.9
YUS	York Mills to Sheppart	13.2	1.9
YUS	Eglinton to Lawrence	13.1	1.9
Bloor-Danforth	Victoria Park to Warden	13	1.9
YUS	Eglinton West to St. Clair West	12.3	1.8
Bloor-Danforth	Broadview to Chester	11.8	1.7
Bloor-Danforth	Woodbine to Main	11.3	1.7
Bloor-Danforth	Chester to Pape	10.9	1.6
YUS	St. George to Mueseum	10.9	1.6
YUS	Spadina to St. George, Spadina line	10.4	1.5
YUS	Sheppard to North York Centre	10.3	1.5
YUS	Wilson to Yorkdale	9.8	1.5
Bloor-Danforth	Castlefrank to Broadview	9.5	1.4
YUS	Yorkdale to Lawrence West	9.5	1.4
Bloor-Danforth	Bathurst to Spadina	9.1	1.3
Bloor-Danforth	Main to Victoria Park	9.1	1.3
Bloor-Danforth	Old Mill to Jane	9.1	1.3
YUS	St. Claire to Davisville	9	1.3
Scarborough SRT	Scarborough Centre to McCowan	8.9	1.3
YUS	Bloor to Rosedale	8.4	1.2
Bloor-Danforth	Keele to Dundas West	8.4	1.2
YUS	Wellesley to Bloor	8.3	1.2
Bloor-Danforth	Lansdowne to Dufferin	8.1	1.2
YUS	Lawrence West to Glencairn	7.7	1.1
YUS	Downsview to Wilson	7.6	1.1
YUS	St. Andrew to Union	7.6	1.1

YUS	Davisville to Eglinton	7.6	1.1
Bloor-Danforth	Christie to Bathurst	7.1	1
YUS	College to Wellesley	7	1
Bloor-Danforth	St. George to Bay	6.7	1
Scarborough SRT	Midtown to Scarobourgh Centre	6.7	1
Bloor-Danforth	Pape to Donlands	6.5	1
YUS	Osgoode to St. Andrew	6.5	1
Bloor-Danforth	Greenwood to Coxwell	6.5	1
YUS	King to Dundas	6.5	1
Bloor-Danforth	Coxwell to Woodbine	6.4	0.9
YUS	Finch Tail	6	0.9
Bloor-Danforth	Yonge to Sheppard	5.4	0.8
Bloor-Danforth	Warden to Kennedy	5.3	0.8
Bloor-Danforth	Royal York to Old Mill	5.2	0.8
Bloor-Danforth	Bayview to Yonge	5.2	0.8
Scarborough SRT	Kennedy to Lawrence East	5.1	0.8
Bloor-Danforth	Kipling to Islington	4.8	0.7
YUS	Glencairn to Eglinton West	4.8	0.7
Bloor-Danforth	Spadina to St. George, Bloor line	4.5	0.7
Bloor-Danforth	Islington to Royal York	4.3	0.6
YUS	Rosedale to Summerhill	4.2	0.6
Bloor-Danforth	High Park to Keele	4.1	0.6
YUS	Dupont to Spadina	4	0.6
YUS	St. Patrick to Osgoode	3.9	0.6
Sheppard	Yonge to Bayview	3.9	0.6
Bloor-Danforth	Ossington to Christie	3.8	0.6
Sheppard	Lesle to Don Mills	3.6	0.5
Scarborough SRT	Lawrence East to Ellesmere	3.2	0.5
Bloor-Danforth	Kipling Tail	2.8	0.4
YUS	Dundas to College	2.8	0.4
YUS	Museum to Queens Park	2.4	0.3
YUS	Queens Park to St. Patric	2	0.3
Bloor-Danforth	Runneymede to High Park	1.4	0.2
YUS	St. Claire West to Dupont	1	0.1

This table was constructed using a combined priority-geometry multiplier with the following weights. For example, a red defect occurring on a tangent track was given the weight 5*10 = 50, and so on. If TTC wishes to revise theses weights according to their subject-matter knowledge, we have provided them the tools to do so.

Geometry	Weight
Tangent	5
Mild curve	7
Tight curve	10

Priority	Weight
Red	10
Yellow	5
Purple	1
Blue	0.5

Brown	0.25
Grey	0.125

Conclusion
We have been advised that the minimal linetest frequency as defined by external regulators is once per year. Since that is currently TTC's testing frequency, we cannot advise less frequent inspections on certain parts of track. Instead, for the sections that are listed as having fewer than one inspection per year, to perform the minimal linetest as required by the external governing body. Then, for the remaining sections, to prioritize based on the list given above.

Reinspection project

Another responsibility of the TTC's Non-destructive testing (NDT) team is to revisit known defects in the subway rail system according to a defined timetable. Depending on the severity, or priority of the defect, the schedule may be every 21 days, or annually, or something in between. When the defects are revisited, the NDT team notes the updated status, and this process is repeated until the defects are resolved.

However, it has been found that some defects remain open for far beyond the originally scheduled resolution date, requiring the NDT team to repeated re-inspect the same defect until it is finally resolved. This may result in some defects being revisited as many as 57 times, consuming the NDT team's valuable resources.

The objective of this project with the TTC was to explore the consequences of less frequent reinspection, validated with their own data.

Summary of previous work
In the previous report, we discussed the model of transitioning from one priority to the next within a given re-inspection period. The results can be seen in the Table 3.

The key takeaways from the previous work is that based on the average number of defects carried each year, a slight increase in the inspection interval can save a significant number of inspections each year. The resulting decrease in reliability is relatively very small. For example, by changing the purple re-inspection interval from 17.4 days to 40 days increases the risk of missing a transition from 0.5% to 1.2%, and yields a saving of 57% in inspection efforts.

In order to focus and update the results, TTC requested that only data from April 1 2016 to March 31 2021 was used.

Table 3 Summary results on transition probabilities

Priority	Re-inspection interval (days)	P(no transition)	Re-inspections / year	Savings (%)	Average # of defects per year	Expected saved inspections per year
Purple	17.4	0.995	21		129	
	21	0.994	17.4	17		464
	40	0.989	9	57		1548
	60	0.984	6.1	71		1922
Blue	36.42	0.991	10		33	
	45	0.989	8.1	19		62.7
	60	0.985	6.1	39		129
	80	0.981	4.6	54		178

Extension of work

With the new dataset, we have the benefit of a model that is already built, but the renewed costs of data cleansing. Resultantly, our most recent efforts have been focussed on preparing and cleansing the new dataset to work seamlessly with the existing dataset.

With the new data stitched up to the existing data, we were able to perform the analysis again, and also observe any changes over the years.

Results

The results of the analysis are as follows.

Transition rate for purple and blue defects:

Priority	Transition rate
Purple	0.00034
Blue	0.0013

Using these transition rates, we were able to illustrate the possibilities with longer reinspection periods. The first reinspection interval under each priority is the current interval.

Priority	Re-inspection interval (days)	P(no transition)	Transitions per 100 inspections	Re-inspections / year	Savings (%)
Purple	17.4	0.994	0.6	21	
	21	0.993	0.7	17.4	17
	40	0.986	1.4	9	57
	60	0.98	2	6.1	71
Blue	36.42	0.954	4.6	10	
	45	0.943	5.7	8.1	19
	60	0.925	7.5	6.1	39
	80	0.901	9.9	4.6	54

The key information is the column labelled “transitions per 100 inspections.” This column indicates the expected number of transitions that occur every 100 inspections. For example, at the current inspection interval of 17.4 days, after 100 inspections at that interval (4.8 years), we can expect to discover 0.6 transitions from purple to a higher priority.

Then, we can compare this baseline to a less-frequent inspection interval. For example, if the inspection interval was every 60 days, then after 100 inspections at that interval (16.4 years), we can expect to discover 2 transitions from purple to a higher priority. By increasing the inspection interval, we’re permitting 1.4 additional transitions to occur. Though this may sound alarming, note that practically speaking, this means that at the reduced inspection rate, 1.4 additional transition will occur every 12 years, with an accompanied saving of 71% in reinspection resources.

We can extend our interpretation with the remainder of the results as needed.

CEA Data Audit Project

Kimia Taghvaei Ganjali

Introduction

Canadian Electricity Association (CEA) is a national leading electricity association in Canada representing many electricity companies. CEA was founded in 1891 and consists of senior executives from its Corporate Utility Members. CEA provides value added projects and services to improve the safety, security, and sustainability of the Canadian electricity industry. These strategic projects are research outputs to improve the performance of utility members.

Every electricity company under CEA wings uses the guidelines provided by CEA to record their equipment operating and outage data along with timestamps. Every system can have different failure modes, hence different types of outages for power plants. The process in which these outages are reported impacts the validity of the reliability measures. A framework is needed to ensure the quality of reported data is not as dependent on organizational and behavioral factors.

The objective of this research thesis is to generate a framework for CEA and their utility members. This framework will help assess their maintenance records and build a consistency throughout all sites working with CEA. To assess the consistency of different CEA sites, the primary question for this research is whether different utility plants are coding their events correctly. Large data sets of all 229 CEA sites have been audited and analyzed to generate a model and perform statistical analysis. Results of the analysis were then compared to see if there are any similarities between the sites or if there are significant differences.

Methodology

The project was broken down to four phases: business understanding, data understanding, modelling, and evaluation. Each phase will be explained in this section as well as a summary of results.

Business understanding

In the early stages of the project, the focus was more on understanding the terminologies used by CEA and the guidelines they had for the sites. CEA manuals and reports were studied to acquire a greater background for the project. The important business terms for this project are listed below:

State code: the state reported for the unit as per the manual
- Available states: 11, 12, ..16
- Outage codes (unavailable states): 21, 22, ...25
- Only reporting changes in states (i.e. no consecutive state 11's)
- The most important states for this study are 24 (maintenance outage) and 25 (planned outage)

Duration: duration of outage in hours
Capacity: capacity of the unit
- Have eight capacity groups and units should be classed based on their group

Data Understanding
Once the objectives and CEA terms were understood, the project moved into a data analysis stage. The meaning behind certain attributes and conditions under which some entries were recorded was discussed with CEA to understand what is required for the analysis. A database was created to merge the large data files from different plants along with other reference tables needed for the analysis. The final dataset was then investigated to identify issues. Issues such as null values, duplicate entries, wrong data types were resolved. The result of the data prep stage was a clean dataset that was then split to create a separate file for every plant.

Modeling
In early iterations of the modelling phase, the focus of the model was on the operating codes but when discussed with the client, there was a need to further analyze the downtime codes (24, 25) and hence the adjustments were made.

The primary research question was whether plants were coding their events correctly and with the focus on state code 24-25, the secondary question was whether these two state codes were being applied consistently across the plants. Furthermore, CEA wanted to investigate if there were any plants that are not discriminating between these two state codes.

There are two main models built for this project namely the data model and stats model. The data model takes the clean data file for every plant as an input and outputs a list of downtime durations for every capacity range. The model looks at the units within each capacity group and creates a list of maintenance outage and planned outage durations for that specific capacity group. If certain plants do not have units within some capacity groups, the list is empty for those ranges. The list is then sorted to create a histogram of all the data points and pass it to the stats model. This data model is created to avoid the manual work for different plants.

The statistical model takes the list of downtime durations for every capacity group and outputs the name of best fitted distribution and its parameters. The model performs a statistical fit analysis by fitting the empirical distribution to the theoretical ones in the scipy stats package. The model compares the fitted distributions by calculating the parameters and error (SSE) and picks the one with the lowest discrepancy.

Initially, all the distributions in the scipy package were studied which resulted in an extensive runtime hence the list was shortened to twenty known distributions.

This model is an efficient method to see the resulting probability distributions for a range of sites. The results were then evaluated to determine if there are statistically significant differences

Evaluation
For each plant, state code, capacity group combinations, there are two scenarios:

1. Resulting distributions are very different.
2. Resulting distributions are similar with different parameters.

In the second case, all those similar distributions were grouped. Additionally, we do the statistical KS test for the same plant to determine if 24, 25 were miscoded.

The hypothesis of the test is that the two samples come from the same distribution. If the difference between the two samples is not significant enough to say that they have different distributions, we fail to reject the null hypothesis.

Miscoded Plants
State code 24 is maintenance outage and should not really take more than a week. Hence, any occasional duration greater than a week was considered an outlier and was discarded from the analysis.

In some cases, many 24 durations were high numbers (more than a week) which indicates that in those plants, state code 24 was miscoded. For all those plants, we performed the KS test to see if 24,25 were the same. The result is in the below figure. In three cases, the KS test yields a positive value indicating that state code 24 and 25 were used interchangeably while in other cases, the KS test was negative meaning that there were some differences between state code 24, 25. However, these plants were still not using state 24 accurately and most data points were high numbers.

Table 4 Miscoded plants

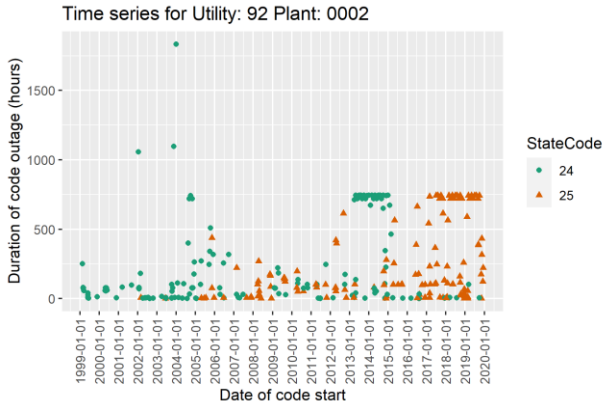
Hydro	Plant #	Note
92	2	pvalue=0.7155036428978084
70	3	KS test is negative but 24 is miscoded
90	6	KS test is negative but 24 is miscoded
15	100	KS test is negative but 24 is miscoded
41	100	KS test is negative but 24 is miscoded
41	120	KS test is negative but 24 is miscoded
41	140	KS test is negative but 24 is miscoded
41	160	KS test is negative but 24 is miscoded
60	208	KS test is negative but 24 is miscoded
50	250	KS test is negative but 24 is miscoded
50	270	KS test is negative but 24 is miscoded
14	310	pvalue=0.45050503595837244
14	320	pvalue=0.23868386968401767
90	416	KS test is negative but 24 is miscoded
60	606	KS test is negative but 24 is miscoded

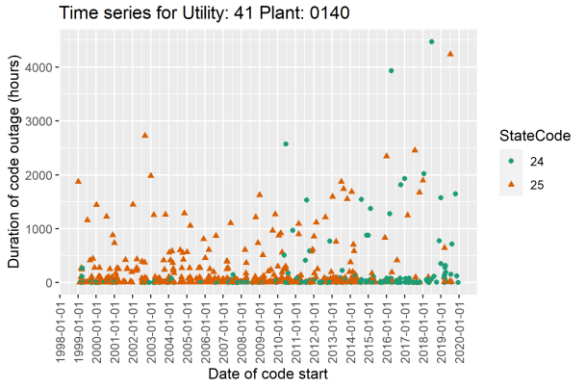
Time series
In this section, we are going to look at the miscoded plants (identified in Section 2.4.3) in more details and do a time series analysis. Broadly speaking, these graphs should

show that state 24's were used as 25's and there is no clear differentiation between the two. Some of these sites were behaving similarly and all the trends identified are discussed below.

In some cases, the historical data shows that there was a shift in the coding practices some time in the past. For example:

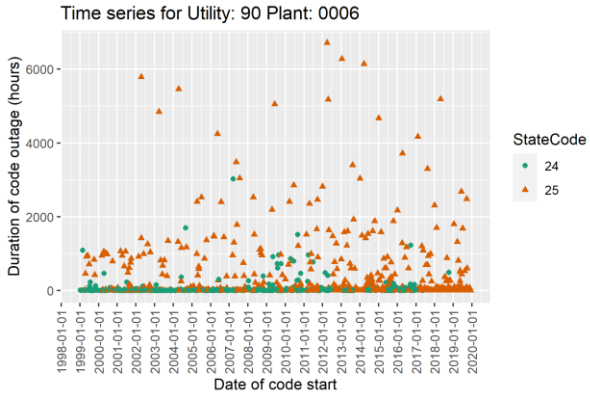
- utility 92, plant 2 in 2015 stopped using 24's and started coding everything as 25. In 2012-2015, they seemed to code a lot of long outages as 24's. 750 hours is about one month, and that is the ceiling you see for state 24
- utility 41, plant 140 changed something in 2014

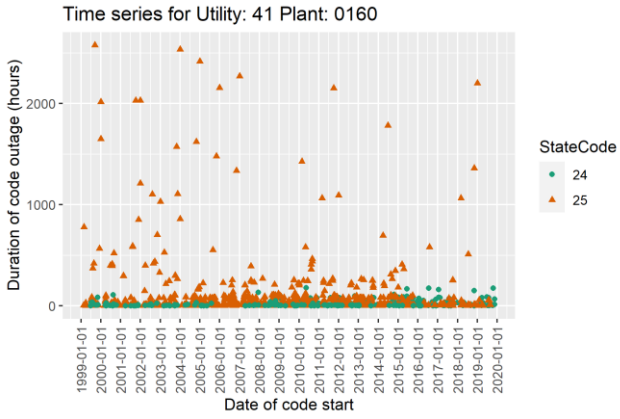




In some cases, they are correctly coding the long outages as 25's, but they are also coding a lot of short outages as 25's too.

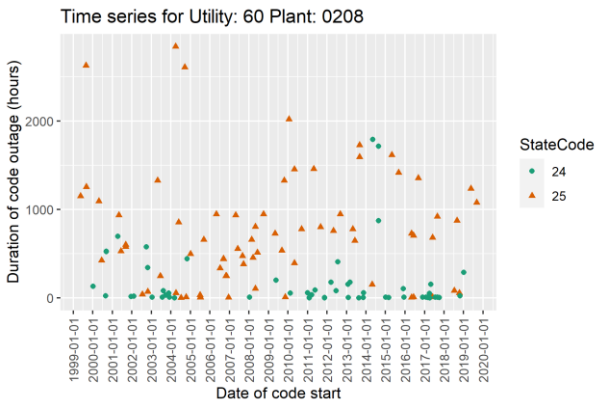
- utility 90, plant 6
- utility 41, plant 160

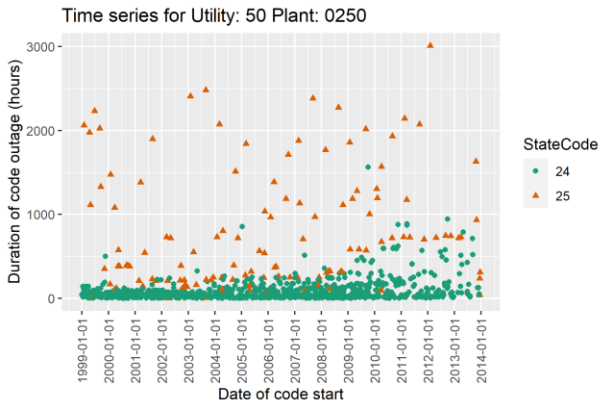




In some cases, they have the right idea and code their shorter outages as maintenance and longer ones as planned outage but perhaps the cut-offs for 24's is too high.

- utility 60 plant 208
- utility 50, plant 250





Recommendations and Future work

After finding the miscoded plants and groups of similar plants, CEA wanted to see some general recommendations for different hydro companies (utilities). A common error across many utilities was having occasional errors in their state code 24 for most plants. While these outliers were discarded from the analysis, it is important to reiterate the definition and limit of state 24 versus 25 to all the plants in these utilities.

After reporting the results for downtime analysis to CEA, there is value in investigating the operating codes (11-16) and understanding the difference between the available states and whether all sites are reporting those consistently. The same method could be used as well as developing other methods of analysis.

Table 5 Utility-level recommendations

Utility	Common errors	Recommendations
14	Most plants had some errors in their state 24's 2 identical 24,25's	Communicate the definition and limits of 24 versus 25
15	Data for 1 plant only, miscoded	Get updated data Do the analysis again for more plants
22	Most plants had some errors in their state 24's	Communicate the definition and limits of 24 versus 25
30	Not enough data for state 24,25	Get updated data Do the analysis again for more plants
41	Most plants had some errors in their state 24's	Communicate the definition and limits of 24 versus 25
43	Old data, did not include in the analysis	-
50	Overall good, only 2 miscoded	More analysis on those 2 plants

60	Most plants had some errors in their state 24's 2 miscoded	Communicate the definition and limits of 24 versus 25
70	Not enough data for state 24,25 1 miscoded	Get updated data Do the analysis again for more plants
80	Not enough data for state 24,25	Get updated data Do the analysis again for more plants
81	Not enough data for state 24,25	Get updated data Do the analysis again for more plants
82	Not enough data for state 24,25	Get updated data Do the analysis again for more plants
90	Most plants had some errors in their state 24's 2 miscoded	Communicate the definition and limits of 24 versus 25
91	Most plants had some errors in their state 24's 2 miscoded 1 identical 24,25's	Communicate the definition and limits of 24 versus 25

Capital Power Criticality Analysis Project Progress Report

Pooyan Sharifi

Project Overview

Capital Power is an independent power generation company who own and operate over 5100MW of power generation across 25 facilities in North America. Amongst these facilities there is a 875MW natural gas fueled power plant located in Brampton, Ontario. This project will involve a criticality analysis of the major systems and sub-systems in order to provide recommendations such as the criticality of components, how to decide which to keep. Additionally, a framework will be provided to help determine which spare parts to keep in inventory and the quantity.

Capital Power has provided an initial dataset containing a list of all work orders conducted on assets within each major system. These systems were defined by Capital Power based on their technical experience with the process equipment. The four major systems identified by Capital Power include:

1. Boiler Feedwater Pumps
2. Condensate Pumps
3. GCW & CCW Pumps
4. GT Hydraulic Life Pump

Boiler Feedwater Pumps

The Boiler Feedwater Pumps and Motors are part of the feedwater system to the Heat Recovery Steam Generator (HRSG) IP and HP steam drums for the production of steam.

The Feedwater Pumps provide feedwater at sufficient pressures and flow rates to maintain normal water level in the HP and IP drums from start-up to full load operation. Each HRSG (3 in total) is equipped with two 100% capacity Feedwater pumps. The Feedwater Pump is a barrel type, eight-stage centrifugal pump manufactured by Sulzer Pumps Inc.

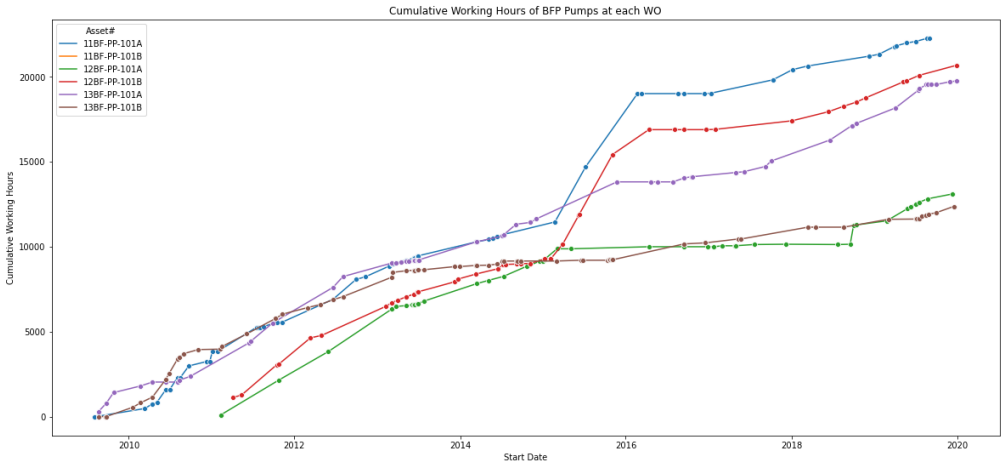


Figure 5: Cumulative Run Hours of Boiler Feedwater Pumps From 2009-2020

The operating run hours of each pump were determined via a status control bit extracted from Capital Power's SCADA system. This control bit (1 when pump is running, 0 when off) was extracted hourly from 2009-2020. Cumulative run hours were able to be computed which were later utilized for further data analysis. It is observed that pumps 11A, 12B and 13A are run the most.

Basic Concepts

A criticality analysis encompasses a process which assigns a criticality rating to assets based on their potential risks. From the initial dataset as provided by Capital Power the assets of the four major systems will be analyzed to identify major components / failure modes. With the given operational run time of said equipment the running hours of each equipment can be utilized to generate a time series of life cycles for each component for the asset. A 2-parameter censored Weibull distribution will be applied to the data in order to estimate important life characteristics for each of the components.

Results and Discussion

Boiler Feedwater Pumps

The table below summarizes the results of fitting a right-censored 2-parameter Weibull distribution. The shape parameter and scale parameter of these distributions are shown below.

Table 6: 2 - Parameter Censored Weibull Parameters for Boiler Feedwater Pumps

Component	Event Description	Shape Parameter β	Scale Parameter η	Comments
Lube Oil (Leaks)	Oil Leaks, low oil tank level, gasket leaking, lube oil pump leak, seal repair	1.289	8221	Wear
Lube Oil (Filters)	In-service high DP, high system back pressure, oil quality deterioration, etc.	1.033	4005	Random failures
Motor (Filters)	Air filter high DP alarm	0.797	3187	Early-life failures

Pump (Leaks)	In-board mechanical seal, o-rings	1.362	2317	Wear

Of the lube oil system there are two components or failure modes. The first being leaks which often occur as the gasket, seal or pump being inspected by an operator and noticing a leak. Additionally, a low oil tank level will also be classified as a failure due to oil leak. The corrective action undertaken is to repair these related equipment. The shape parameter for this failure mode is 1.289 which suggests the failure rate increases over time. This could be due to gradual wear of related parts such as the seal or gasket.

In addition to oil leaks another major component or failure mode is of the lube oil filters. These filters eventually build up debris over time which increase the differential pressure across the filter and reduce the flow rate of the lube oil. This reduction in flow could result in an increase in lube oil temperature which could result in faster lube oil degradation or tripping of equipment due to high lube oil temperature. The shape parameter was approximately equal to one which suggests a constant failure rate over time. This would coincide with the idea that filters typically get clogged in roughly the same number of operating hours over time.

The next major component of the boiler feedwater pumps is the motor which had two distinguishable failure modes. The first being the air filters of the motor which typically failed and required replacement when clogged resulting in a high differential pressure alarm. This component had a shape parameter of 0.797 which suggests it has early-life failures. The next component is of pump leaks which involve failure of the mechanical seal or O-rings on the pump. These had a shape parameter of 1.362 which suggest a failure rate that increases over time. This could be due to gradual wear of the seal and O-rings over time.

A condition-based maintenance policy for a two-unit system subject to dependent soft and hard failures: A reinforcement learning approach

Seyedvahid Najafi

Introduction

Manufacturing systems are composed of multiple components that interact with each other in different ways. The components wear out with age and usage, and lack of an appropriate asset management plan can incur considerable operation and maintenance (O&M) costs and lead to industrial accidents, and safety hazards. Designing an appropriate program for maintaining complicated systems is a common concern in academic and professional circles, increasing workplace safety and reducing O&M costs.

Condition-based maintenance (CBM), which prescribes actions when the actual state of a system is known, has recently received special attention to solving multi-unit maintenance optimization problems. Several CBM models have been designed proposing only the replacement of units, although repair actions are performed with different quality levels. When a unit fails, minimal repair can be performed, which returns it to the same state just before failure (as-bad). The other type of repair, which is referred to as general repair, improves a unit's health, but it does not return it to the as-new condition [1,2].

Yousefi [3] proposed a CBM policy for a multi-component system subject to deterioration and random shocks. The CBM proposes repair and replacement actions based on three degradation thresholds, where repair reduces the degradation by one level. The problem is formulated as a Markov decision process, and the Q-learning algorithm is applied to find the optimal policy. Jafari et al. [4] proposed a threshold-based CBM for a multi-unit system, in which perfect repair is allowed. Najafi and Makis [5] proposed an algorithm to find the optimal CBM for a two-unit system, where one unit is under condition monitoring, and age-based maintenance is applied on the other unit.

We propose an opportunistic CBM algorithm for a two-unit series system subject to soft and hard failures. The problem is formulated in the SMDP framework, and a reinforcement algorithm is applied to find the optimal policy and the expected long-run cost per unit time. Unlike traditional CBM approaches, no thresholds for maintenance intervention are considered; instead, the proposed policy maps condition monitoring data directly with the

repair/replacement actions. Using reinforcement learning enables us to consider both the age and deterioration of the units subject to condition monitoring. Moreover, in this study, general repair is allowed and can be performed at different quality.

System Description

The system consists of two major units under condition monitoring, and the failure of a unit leads to whole system failure, which is self-announcing. The state space of the system can be expressed as:

$$S = \{n_i, z_i, f_i | n_i \in \{0, \dots, \bar{N}_i\}, z_i \in \{0, \dots, \bar{Z}_i\}, f \in \{0,1\}, i \in \{0,1\}\} \quad (1)$$

where n_i, z_i and f_i represent the age, deterioration, and status of unit $i \in \{0,1\}$, respectively. N_i indicates the maximum useful age of unit i, $\bar{Z}_i$ is the maximum deterioration of unit i, and f represents the operational (0) or failure (1) status of the system.

Actions (a_1, a_2) are performed on unit 1 and unit 2, respectively, where $a_i \in \{0, \dots, Z_i\} \cup \{MR\}$ and actions consist of doing nothing (0), general repair ($a_i \in \{1, \dots, Z_i - 1\}$), minimal repair (MR), and replacement (Z_i).

The deterioration of each unit is described by the gamma process $\{X_t | t \geq 0\}$ whose probability distribution function is expressed as:

$$Ga(x \mid \alpha(t), \beta) = \frac{\beta^{\alpha(t)} x^{\alpha(t)-1} e^{-\beta x}}{\Gamma(\alpha(t))}, x \geq 0 \quad (2)$$

where the shape and scale parameters of the gamma process are $\alpha(t) > 0$ and $\beta > 0$, respectively, and the gamma function is:

$$\Gamma(\alpha(t)) = \int_0^\infty z^{\alpha(t)-1} e^{-z} dz \quad (3)$$

The hazard rate is widely applied for condition monitoring purposes. The proportional hazards model (PHM) estimates the risk of failure, integrating the effect of both age and deterioration. The general form of the PHM can be represented as follows [6]:

$$h(t, X_t) = h_0(t) \psi(X_t) \quad (4)$$

where $h_0(t)$ is the baseline function and depends on age only, and $\psi(X_t)$ is a function depending on environmental factors or health data referred to as covariates.

A Weibull PHM distribution is applied to estimate the hazard rate:

$$h(t, X_t) = \frac{k t^{k-1}}{\lambda^k} \cdot \exp(\theta X_t) \quad (5)$$

where λ is the scale parameter, k is the shape parameter, and θ is the regression coefficient vector of the covariates.

Sojourn Times

Different from soft failure that yields defective products, hard failure occurs randomly and stops the unit from operation. The remaining useful life of units depends on both age and deterioration and is estimated by [7]:

$$\hat{L}(t) = \hat{E}(T - t | T \geq t) \quad (4)$$

The remaining useful life can be estimated using the PH function [8]:

$$\hat{L}(t) \approx \int_0^{\infty} exp\left(-\int_t^{t+\tau} h(s|z_t)ds \right) d\tau \quad (7)$$

When no action is performed on each unit, the sojourn time is computed as $\nu_t = \min\{\widehat{L_\iota}(t)\}$, where $\widehat{L_\iota}$ is the remaining useful life of unit $i \in \{1,2\}$. If $\nu_t < \Delta$, a hard failure occurs, otherwise, $\nu_t = \Delta$, meaning that both units will survive over the inspection epoch. When a maintenance action is performed, sojourn time is computed using the following time parameters:

Notation	Definition
$T_{PGi}(\delta)$	preventive/opportunistic general repair time of unit i
T_{PRi}	preventive/opportunistic replacement time of unit i
T_{HRi}	corrective replacement time of unit i when hard failure occurs
T_{SRi}	corrective replacement time of unit i when soft failure occurs
$T_{HGi}(\delta)$	corrective general repair time of unit i when hard failure occurs
$T_{SGi}(\delta)$	corrective general repair time of unit i when soft failure occurs
T_{Mi}	minimal repair time of unit i

Table 1. time parameters notation and definition

where δ is the improvement made by general repair that reduces the unit deterioration from z to z'. Preventive general repair time is a function of the repair level, and the higher the improvement level, the longer the maintenance duration:

$$T_{PGi}(\delta) = \left(\frac{\delta}{Z_i}.j\right).T_{PRi} \quad (8)$$

To what extent repair time is proportional to the replacement time depends on the system structure and is captured by the repair time coefficient $0 < j \leq 1$. The other repair times are computed similarly, but corrective general repair times when soft and hard failure occur are proportional to T_{SRi} and T_{HRi}, respectively.

Reward structure

In this study, O&M costs are modeled as negative rewards, and the following cost parameters are considered:

Notation	Definition
$C_{PGi}(\delta)$	preventive/opportunistic general repair cost of unit i
C_{PRi}	preventive/opportunistic replacement cost of unit i
C_{HRi}	corrective replacement cost of unit i when hard failure occurs
C_{SRi}	corrective replacement cost of unit i when soft failure occurs
$C_{HGi}(\delta)$	corrective general repair cost of unit i when hard failure occurs
$C_{SGi}(\delta)$	corrective general repair cost of unit i when soft failure occurs

C_{Mi}	minimal repair cost of unit i
C_S	system setup cost
C_L	production loss per unit
C_I	inspection cost

Table 2. cost parameters notation and definition

When no action is taken, no cost is incurred; otherwise, the system costs depend on the action type. For example, the preventive general repair cost is calculated using the following equation:

$$C_{PGi}(\delta) = C_S + C_I + \left(\frac{\delta}{Z_i}.h\right).C_{PRi} + C_L.T_{PGi}(\delta) \quad (9)$$

Repair costs are a function of improvement level, and higher quality actions incur higher maintenance costs. However, repair costs are also influenced by other technical factors and repair complexity, which is incorporated by the repair cost coefficient $0 < h \leq 1$.

Proposed Policy

The system is inspected on an equidistant inspection interval of Δ, where the condition of each unit is monitored. If deterioration of unit i exceeds its soft failure threshold m_i, a maintenance action is performed to return the deterioration to somewhere below the threshold. Otherwise, a preventive action can be performed to replace or improve the system condition. When hard failure occurs, the failed unit is rectified immediately, and there is an opportunity to improve the other unit.

Reinforcement learning algorithm

This study considered the long-run average cost optimality criterion and used the SMART (Semi-Markov Average Reward Technique) algorithm to find the optimal policy [9]. The SMART algorithm learns the value of action a in particular state s under policy π, i.e., $R(s,a)$, which is derived from the Bellman optimality equation for average reward SMDPs [10]:

$$R_\pi(s,a) = r(s,a) - \rho_\pi.v(s,a) + \sum p_{s,s'}(a)\, max_{a'}\{R_\pi(s',a')\} \quad (5)$$

where $p_{s,s'}(a)$ is the transition probability from state s to s' when action a is chosen, and ρ_π is the average reward. The algorithm updates the utility of action values using temporal differences between the current and next state action-values. The above equation can be rewritten as follows if the expected reward r and the expected sojourn time v is replaced with their observed values:

$$R_\pi(s,a) \overset{\alpha_i}{\leftarrow} r_{imm}(s,a,s') - \rho_\pi.v_{imm}(s,a,s') + max_{a'}\{R_\pi(s',a')\} \quad (6)$$

Where $r_{imm}(s,a,s')$ and $v_{imm}(s,a,s')$ are the observed immediate reward and sojourn time resulted from taking action a in state s, and α_i is the learning rate at time i. The average reward is estimated as the ratio of cumulative immediate reward to the cumulative sojourn times:

$$\rho = \frac{\sum_{i=0}^{n} r_{imm}(s,a,s')}{\sum_{i=0}^{n} v_{imm}(s,a,s')} = \frac{R_{Total}}{T_{Total}} \quad (7)$$

The SMART algorithm is as follows, where α_i and ϵ_i are the learning and the exploration rates that decay according to a Darken-Chang-Moody search-then-converge procedure [11].

1. Set maximum time step I_{Max} and initialize action values $R_{new}(s,a) = 0$. Set the initial state (0,0,0), total reward $R_{Total} = 0$, total time $T_{Total} = 0$, and average reward $\rho = 0$.

2. While $i < I_{Max}$ do:

2.1. Choose a random action with probability ϵ_i; otherwise, execute action a that maximizes $R_i(s,a)$.

2.2. Execute action a and let the next state s', generate sojourn time $v_{imm}(s,a,s')$ and immediate reward $r_{imm}(s,a,s')$.

2.3. Obtain $R_{new}(s,a)$ following the update rule:

$$R_{new}(s,a) \overset{\alpha_i}{\leftarrow} (1-\alpha_i)R_{old}(s,a) + \alpha_i\{r_{imm}(s,a,s') - \rho.v_{imm}(s,a,s') + max_{a'}(R_{old}(s',a))\}$$

3. If a nonrandom action is selected in step 2.1:

Update total reward R_{Total} with $R_{Total} + r_{imm}(s,a,s')$

Update total time T_{Total} with $T_{Total} + v_{imm}(s,a,s')$

Update average reward ρ with $R_{Total}/\ T_{Total}$.

4. set current state s to the new state s' and update i with i+1.

Numerical example

In this section, system parameters, a numerical example, and alternative scenarios are presented. The system consists of two units in series that are both subject to condition monitoring. The lifetime of the units follows Weibull distributions with K=[30,25] B=[4.2,3.5], and $\boldsymbol{\theta}$ =[1,1.5], which represent the scale (months), shape and regression coefficient vectors, where the ith element describes unit i. Maximum age, deterioration levels, and soft failure thresholds for the units are 20 months, 8, and 5, respectively. Setup cost, inspection cost, and production loss are \$50, \$20, \$10, respectively, and maintenance costs are represented as follows:

Notation	Unit 1	Unit 2
C_{PRi}	\$4600	\$2500
C_{HRi}	\$5000	\$3000
C_{SRi}	\$4800	\$2700
C_{Mi}	\$400	\$250

Table 3. cost parameters

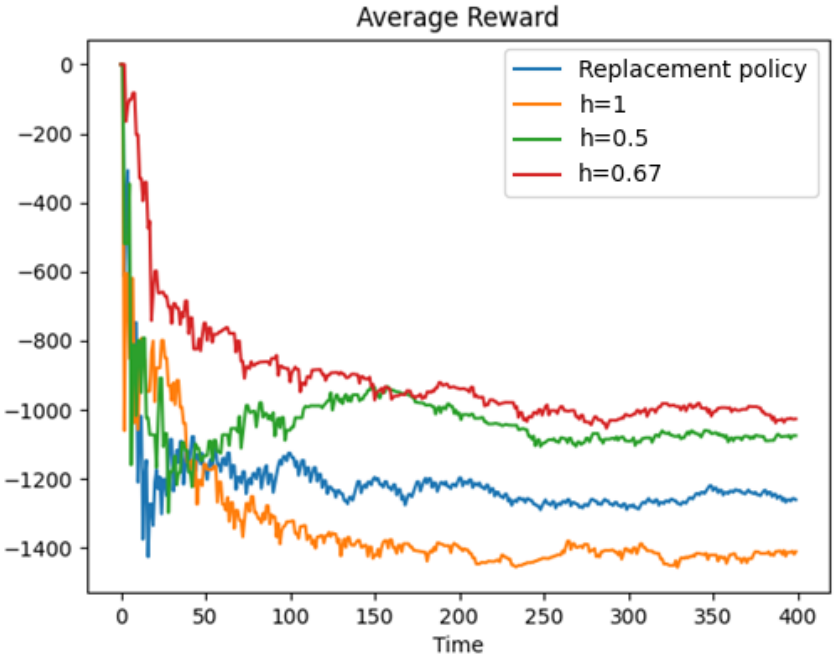
The time parameters are as follows:

Notation	Unit 1 (month)	Unit 2 (month)
T_{PRi}	0.023	0.019
T_{HRi}	0.033	0.029
T_{SRi}	0.026	0.021

T_{Mi}	0.018	0.011

Table 4. time parameters

Note that general repair times and costs are computed according to Sections 2.1 and 2.2, respectively. Given the abovementioned parameters, the proposed policy is compared with the replacement policy, in which no repair actions are allowed, and units can be replaced only. In this experiment, the repair time coefficient is $j = 1$, and three values for the repair cost coefficient (h=0.5, 0.67 and 1) are considered. The following graph shows the average reward under the replacement policy and the proposed policy when h takes 1, 0.5, and .67. The inspection interval is three months for all policies.



Graph 1. The long-run average cost of repair and replacement policies

As shown in the graph the repair policy does not necessarily results in a lower long-term average reward than the replacement policy. The replacement policy incurs the average cost of $1202 and outperforms the repair policy when $h = 1$ with the average cost of $1398. However, the repair policy converges to a lower average cost when h decreases to 0.67 and 0.5, resulting in the average cost of $925 and $995, respectively. The findings show that the repair policies with $h = 0.5$ and $h = 0.65$ accomplish 20% and 23% cost reductions, respectively, compared to the replacement policy.

Conclusion

In this paper, an opportunistic maintenance policy with general repair is proposed for a two-unit series system, in which the condition of both units is monitored. The maintenance problem is formulated in the SMDP framework, and the condition under which the proposed maintenance policy minimizes the long-run average cost per unit time is discussed. The policy prescribes preventive, corrective, and opportunistic actions consisting of minimal repair, general repair, and replacement. The novelty of this research is the inclusion of general repair in the maintenance modeling of two-unit systems in the SMDP framework, where repair actions

with different quality are performed without considering maintenance thresholds. A numerical example confirms the superiority of the proposed policy over a similar policy in which general repair is not considered.

To further the research, we plan to study systems with a higher number of units to enhance the practical value of the research, which requires finding an approach capable of managing a large state space. Furthermore, providing an analytical solution for the problem that shows the optimal policy will confirm the validity of the proposed approach.

References

[1] Kijima M, Morimura H, Suzuki Y. Periodical replacement problem without assuming minimal repair. Eur J Oper Res 1988;37:194–203.

[2] Pham H, Wang H. Imperfect maintenance. Eur J Oper Res 1996;94:425–38.

[3] Yousefi N, Tsianikas S, Coit DW. Reinforcement learning for dynamic condition-based maintenance of a system with individually repairable components. Qual Eng 2020;32:388–408.

[4] Jafari L, Naderkhani F, Makis V. Joint optimization of maintenance policy and inspection interval for a multi-unit series system using proportional hazards model. J Oper Res Soc 2018;69:36–48.

[5] Najafi S, Makis V. Comparison of Two Maintenance Policies for a Two-Unit Series System Considering General Repair. Int J Ind Manuf Eng 2020;14:558–63.

[6] Cox DR. Regression models and life-tables. J R Stat Soc Ser B 1972;34:187–202.

[7] Banjevic D, Jardine AKS. Calculation of reliability function and remaining useful life for a Markov failure time process. IMA J Manag Math 2006;17:115–30.

[8] Liao H, Zhao W, Guo H. Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model. RAMS'06. Annu. Reliab. Maintainab. Symp. 2006., IEEE; 2006, p. 127–32.

[9] Mahadevan S, Marchalleck N, Das TK, Gosavi A. Self-improving factory simulation using continuous-time average-reward reinforcement learning. Mach. Learn. Work. THEN Conf., MORGAN KAUFMANN PUBLISHERS, INC.; 1997, p. 202–10.

[10] Mahadevan S, Theocharous G. Optimizing Production Manufacturing Using Reinforcement Learning. FLAIRS Conf., vol. 372, 1998, p. 377.

[11] Darken C, Chang J, Moody J. Learning rate schedules for faster stochastic gradient search. Neural networks signal Process., vol. 2, Citeseer; 1992.

A novel GRU-driven Stochastic Degradation Process for Battery Forecasting

Zihan Zhang

In recent decades, there has been significant growth in the development of rechargeable battery-powered devices, such as electric vehicles, leading to a huge demand for batteries with high reliability and quality. End of life (EoL) is a critical indicator of battery health; it can be estimated by either adaptive stochastic processes or advanced machine learning techniques. However, these approaches either follow the degradation path having a specific form or lack stochastic interpretation due to its black-box nature. To address these challenges, an GRU-driven stochastic degradation process is proposed that can formulate battery degradation, in which drift fluctuation is controlled by a recursive Gaussian distribution with its mean learnt from an GRU-driven degradation pattern. Due to the non-Markovian state transitions, a sampling-based expectation maximization algorithm is developed to estimate model parameters based on historical observations.

Introduction

As Li-ion batteries have been widely used in electric vehicles (EVs) and hybrid EVs, the reliability of battery has become increasingly critical to the safe and reliable operation of a vehicle. This has led to increasing efforts of enhancing battery health management. [1] The prediction of state-of-health (SoH), a critical indicator of battery state, has been recognized as the central metric in battery health management and prognostic methods for SoH can be classified as model-based and data-driven forecasting methods.

Model-based forecasting methods include electrochemical model, equivalent circuit model, empirical model and stochastic model. [2-3] The former two models attempt to utilize the physical or chemical dynamics of the battery properties by exploring complex aging mechanisms. [4] However, the high cost of measuring the internal degradation parameters makes their practical application prohibitively expensive. [5] In contrast, empirical models may be useful, because the mathematical forms of the degradation trajectory can be mined from historical battery degradation processes, such as linear model [6] and exponential model [7]. Similarly, stochastic models can capture the degradation uncertainty based on nonlinear Wiener processes with specific drift terms, such as linear [8] and exponential [9]. Furthermore, advanced filtering techniques, such as Kalman filter [10] and extended Kalman filter (EKF) [11], can help either empirical or stochastic models reduce impacts of measurement noise and

external disturbance to further increase forecasting accuracy. However, without prior knowledge of the underlying degradation mechanisms and limited by specific expressions of degradation path (i.e., mathematical form in empirical models and drift terms in stochastic models), it is difficult to generalize them to batteries of diverse types.

Data-driven forecasting methods rely only on historical measurements to establish a degradation model, without prior knowledge of inherent degradation behavior. Optimization methods for data-driven models integrating kernel techniques are popular for improving the prediction efficiency, such as support vector machine (SVM) [12], relevant vector machine (RVM) [13] and Gaussian process regression (GPR) [14]. But it is difficult to generalize them due to their sensitivity to the kernel parameters [15]. To overcome this drawback, neural network-based methods have been implemented, such as autoregressive model [16], and Elman neural network [17]. But their prediction power is limited to short-data windows. Recurrent neural networks (RNNs) can model long-range dependencies [18], but it is hard to implement back propagation in practice because the squashing nonlinear activation function causes the prediction power of the model exponentially decaying over time. Besides, gated recurrent unit (GRU) achieves more efficient operation for real-time prediction with less training parameters than other RNNs, such as LSTM (Long Short-term Memory). [19] Although GRU and LSTM exhibit better performance in learning long-term dependence than the traditional RNN, they cannot describe stochastic degradation characteristics.

To close the gap, a GRU-driven degradation process is proposed for tracking battery degradation along with its increasing age. Specifically, the hidden state of the GRU neural network is introduced into the degradation process as a pattern term for capturing long dependencies among degradation rates. Then, a mapping function maps the pattern term to a mean of the degradation rate space. Thus, degradation rate is updated based on the entire degradation history instead of only the last time step, leading to a non-Markovian transitions. To cope with the non-Markovian transition, we developed a joint-learning sampling-based algorithm to estimate model parameters under the framework of Expectation Maximization (EM) method. Specifically, Sequential Monte Carlo (SMC) embedded in E-step provides posterior samples of degradation rates to M-step. Finally, a forecasting algorithm is proposed to estimate battery EoL based on the generative process of predicted degradation trajectory.

The remainder of this paper is organized as follows. Section II presents the model formulation, where a novel GRU-driven degradation process is developed. Section III investigates the parameter estimation under the framework of Expectation Maximization method. Section IV explains how EoL can be estimated. Section V validates the proposed model using a case study of CALCE dataset. Finally, Section VI concludes the paper.

Pattern-driven Degradation Process

Let $\{S_t;\ t \geq 0\}$ be a discrete-cycle battery degradation process, where S_t represents the battery SoH at the charge and discharge cycle t and can be calculated by $S_t = C_t/C_0$, where C_t is the capacity at cycle t and C_0 is the initial capacity.

The evolution of the degradation process is controlled by degradation rate κ_t which represents the degradation speed. Additionally, the update of degradation rate is assumed to follow a recursive Gaussian distribution which is Markovian, i.e., $p(\kappa_t|\kappa_{t-1}) = \mathcal{N}(\kappa_t; \kappa_{t-1}, \xi^2)$. Note that ξ is the time-invariant standard deviation and the initial degradation state is κ_0, which are to be estimated.

We introduce a GRU-driven pattern term $\boldsymbol{Q}_t$ to model long dependency in degradation rate κ_t, that is

$$\boldsymbol{Q}_t = \mathbf{GRU}(\boldsymbol{Q}_{t-1}, \kappa_{t-1}; \boldsymbol{\theta}_G), \quad (1)$$

where $\boldsymbol{Q}_t$ characterizes the time-varying degradation pattern, and $\boldsymbol{\theta}_G$ is the vector of hyper parameters of GRU neural network. Because (1) incorporates the entire evolution information, the Markovian assumption of traditional update of degradation rate will be violated. To enable the degradation rate κ_t controlled by the degradation pattern $\boldsymbol{Q}_t$, we use a mapping function $h_\mu(\cdot)$ to project $\boldsymbol{Q}_t$ from pattern space to the mean of the Gaussian space, i.e.,

$$p(\kappa_t|\kappa_{1:t-1}) = \mathcal{N}\left(\kappa_t; h_\mu(\boldsymbol{Q}_t; \boldsymbol{\theta}_h), \xi^2\right), \quad (2)$$

where $\boldsymbol{\theta}_h$ is the parameter vector of mapping function. For notational convenience, $\boldsymbol{\theta}_G$ and $\boldsymbol{\theta}_h$ are omitted in expressions of (1) and (2) in the rest of the paper.

By combing (1) and the mapping function $h_\mu(\cdot)$ in a single GRU-driven (Gd) neural network, we develop our GRU-driven degradation process model as follows:

$$\kappa_t \sim \mathcal{N}\left(Gd_{(\boldsymbol{\theta}_G, \boldsymbol{\theta}_h)}(\boldsymbol{Q}_{t-1}, \kappa_{t-1})), \xi^2\right), \quad (3\text{-a})$$

$$S_t = S_{t-1} + \kappa_t + \int_{t-1}^{t} \varphi_\tau \, d\Lambda_\tau, \quad (3\text{-b})$$

where φ_t is the time-invariant diffusion parameter for battery degradation, Λ represents standard Brownian motion. Due to the stability of φ_t, it can be reduced to a constant φ.

Compared with traditional methods proposed by Zhai, et al. [20], the proposed model can not only describe the degradation dynamics due to the adaptive degradation rate in the degradation process but also facilitate updating based on the entire historical degradation information thanks to the strong dependence-learning ability of GRU.

Parameter Estimation

Due to non-Markovian state transition in the proposed model, a joint-learning sampling-based expectation maximization algorithm is proposed to estimate model parameters. The model parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_\kappa, \varphi)$, where $\boldsymbol{\theta}_\kappa = (\kappa_0, \xi, \boldsymbol{\theta}_G, \boldsymbol{\theta}_h)$, can be learnt by maximizing the log-likelihood of $\boldsymbol{\Theta}$, i.e., $\boldsymbol{\Theta}^* = \text{argmax}_{\boldsymbol{\Theta}} \ell(\boldsymbol{\Theta})$, under EM framework.

Sampling based E-step:

The variational lower bound of the log probability $\ell(\boldsymbol{\Theta})$, $L_{\boldsymbol{\Theta}}$, can be expressed as $L_{\boldsymbol{\Theta}} = E_{\kappa_{1:T} \sim q(\kappa_{1:T})}\left\{\log\left[\frac{p(S_{1:T}, \kappa_{1:T}; \boldsymbol{\Theta})}{q(\kappa_{1:T})}\right]\right\}$. Instead of maximizing $\ell(\boldsymbol{\Theta})$, we can just maximize $L_{\boldsymbol{\Theta}}$. Furthermore, optimal $\boldsymbol{\Theta}^*$ can be obtained by $\boldsymbol{\Theta}^* = \text{argmax}_{\boldsymbol{\Theta}} E_{\kappa_{1:T} \sim q(\kappa_{1:T})}[\log p(S_{1:T}, \kappa_{1:T}; \boldsymbol{\Theta})]$.

Because the state transition is non-Markovian, the closed form of optimal $\boldsymbol{\Theta}^*$ cannot be found by following the conventional EM algorithm. Thus, $\kappa_{1:T}$ need to be sampled from the posterior $p\left(\kappa_{1:T}|S_{1:T}, \boldsymbol{\Theta}^{old}\right)$.

Using the variational lower bound, posterior samples $\kappa_{1:T}$ can be obtained from $\kappa_{1:T} \sim q(\kappa_{1:T})$ rather than the original intractable posterior $p\left(\kappa_{1:T}|S_{1:T}, \boldsymbol{\Theta}^{old}\right)$. Here, SMC is adopted to implement the sampling. Specifically, $p(S_{1:t}, \kappa_{1:t})$ is the unnormalized target distribution at cycle

t. For notational convenience, $\boldsymbol{\Theta}$ is omitted in the remaining E-step., Then posterior samples $\kappa_{1:T}$ can be given by Algorithm 1.

Algorithm 1: Posterior sampling based on SMC algorithm.

Initialization: Set $\kappa_0^n = \kappa_0$ and $\omega_0^n = 1/N$ for $n = 1, \dots, N$.

1 For $t = 1, \dots, T$ do

2 Draw ancestor $a_{t-1}^n \sim \mathcal{F}(\cdot|\boldsymbol{W}_{t-1})$ for $n = 1, \dots, N$, where $\mathcal{F}(\cdot|\boldsymbol{W}_{t-1})$ is the discrete probability distribution;

3 Draw rate $\kappa_t^n \sim p(\cdot|\kappa_{1:t-1}^{a_{t-1}^n}, S_t)$ for $n = 1, \dots, N$, where $p(\kappa_t|\kappa_{1:t-1}, S_t) \propto \mathcal{N}(\kappa_t; \mu_{\kappa_t}, \sigma_{\kappa_t}^2)$, $\sigma_{\kappa_t}^2 = (\varphi^{-2} + \xi^{-2})^{-1}$ and $\mu_{\kappa_t} = \sigma_{\kappa_t}^2[\varphi^{-2}(S_t - S_{t-1}) + \xi^{-2}h_\mu(\boldsymbol{Q}_t)]$

4 Set $\kappa_{1:t}^n = (\kappa_{1:t-1}^{a_{t-1}^n}, \kappa_t^n)$ for $n = 1, \dots, N$;

5 Compute the weight ω_t^n for $n = 1, \dots, N$, where $p(S_t|\kappa_{1:t-1}) \propto \mathcal{N}(S_t; \mu_{S_t}, \sigma_{S_t}^2)$, $\mu_{S_t} = h_\mu(\boldsymbol{Q}_t) + S_{t-1}$ and $\sigma_{S_t}^2 = \varphi^2 + \xi^2$

6 Compute the normalized the weight W_t^n for $n = 1, \dots, N$;

7 End for

8 Sample $\chi \sim \mathcal{F}(\cdot|\boldsymbol{W}_T)$, and trace backward to obtain the particle path $\kappa_{1:T}^\chi$.

Joint-learning M-step:

Based on the previous step, $\boldsymbol{\Theta}^*$ can be given by $\boldsymbol{\Theta}^* = \text{argmax}_{\boldsymbol{\Theta}} \left[\log p(\kappa_{1:T}^\chi) + \sum_{t=1}^T \log p(S_t|\kappa_t^{B_t^\chi})\right]$. Due to independence between $\boldsymbol{\Theta}_\kappa$ and φ, parameter estimation for $\boldsymbol{\Theta}$ can be divided into two subproblems: $\boldsymbol{\Theta}_\kappa^* = \text{argmax}_{\boldsymbol{\Theta}_\kappa} \log p(\kappa_{1:T}^\chi; \boldsymbol{\Theta}_\kappa)$, and $\varphi^* = \text{argmax}_\varphi \sum_{t=1}^T \log p(S_t|\kappa_t^{B_t^\chi}; \varphi)$, which represents the maximum likelihood estimations of a GRU network and (3-b), respectively.

Based on the statistical properties of Brownian motion, (3-a) can be rewritten as a Gaussian distribution. Thus, $\boldsymbol{\Theta}_\kappa$ and φ can be jointly estimated by adopting the negative log-likelihood of (3-a) to be the loss function of the neural network. Specifically, the expectation of (3-a) (if ignoring the constant term S_{t-1}) can be expressed as $E(S_t) = Gd_{(\boldsymbol{\theta}_G, \boldsymbol{\theta}_h)}(\boldsymbol{Q}_{t-1}, \kappa_{t-1}))$. Furthermore, φ and ξ can be estimated by an extended Gd neural network. By Bayesian theorem, the variance of S_t can be expressed as $\text{Var}(S_t) = \xi^2 + \varphi^2$. When constant terms are ignored, the log-likelihood of the Gaussian distribution can be formulated as $L \propto \frac{1}{2}(\ln(\xi^2 + \varphi^2) + (\xi^2 + \varphi^2)^{-1}(S_t - Gd_{(\boldsymbol{\theta}_G, \boldsymbol{\theta}_h)}(\boldsymbol{Q}_{t-1}, \kappa_{t-1}))^2)$. Thus, parameter estimation problem is transformed into a neural network training problem as shown in Fig. 1.

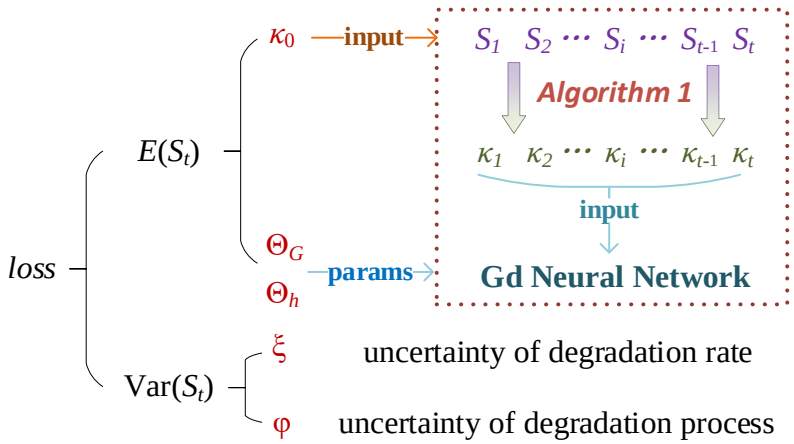


Fig. 1. Parameter estimation of the proposed model.

EoL Forecasting

After obtaining all parameters, the proposed model can be used for battery degradation forecasting. Specifically, battery EoL refers to the number of cycles required for the battery capacity to fall below a specific malfunction threshold (usually 70-80%, 80% is used in this work) under a certain charge and discharge setting. Obviously, the initial SoH is always 1 based on the definition of SoH. Then battery EoL can be predicted based on the generated SoH series following Algorithm 2.

Algorithm 2: Battery EoL forecasting.

For $i = 1, \ldots, I$ do

While $S_{T+i} > \zeta$ do

  Perform pattern transition: (1);

  Predict degradation rate: (2);

  Predict SoH S_{t+i}: (3-b);

  $k = k + 1$;

End while

Calculate $EoL_i = T + k$.

Based on parameters learned from historical SoH series, the proposed model can be formulated and EoL can be predicted correspondingly.

References

[1] Mehmet Ugras Cuma, Tahsin Koroglu. A comprehensive review on estimation strategies used in hybrid and battery electric vehicles. Renewable and Sustainable Energy Reviews. 2015, 42: 517-531.

[2] Bhaskar Saha1, Kai Goebel, Jon Christophersen. Comparison of prognostic algorithms for estimating remaining useful life of batteries. 2009, 31 (3/4): 293-308.

[3] Changfu Zou, Lei Zhang, Xiaosong Hu, Zhenpo Wang, Torsten Wik, Michael Pecht. A review of fractional-order techniques applied to lithium-ion batteries, lead-acid batteries, and supercapacitors. 2018, 390: 286-296.

[4] Chao Lyu, Qingzhi Lai, Tengfei Ge, Honghai Yu, Lixin Wang, Na Ma. A lead-acid battery's remaining useful life prediction by using electrochemical model in the Particle Filtering framework. 2017, 120: 975-984.

[5] Jingliang Zhang, Jay Lee. A review on prognostics and health monitoring of Li-ion battery. 2011, 196: 6007-6014.

[6] Chao Hu, Hui Ye, Gaurav Jain, Craig Schmidt. Remaining useful life assessment of lithium-ion batteries in implantable medical devices. Journal of Power Sources. 2018, 375: 118-130.

[7] Lijun. Zhang, Zhongqiang Mu, Changyan Sun. Remaining useful life prediction for Lithium-ion batteries based on exponential model and particle filter. IEEE Access. 2018, 6: 17729-17740.

[8] Guangzhong Dong, Zonghai Chen, Jingwen Wei, Qiang Ling. Battery health prognosis using Brownian motion modeling and particle filtering. IEEE Transactions on Industrial Electronics. 2018, 65(11): 8646-8655.

[9] Dong Wang, Yang Zhao, Fangfang Yang, Kwok-Leung Tsui. Nonlinear-drifted Brownian motion with multiple hidden states for remaining useful life prediction of rechargeable batteries. Mechanical Systems and Signal Processing. 2017, 93: 531-544.

[10] Xiaosheng Si. An adaptive prognostic approach via nonlinear degradation modeling: application to battery data. IEEE Transactions on Industrial Electrics. 2015, 62(8): 5082-5096.

[11] Liming Deng, Yucheng Hsu, Hanxiong Li. An improved model for remaining useful life prediction on capacity degradation and regeneration of lithium-ion battery. Annual Conference of the Prognostics and Health Management Society. 2017.

[12] Meru A. Patil, Piyush Tagade, Krishnan S. Hariharan, Subramanya M. Kolake, Taewon Song, Taejung Yeo, Seokgwang Doo. A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation. Applied Energy. 2015, 159: 285-297.

[13] YongZhi Zhang, Rui Xiong, HongWen He, Michael Pecht. Validation and verification of a hybrid method for remaining useful life prediction of lithium-ion batteries. Journal of Cleaner Production. 2019, 212: 240-249.

[14] Gozde O. Sahinoglu, Milutin Pajovic, Zafer Sahinoglu, Yebin Wang, Philip V. Orlik, Toshihiro Wada. Battery state-of-charge estimation based on regular/recurrent Gaussian process regression. IEEE Transactions on Industrial Electronics. 2018, 65 (5): 4311-4321.

[15] Jianbo Yu. State of health prediction of lithium-ion batteries: Multiscale logic regression and Gaussian process regression ensemble. Reliability Engineering and System Safety. 2018, 174: 82-95.

[16] Yapeng Zhou, Miaohua Huang, Yupu Chen, Ye Tao. A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. Journal of Power Sources. 2016, 321: 1-10.

[17] Xiaoyu Li, Lei Zhang, Zhenpo Wang, Peng Dong. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. Journal of Energy Storage. 2019, 21: 510-518.

[18] Yongzhi Zhang, Rui Xiong, Hongewn He, Michael G. Pecht. Long short-term memory recurrent neural network for remaining useful life prediction of Lithium-ion batteries. IEEE Transactions on Vehicular Technology. 2018, 67(7): 5695-5705.

[19] Wenxian Duan, Chuanxue Song, Silun Peng, Feng Xiao, Yulong Shao, Shixin Song. An improved gated-recurrent unit network model for state-of-charge estimation of Lithium-ion battery. Energies. 2020, 13(23): 6366.

[20] Qingqing Zhai, Zhisheng Ye. RUL prediction of deteriorating products using an adaptive wiener process model. IEEE Transactions on Industrial Informatics. 2017, 13(6): 2911-2921.

Implementation of Hybrid Prognostic Framework – Adaptive Degradation Model with Deep Learning Trajectory

Dhavalkumar Patel

Objective

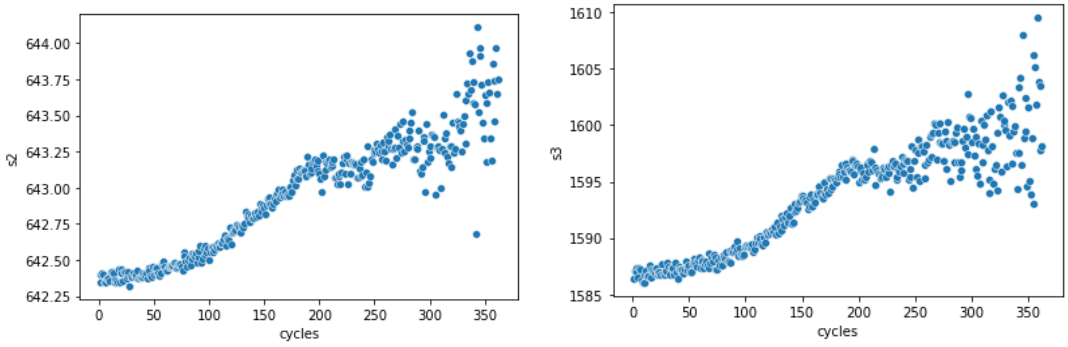
Estimation of the remaining useful life (RUL) and the degradation process plays an important role in the predictive maintenance of a component. The project focuses on the implementation of the hybrid prognostic framework as provided in the research paper; the developed program follows this framework.

- Dataset

NASA C-MAPSS turbofan engine degradation data is used for the project. The dataset comprises of four different engine sets: FD001, FD002, FD003, and FD004 with 100, 260, 100 and 260 engine units respectively. Each units have varying cycles and respective sensor readings.

- Initial Data Exploration

The presented plots are for FD001 dataset. The sensor readings of all the engine units in this dataset are averaged for each running cycle. The scatter plots can be used to visualize the trends in the sensor readings along with descriptive statistics. The sensor readings s2, s3, s4, s11, and s17, have similar increasing trends in the data which is evident in the below plots. These sensors will later be considered for Health Index construction. The sensor readings s8, s13, and s15 also follow a similar trend; however, they may not be considered as their spread of data is not significant.



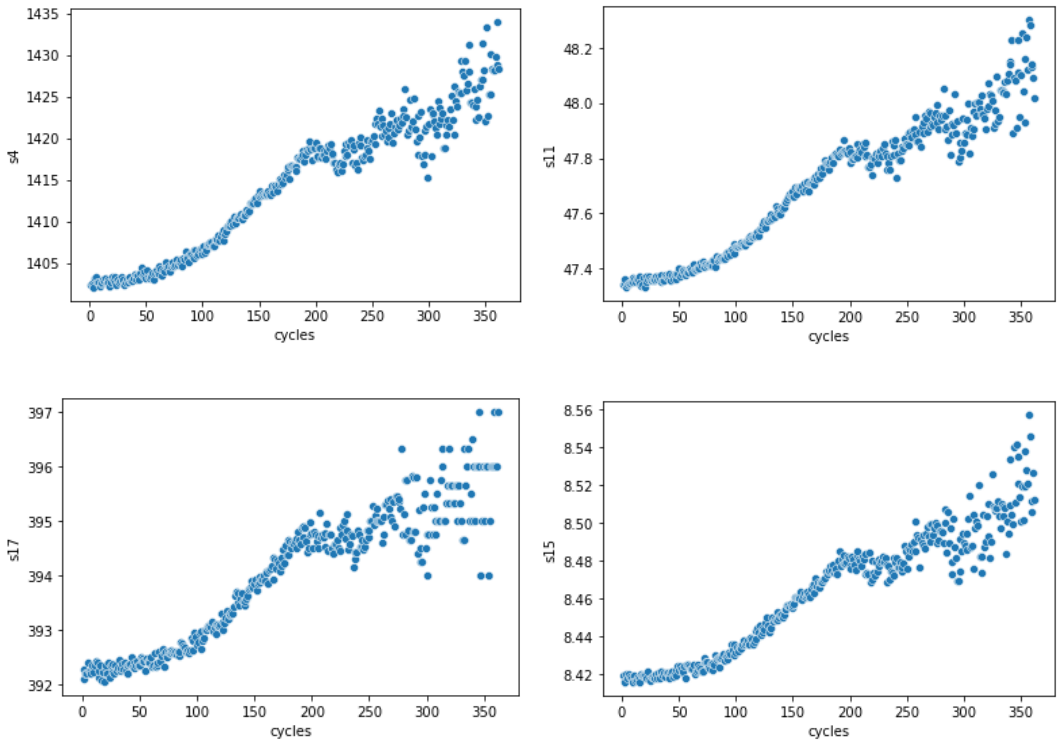
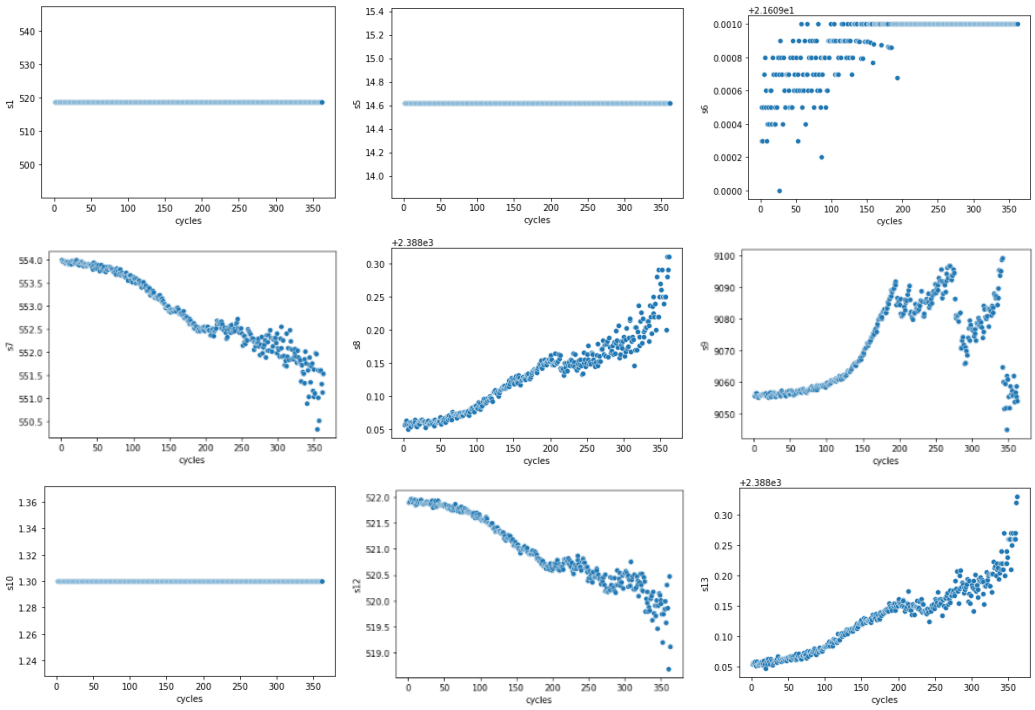


Fig 6: Trend plots for sensors': ('s2', 's3', 's4', 's11', 's17', 's15').

Below are the readings for other sensors which were not considered for health index calculation.



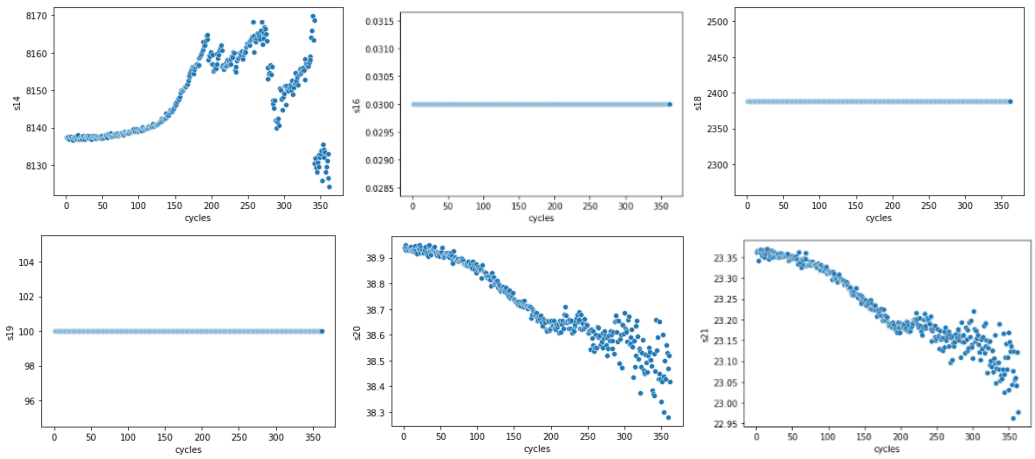


Fig 7: Trend plots for sensors which are not considered for Health Index Calculation

- Remaining Useful Life (RUL)

The remaining useful life is calculated based on the formula presented in the research paper. $N(t_i^j) = \frac{t_i^j - min_j(t_i^j)}{max_j(t_i^j) - min_j(t_i^j)}$, where t_i^j is the running cycles of the engine j at time i. The remaining useful life distributions of each engine set is shown in the following image.

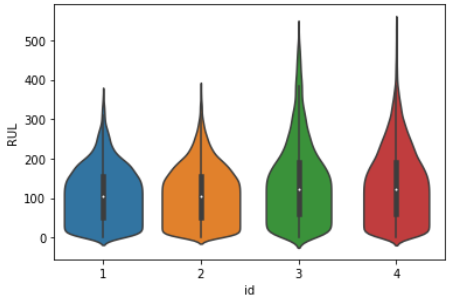


Fig 8: Violin plots for RUL distributions for each engine dataset (FD001, FD002, FD003, and FD004).

- Data Normalization

MinMaxScalar is employed to normalize the data. The range of the normalized data is [0, 2] for the sensor readings and the normalized cycles. The RUL is normalized in the range of [0, 1] as it is used as a label for Health Index calculation.

- Health Index (HI) Construction

The Health Index is constructed using stacking ensemble algorithm. Several algorithms including Extra Trees Regressor (etr), Random Forest Regressor (rfr), and Support Vector Regression (svr) were evaluated and compared with the stacking model. The architecture of the model includes three mentioned base models (level-0 models), and a level-1 model (Linear Regression) which combines the predictions of level-0 models. All the models were evaluated, and the stacking model performed slightly better than the other models.

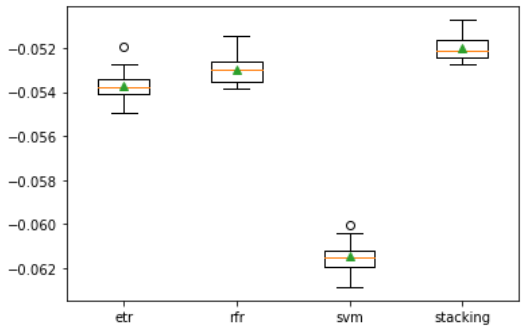


Fig 9: Box plots of Negative Mean Absolute Error for each model on FD002 engine set.

Health Index values are constructed for the complete dataset using the stacking model. The plots below compares the original normalized RUL values (presented in red) with constructed Health Index (presented in blue). Individual models were developed for each engine set.

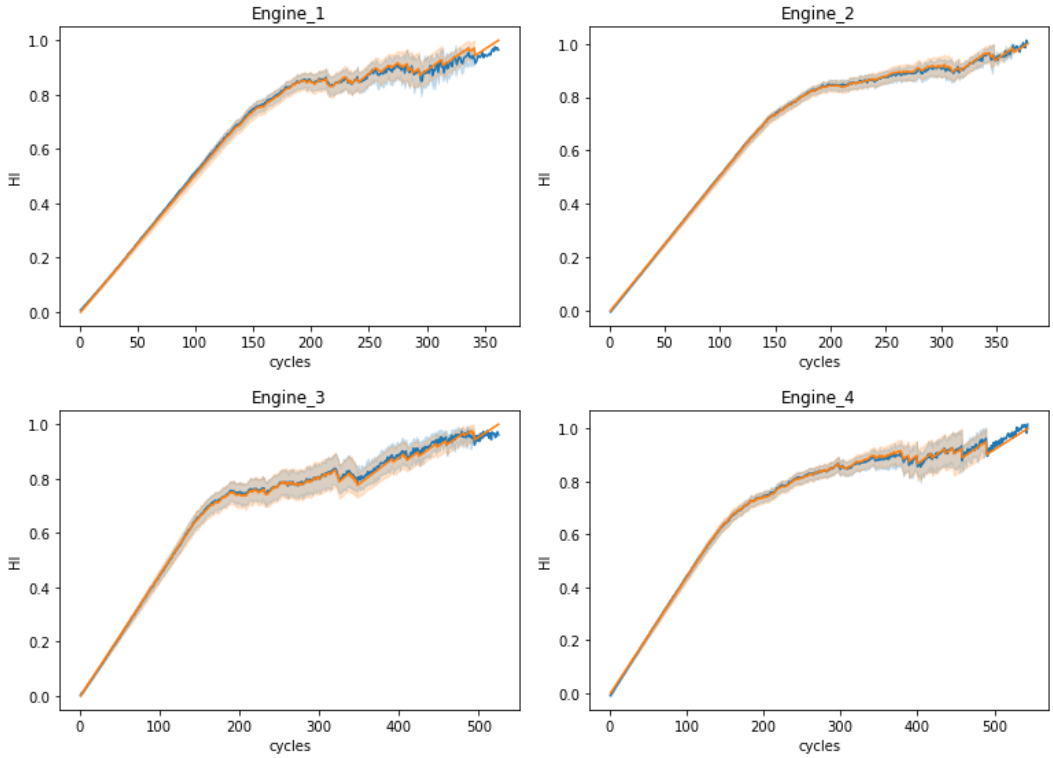


Fig 10: Normalized RUL values (in red) and Constructed Health Index (in blue)

- Model Architecture: (Encoder—LSTM, Decoder—CNN)
The model architecture is structured as presented in the paper.

LSTM Encoder:
Input: Historical Normalized Cycles and respective Historical Health Index. The model can handle data with variable input and output length.

Output: The last hidden state of the LSTM model is used as the encoded health vector.

CNN Decoder:

Input: The encoded signal is stacked multiple times (*output length) and the future normalized running cycles are appended to this stacked data.

Output: The CNN model outputs the future degradation signals.

The sample model architecture is presented below. The parameters can be adjusted according to our requirements.

```
encoder_decoder(
  (encoder): lstm_encoder(
    (lstm): LSTM(2, 20, batch_first=True)
  )
  (decoder): cnn_decoder(
    (cnn_layers): Sequential(
      (0): Conv1d(400, 400, kernel_size=(3,), stride=(1,), padding=(1,))
      (1): ReLU(inplace=True)
      (2): Conv1d(400, 400, kernel_size=(14,), stride=(1,), padding=(1,))
      (3): ReLU(inplace=True)
      (4): Conv1d(400, 400, kernel_size=(12,), stride=(1,), padding=(1,))
      (5): ReLU(inplace=True)
    )
  )
)
```

Fig 11: Description of Model Architecture

- Future Steps

1. Tune Model Hyperparameters
2. Define functions for calculation of evaluation metrics: RMSE, PCIP, and MPIW
3. Separate the DNN model for comparative studies.
4. Apply the developed program for new data.
5. Incorporate the adaptive wiener and wiener model for comparative studies.

Meta-free representation learning for few-shot learning

Kuilin Chen & Chi-Guhn Lee

Introduction

Currently, the vast majority of few-shot learning methods are within the general paradigm of meta-learning (a.k.a. learning to learn) [55, 4, 61], which learns multiple tasks in an episodic manner to distill transferrable knowledge. It is believed that meta-learning can learn unseen tasks with limited data more quickly and accurately using the transferrable knowledge, compared with training from scratch. Examples of transferrable knowledge could be good initialization of model parameters [57] or good embeddings of input data [16]. Meta-learning has been successfully applied to a number of domains, including few-shot classification [57, 44, 24], few-shot regression [16, 17, 36] and meta-reinforcement learning [69, 18]. In this article, we focus on few-shot regression and classification problems, and aim to tackle three problems in existing few-shot learning methods.

First, episodic meta-training methods are slow to converge, prone to over-fitting and tricky to implement [1]. Furthermore, recent studies [8, 11] cast doubt on whether it is the episodic meta-learning algorithm or the learned representation that is responsible for fast adaption to new tasks. It is found that the effectiveness of optimization-based meta-learning algorithms is due to reusing high-quality representation, instead of rapid learning of task-specific representation [50]. After thorough literature review, we do not find any theories to support the assumption that meta-learned representation generalizes better than the one from simple transfer learning, except some limited empirical case studies [22]. Although transfer learning and meta-learning are treated separately in the few-shot learning literature, we derive the link between them. Then, a new transfer learning method for few-shot learning is developed by finding representation that generalizes well on meta-test tasks. The proposed method is easy to implement and fast to converge because it does not require episodic meta-training.

Second, we address limitations in transfer learning methods for few-shot learning. Existing transfer learning approaches are restricted to few-shot classification problems [11, 14, 62, 37]. To the best of our knowledge, no transfer learning method is developed to achieve similar performance to meta-learning in few-shot regression. In the proposed MFRL, stochastic weight averaging (SWA) [31] improves the generalization capability of the representation for both regression and classification problems, because SWA is agnostic to loss function types.

Compared with most transfer learning approaches in few-shot learning that improve the quality of representation in an ad-hoc way, the proposed MFRL can be understood intuitively from the perspective of loss landscape geometry in modern deep neural networks.

Third, we calibrate the probabilistic output from the model learned upon few-shot samples in the meta-test phase. Few-shot learning models can be used in risk-averse applications such as medical diagnosis [46]. The diagnosis decision is made on not only point estimation but also probabilities associated with the prediction. The risk of making wrong decisions is significant when using uncalibrated models [3]. Despite a plethora of few-shot learning methods to improve the point estimation accuracy, few methods are developed to get probabilistic models with well calibrated uncertainty. Different from previous methods that integrate Bayesian learning into episodic meta-training [25, 17, 69, 58], our method only requires training an appropriate linear layer in the meta-test phase to get models with well calibrated uncertainty. Hierarchical Bayesian linear models are used to properly capture the uncertainty from very limited training samples in few-shot regression. In parallel, we scale the softmax output to make the few-shot classification model well calibrated.

Our contributions in this work are summarized as follows:

- We derive the relation between transfer learning and episodic meta-learning, and propose a transfer learning method that can handle both few-shot regression and classification problems with performance exceeding SOTA.
- We provide an explainable perspective to tackle few-shot learning problem based on geometry of loss landscape in modern deep neural networks.
- The proposed method results in well calibrated uncertainty in few-shot learning models, while preserving SOTA accuracy.
- Compared with existing few-shot learning methods, the proposed method is much easier to implement and more computationally efficient.

Results

We follow the standard setup in few-shot learning literature. The model is trained on a meta-training dataset and hyperparameters are selected based on the performance on a meta-validation dataset. The final performance of the model is evaluated on a meta-test dataset. The proposed method is applied to few-shot regression and classification problems, and compared against a wide range of alternative methods.

Few-shot regression

Sine waves [16] and head pose estimation [45] datasets are used to evaluate the performance of MFRL in few-shot regression.

The results for few-shot regression are summarized in the table for sine waves and head pose estimation. In the sine wave few-shot regression task, MFRL outperforms all meta-learning methods. It demonstrates that the high-quality representation can be learned in supervised learning, without episodic meta-learning. Although DKT with a spectral mixture (SM) kernel achieves very high accuracy, the good performance should be attributed to strong inductive bias to periodic functions in the SM kernel [66]. In the head pose estimation experiment, the proposed method also achieves much better accuracy than MAML, Bayesian MAML and DKT. In both few-shot regression problems, SWA results in improved accuracy, indicating that SWA can improve the quality of features and facilitate the learning of downstream tasks. In addition,

uncertainty is correctly estimated by the hierarchical Bayesian linear model with learned features using just 10 training samples.

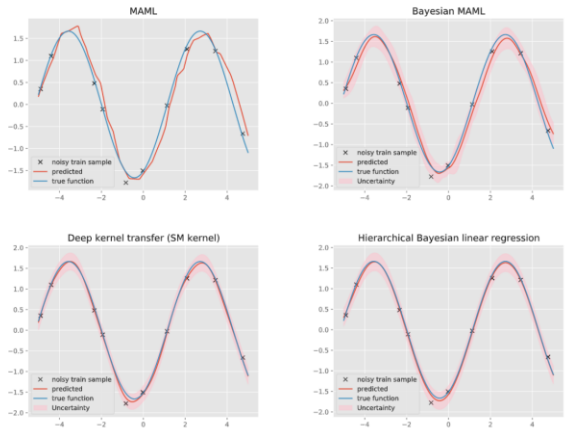


Figure 2: Qualitative comparison between different methods for sinusoidal function approximation and uncertainty quantification using 10 training samples. The standard deviation of data generation noise is 0.1, and the estimated noise standard deviation is 0.093 in MFRL.

Sine wave (2-layer MLP)	MSE
MAML [16]	0.67 ± 0.06
Bayesian MAML [69]	0.54 ± 0.05
ALPaCA [27]	0.14 ± 0.09
R2D2 [5]	$0.46 \pm$ NA
DKT + RBF [45]	1.38 ± 0.03
DKT + Spectral [45]	0.08 ± 0.06
MFRL	$\mathbf{0.016 \pm 0.008}$
Head pose (3-layer Conv Net)	MSE
MAML [16]	0.21 ± 0.01
Bayesian MAML [69]	0.18 ± 0.01
DKT + Spectral [45]	0.10 ± 0.01
MFRL	$\mathbf{0.027 \pm 0.005}$

Table 1: 10-shot regression on sine waves and head pose estimation.

Few-shot classification

We conduct few-shot classification experiments on four widely used few-shot image recognition benchmarks: miniImageNet [51], tieredImageNet [53], CIFAR-FS [5] and FC100 [43].

During testing, we conduct 5 independent runs of 600 randomly sampled few-shot classification tasks from test classes, and calculate the average accuracy. Each task contains 5 classes, 1 x 5 or 5 x 5 support samples and 75 query samples. A logistic regression model is learned only using the support samples. The classification accuracy is evaluated on the query samples. The results of the proposed method and previous SOTA methods using similar backbones are reported in tables. The proposed method achieves the best performance in most of the experiments, when compared with previous SOTA methods. Our method is closely related to Baseline++ [8] and fine-tuning on logits [11]. Baseline++ normalizes both classification weights and features, while the proposed method only normalizes features. It allows our method to find a more accurate model in a more flexible hypothesis space, given high-quality representation. Compared with fine-tuning on logits, our method obtains better results by learning a new logistic regression model on features, which store richer information about the data. Some approaches [54, 59] pretrain a C-class classification model on all training data, and then apply highly sophisticated meta-learning techniques to the pretrained model to achieve SOTA performance. Our approach with SWA outperforms those pretrained-then-meta-learned models, which demonstrates that SWA obtains high-quality representation that generalizes well to unseen tasks. Compared with improving representation quality for few-shot classification via self-distillation [62], the computational cost of SWA is significantly smaller because it does not require training models from scratch multiple times. Moreover, SWA can be applied to find good representation for both few-shot regression and classification, while previous transfer learning approaches can only handle few-shot classification problems [62, 37].

Table 2: Few-shot classification results on miniImageNet and tieredImageNet.

Method	Backbone	miniImageNet 5-way		tieredImageNet 5-way	
		1-shot	5-shot	1-shot	5-shot
Matching Net [64]	ResNet-12	63.08 ± 0.80	75.99 ± 0.60	68.50 ± 0.92	80.60 ± 0.71
Proto Net [57]	ResNet-12	60.37 ± 0.83	78.02 ± 0.57	65.65 ± 0.92	83.40 ± 0.65
MAML [16]	ResNet-12	56.58 ± 1.84	70.85 ± 0.91	-	-
AdaResNet [40]	ResNet-12	56.88 ± 0.62	71.94 ± 0.57	-	-
TADAM [43]	ResNet-12	58.50 ± 0.30	76.70 ± 0.30	-	-
Baseline++ [8]	ResNet-12	60.83 ± 0.81	77.81 ± 0.76	68.64 ± 0.86	80.47 ± 0.67
TapNet [71]	ResNet-12	61.65 ± 0.15	76.36 ± 0.10	63.08 ± 0.15	80.26 ± 0.12
Variational FSL [72]	ResNet-12	61.23 ± 0.23	77.69 ± 0.17	-	-
MetaOptNet [35]	ResNet-12	62.64 ± 0.61	78.63 ± 0.46	65.99 ± 0.72	81.56 ± 0.53
Ensemble [14]	ResNet-18	59.48 ± 0.65	75.62 ± 0.42	-	-
DSN [56]	ResNet-12	62.64 ± 0.66	78.83 ± 0.45	66.22 ± 0.75	82.79 ± 0.48
DKT [45]	ResNet-12	61.29 ± 0.57	76.25 ± 0.51	67.21 ± 0.52	79.69 ± 0.53
MFRL	ResNet-12	$\mathbf{64.11 \pm 0.82}$	$\mathbf{81.65 \pm 0.54}$	$\mathbf{71.43 \pm 0.81}$	$\mathbf{86.61 \pm 0.60}$
LEO [54]	WRN-28-10	61.76 ± 0.08	77.59 ± 0.12	66.33 ± 0.03	81.44 ± 0.09
Fine-tune [11]	WRN-28-10	57.73 ± 0.62	78.17 ± 0.49	66.58 ± 0.70	85.55 ± 0.48
Inductive SIB [30]	WRN-28-10	60.12 ± 0.56	78.17 ± 0.35	69.20 ± 0.58	84.96 ± 0.36
MetaFun [67]	WRN-28-10	$\mathbf{64.13 \pm 0.13}$	80.82 ± 0.17	67.27 ± 0.14	83.28 ± 0.12
MFRL	WRN-28-10	63.57 ± 0.81	$\mathbf{81.34 \pm 0.59}$	$\mathbf{71.38 \pm 0.86}$	$\mathbf{86.17 \pm 0.67}$

The proposed method not only achieves high accuracy in few-shot classification, but also makes the classification uncertainty well calibrated. Reliability diagrams can be used to check model calibration visually, which plot an identity function between prediction accuracy and confidence when the model is perfectly calibrated [9]. Figure shows the classification reliability diagrams along with widely used metrics for uncertainty calibration, including expected calibration error (ECE) [26], maximum calibration error (MCE) [41] and Brier score(BRI) [7]. ECE measures the average binned difference between confidence and accuracy, while MCE measures the maximum difference. BRI is the squared error between the predicted probabilities and one-hot labels. It is obvious that Baseline++ and DKT are significantly under-confident because cosine similarity is limited between -1 to 1 and flattens the peakedness of predictive probabilities. MAML is over-confident because tuning a deep neural network on few-shot data is prone to over-fitting. Meanwhile, Proto net and matching net are better calibrated than MAML because they do not fine-tune the entire network during testing. Nevertheless, they are still slightly over-confident. The results indicate that MFRL with a global temperature scaling factor can learn well calibrated models from very limited training samples.

Conclusions

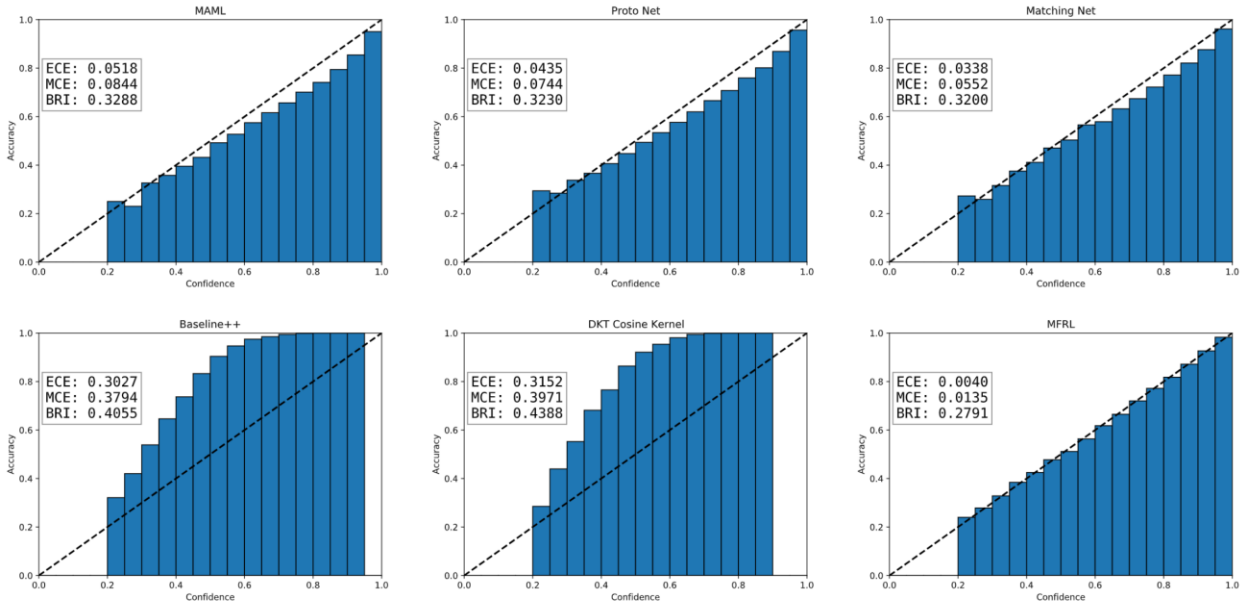


Figure 3: Reliability diagrams for 5-way 5-shot classification on miniImageNet

In this article, we derive the relation between transfer learning and episodic meta-learning. Based on the geometry of loss landscape of modern deep neural networks, a transfer learning approach is developed to learn representation that generalizes well to unseen tasks using SWA. The proposed method can be applied to both classification and regression tasks. Extensive experiments show that our method not only outperforms other SOTA methods on various datasets but also correctly quantifies the uncertainty in prediction.

References

[1] Antreas Antoniou, Harrison Edwards, and Amos Storkey. How to train your maml. In International Conference on Learning Representations, 2018.

[2] Ben Athiwaratkun, Marc Finzi, Pavel Izmailov, and Andrew Gordon Wilson. There are many consistent explanations of unlabeled data: Why you should average. In International Conference on Learning Representations, 2019.

[3] Edmon Begoli, Tanmoy Bhattacharya, and Dimitri Kusnezov. The need for uncertainty quantification in machine-assisted medical decision making. Nature Machine Intelligence, 1(1):20{23, 2019.

[4] Y Bengio, S Bengio, and J Cloutier. Learning a synaptic learning rule. In IJCNN-91-Seattle International Joint Conference on Neural Networks, volume 2, pages 969{vol. IEEE, 1991.

[5] Luca Bertinetto, Joao F Henriques, Philip Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In International Conference on Learning Representations, 2019.

[6] Glenn W Brier. Varification of forecasts expressed in terms of probability. Monthly weather review, 78(1):13, 1950.

[7] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In International Conference on Learning Representations, 2019.

[8] Morris H DeGroot and Stephen E Fienberg. The comparison and evaluation of forecasters. Journal of the Royal Statistical Society: Series D (The Statistician), 32(1-2):1222, 1983.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A largescale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248-255. Ieee, 2009.

[10] Guneet Singh Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A baseline for few-shot image classification. In International Conference on Learning Representations, 2020.

[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations, 2021.

[12] Jan Drugowitsch. Variational bayesian inference for linear and logistic regression. stat, 1050:16, 2014.

[13] Nikita Dvornik, Cordelia Schmid, and Julien Mairal. Diversity with cooperation: Ensemble methods for few-shot classification. In Proceedings of the IEEE International Conference on Computer Vision, pages 3723-3731, 2019.

[14] Nanyi Fei, Zhiwu Lu, Tao Xiang, and Songfang Huang. fMELRg: Meta-learning via modeling episode-level relationships for few-shot learning. In International Conference on Learning Representations, 2021.

[15] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, pages 1126{1135. JMLR. org, 2017.

[16] Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic model-agnostic meta-learning. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, pages 9537-9548, 2018.

[17] Sebastian Flennerhag, Andrei A Rusu, Razvan Pascanu, Francesco Visin, Hujun Yin, and Raia Hadsell. Meta-learning with warped gradient descent. In International Conference on Learning Representations, 2020.

[18] Victor Garcia and Joan Bruna Estrach. Few-shot learning with graph neural networks. In 6th International Conference on Learning Representations, ICLR 2018, 2018.

[19] Timur Garipov, Pavel Izmailov, Dmitrii Podoprikhin, Dmitry Vetrov, and Andrew Gordon Wilson. Loss surfaces, mode connectivity, and fast ensembling of dnns. Advances in Neural Information Processing Systems, 2018:8789{8798, 2018.

[20] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4367-4375, 2018.

[21] Micah Goldblum, Steven Reich, Liam Fowl, Renkun Ni, Valeriia Cherepanova, and Tom Goldstein. Unraveling meta-learning: Understanding feature representations for few-shot tasks. In Hal Daume III and Aarti Singh, editors, Proceedings of the 37th International Conference on Machine Learning, volume 119 of Proceedings of Machine Learning Research, pages 3607{3616, Virtual, 13-18 Jul 2020. PMLR.

[22] Shaogang Gong, Stephen McKenna, and John J Collins. An investigation into face pose distributions. In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pages 265{270. IEEE, 1996.

[23] J Gordon, J Bronskill, M Bauer, S Nowozin, and RE Turner. Meta-learning probabilistic inference for prediction. In International Conference on Learning Representations (ICLR 2019). OpenReview. net, 2019.

[24] Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, and Thomas Griths. Recasting gradient-based meta-learning as hierarchical bayes. In International Conference on Learning Representations, 2018.

[25] Chuan Guo, Geo Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In International Conference on Machine Learning, pages 1321-1330. PMLR, 2017.

[26] James Harrison, Apoorva Sharma, and Marco Pavone. Meta-learning priors for efficient online bayesian regression. In International Workshop on the Algorithmic Foundations of Robotics, pages 318{337. Springer, 2018.

[27] Haowei He, Gao Huang, and Yang Yuan. Asymmetric valleys: Beyond sharp and at local minima. In Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pages 2549-2560, 2019.

[28] Matthew D Homan and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. J. Mach. Learn. Res., 15(1):1593{1623, 2014.

[29] Shell Xu Hu, Pablo Garcia Moreno, Yang Xiao, Xi Shen, Guillaume Obozinski, Neil Lawrence, and Andreas Damianou. Empirical bayes transductive meta-learning with synthetic gradients. In International Conference on Learning Representations, 2020.

[30] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. In 34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018, pages 876{885. Association For Uncertainty in Artificial Intelligence (AUAI), 2018.

[31] Ghassen Jerfel, Erin Grant, Thomas L Griths, and Katherine Heller. Reconciling meta-learning and continual learning with online mixtures of tasks. Advances in Neural Information Processing Systems, 32, 2019.

[32] Jaekyeom Kim, Hyoungseok Kim, and Gunhee Kim. Model-agnostic boundary-adversarial sampling for test-time generalization in few-shot learning. In European conference on computer vision. Springer, 2020.

[33] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In ICML deep learning workshop, volume 2. Lille, 2015.

[34] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 10657-10665, 2019.

[35] Yoonho Lee and Seungjin Choi. Gradient-based meta-learning with learned layerwise metric and subspace. In International Conference on Machine Learning, pages 2927{2936, 2018.

[36] Puneet Mangla, Mayank Singh, Abhishek Sinha, Nupur Kumari, Vineeth N Balasubramanian, and Balaji Krishnamurthy. Charting the right manifold: Manifold mixup for few-shot learning. In 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 2207-2216. IEEE, 2020.

[37] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A simple neural attentive meta-learner. In International Conference on Learning Representations, 2018.

[38] Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte carlo radient estimation in machine learning. Journal of Machine Learning Research, 21(132):1-62, 2020.

[39] Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, and Adam Trischler. Rapid adaptation with conditionally shifted neurons. In International Conference on Machine Learning, pages 3664{3673. PMLR, 2018.

[40] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 29, 2015.

[41] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. arXiv preprint arXiv:1803.02999, 2018.

[42] Boris Oreshkin, Pau Rodrguez Lopez, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. In Advances in Neural Information Processing Systems, pages 721-731, 2018.

[43] Eunbyung Park and Junier B Oliva. Meta-curvature. Advances in Neural Information Processing Systems, 32:3314-3324, 2019.

[44] Massimiliano Patacchiola, Jack Turner, Elliot J Crowley, Michael O'Boyle, and Amos J Storkey. Bayesian meta-learning for the few-shot setting via deep kernels. Advances in Neural Information Processing Systems, 33, 2020.

[45] Viraj Prabhu, Anitha Kannan, Murali Ravuri, Manish Chaplain, David Sontag, and Xavier Amatriain. Few-shot learning for dermatological disease diagnosis. In Machine Learning for Healthcare Conference, pages 532-552. PMLR, 2019.

[46] Hang Qi, Matthew Brown, and David G Lowe. Low-shot learning with imprinted weights. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5822-5830, 2018.

[47] Limeng Qiao, Yemin Shi, Jia Li, Yaowei Wang, Tiejun Huang, and Yonghong Tian. Transductive episodic-wise adaptive metric for few-shot learning. In Proceedings of the IEEE International Conference on Computer Vision, pages 3603-3612, 2019.

[48] Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan L Yuille. Few-shot image recognition by predicting parameters from activations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7229-7238, 2018.

[49] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. In International Conference on Learning Representations, 2020.

[50] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net, 2017.

[51] Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Few-shot learning with embedded class models and shot-free meta training. In Proceedings of the IEEE International Conference on Computer Vision, pages 331-339, 2019.

[52] Mengye Ren, Eleni Triantallou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B Tenenbaum, Hugo Larochelle, and Richard S Zemel. Meta-learning for semi-supervised few-shot classification. In International Conference on Learning Representations, 2018.

[53] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. In International Conference on Learning Representations, 2019.

[54] Jurgen Schmidhuber. Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook. PhD thesis, Technische University at Munchen, 1987.

[55] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. Adaptive subspaces for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4136-4145, 2020.

[56] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In Advances in Neural Information Processing Systems, pages 4077-4087, 2017.

[57] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 403-412, 2019.

[58] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1199-1208, 2018.

[59] Sebastian Thrun and Lorien Pratt. Learning to learn. Springer Science & Business Media, 1998.

[60] Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B Tenenbaum, and Phillip Isola. Rethinking few-shot image classification: a good embedding is all you need? In European conference on computer vision. Springer, 2020.

[61] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. Journal of machine learning research, Jun:211-244, 2001.

[62] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In Advances in neural information processing systems, pages 3630-3638, 2016.

[63] Florian Wenzel, Kevin Roth, Bastiaan Veeling, Jakub Swiatkowski, Linh Tran, Stephan Mandt, Jasper Snoek, Tim Salimans, Rodolphe Jenatton, and Sebastian Nowozin. How good is the bayes posterior in deep neural networks really? In International Conference on Machine Learning, pages 10248{10259. PMLR, 2020.

[64] Andrew Wilson and Ryan Adams. Gaussian process kernels for pattern discovery and extrapolation. In International conference on machine learning, pages 1067-1075, 2013.

[65] Jin Xu, Jean-Francois Ton, Hyunjik Kim, Adam Kosiorek, and Yee Whye Teh. Metafun: Meta-learning with iterative functional updates. In International Conference on Machine Learning, pages 10617-10627. PMLR, 2020.

[66] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Few-shot learning via embedding adaptation with set-to-set functions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8808-8817, 2020.

[67] Jaesik Yoon, Taesup Kim, Ousmane Dia, Sungwoong Kim, Yoshua Bengio, and Sungjin Ahn. Bayesian model-agnostic meta-learning. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, pages 7343-7353, 2018.

[68] Sung Whan Yoon, Jun Seo, and Jaekyun Moon. Tapnet: Neural network augmented with task-adaptive projection for few-shot learning. In ICML 2019 (International Conference on Machine Learning). ICML, 2019.

[69] Sung Whan Yoon, Do-Yeon Kim, Jun Seo, and Jaekyun Moon. Xtarnet: Learning to extract task-adaptive representation for incremental few-shot learning. In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event, volume 119 of Proceedings of Machine Learning Research, pages 10852-10860. PMLR, 2020.

[70] Jian Zhang, Chenglong Zhao, Bingbing Ni, Minghao Xu, and Xiaokang Yang. Variational few-shot learning. In Proceedings of the IEEE International Conference on Computer Vision, pages 1685-1694, 2019.

[71] Manli Zhang, Jianhong Zhang, Zhiwu Lu, Tao Xiang, Mingyu Ding, and Songfang Huang. IEPT: Instance-level and episode-level pretext tasks for few-shot learning. In International Conference on Learning Representations, 2021.

Automatic Airport Xray Baggage Scanner via Adversarial Domain Adaptation

Jahyun (Lucrece) Shin

Background

The aim of this research is to build an automatic airport security baggage scanner. Currently, human operators at airports are inspecting if travellers' baggage contains any harmful objects. It would be both time and cost-effective to implement a "smart" scanner that can automatically perform this scanning process to detect any harmful objects.

Data

Xray Images

Incheon International Airport located in Incheon, South Korea provided us with Xray images of scanned baggages containing harmful objects, as shown in Figure 1. The provided Xray images were labeled with seven classes: gun, knife, hard disk, phone, battery, usb, and shuriken. During initial stages, we decided to consider only two classes, gun and knife.

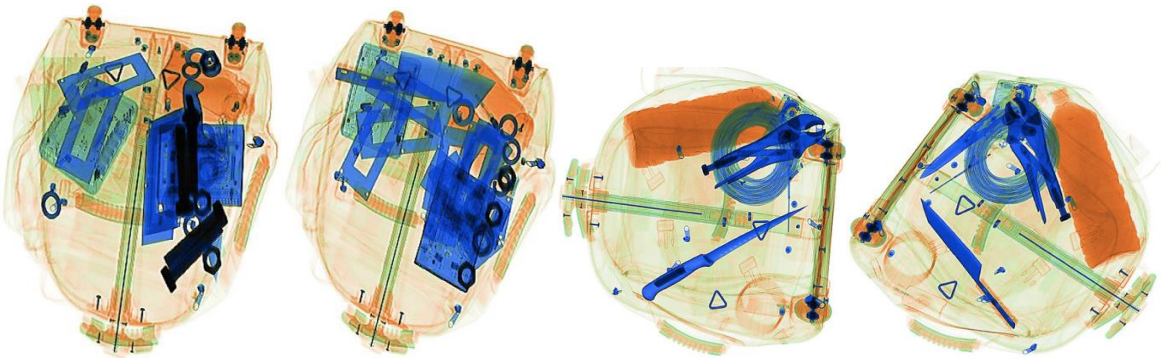
Figure 1. ***Examples of scanned Xray images provided by Incheon International Airport.***

Removal of Duplicates. The given Xray images contained many duplicates, in the form of different rotations of the same image. In order to calculate the number of unique images, all such duplicates were removed.

Removal of Whitespace. Most gun and knife Xray images came with a big white space that filled nearly three quarters of the entire image space. Since this is a waste of information for

such high-dimensional data, they were cropped tightly to only contain the useful content of the image.

Removal of Kitchen Knife Images. The given knife class is a mix of three different categories of knife: cutter knife, other knife, and kitchen knife. However, kitchen knife images are very clogged with a large overlap with other objects in the bag, as shown in Figure 2 (a) and (b). Although the final goal will be to detect the object in such clogged environments as well, for initial stages of model development, these images were considered as a source of noise, since the images of other categories of knife at least have a clear and isolated shape of knife, as shown in Figure 2 (c) and (d). Thus, kitchen knife images were removed from the dataset.



(a) (b) (c) (d)

Figure 2. ***Xray images of knife class.*** *(a), (b): kitchen knife images, (c), (d): other knife images*

Google Images

The biggest issue faced in this topic was that there were not a sufficient number of Xray images to build a robust model without overfitting. The suggested solution was to scrape Google images of the same objects, as shown in Figure 3, which are openly available in tens of thousands of quantities from the internet. When a model is trained with ample Google images, we can develop a technique to adapt the model to perform well for Xray images as well.

Table 1 summarizes the number of Google images and Xray images, which shows that there are much more unique and labeled Google images available than Xray images.

Table 1. ***Number of Google and Xray images.***

Class	No. of Google images	No. of Xray images (original)	No. of Xray Images (no duplicates or kitchen knife)
Knife	**1111**	550	**83**
Gun	**1045**	1050	**111**

Figure 3. ***Examples of scraped Google images of gun and knife.***

Methodology

ResNext-50 Backbone

In this problem, we take the image classification approach, where we will classify the whole image an object. For the main backbone model for image classification, we use Resnext-50 [3] for its relatively few number of parameters than other models such as vgg or AlexNet. It also has lower top-1 and top-5 error than the regular ResNet-50 or -101 [5].

Vision Transformer [7] Backbone

We also experiment with using a fairly recently introduced image classification model named Vision Transformer [7]. Transformer [8] via self-attention has received enormous popularity in applications in natural language processing. Although the convolutional neural networks are still the most popular building blocks in computer vision, some research has been carried out for using transformers for image classification tasks. A recognized model is called Vision Transformer [7], ViT for short, developed by Google. As shown in Figure 4, the model divides an image into 16x16 pixel patches, thus creating 196 image "tokens", and calculate each one's attention to one another using a transformer. Such method was motivated by the idea that ViT imposes a lesser extent of structural bias than CNNs, thus enabling better generalization when trained with a very large amount of images. As expected, ViT resulted in a better classification result than ResNet when trained with a very large dataset (21k images) [7].

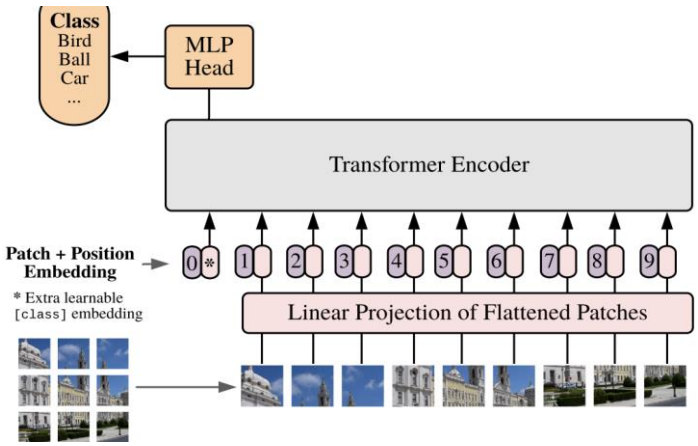


Figure 4. ***Vision Transformer (ViT).***

"Benign" Class

At first, a ResNext-50 model was trained with Google images of two classes (gun and knife). However, when the same model was tested with Google or Xray images that do not contain gun

or knife, it classified nearly all of them as knife, with 90 to 100% confidence. This result was alarming, since most baggage at the airport should not contain a gun or knife, and having such a high false alarm rate would be very inefficient.

Consequently, a third "benign" class was introduced, which represents all objects in the world that are not gun or knife. For this class, google images of any random objects not related to gun and knife were scraped. For Xray domain, given images of classes other than gun and knife (battery, usb, hard disk, and phone) were used.

Adversarial Discriminative Domain Adaptation (ADDA)

Since there are not enough Xray images to build a robust model, we use the google images with rich annotations as training data, and use the small amount of Xray images only for testing the model. Consequently, the specific problem is to reduce the difference in the distributions of Google images and Xray images, so that the same model will work well on both types. Since Xray images have inherently much different characteristics than Google images, a model trained only on the Google images might provide disappointing outcomes when tested on Xray images [2]. This issue is known as *domain shift* [2].

In order to accommodate for domain shift, we introduce Adversarial Discriminative Domain Adaptation (ADDA) [1]. Overview of the model is shown in Figure 5. During training stage, the model first pre-trains a source encoder and a source classifier on the *class labels* of labeled source images. Next, a target encoder, which has the same structure with the source encoder, is initialized with the pre-trained weights of the source encoder from the first step. It then gets trained while the source encoder weights are frozen.

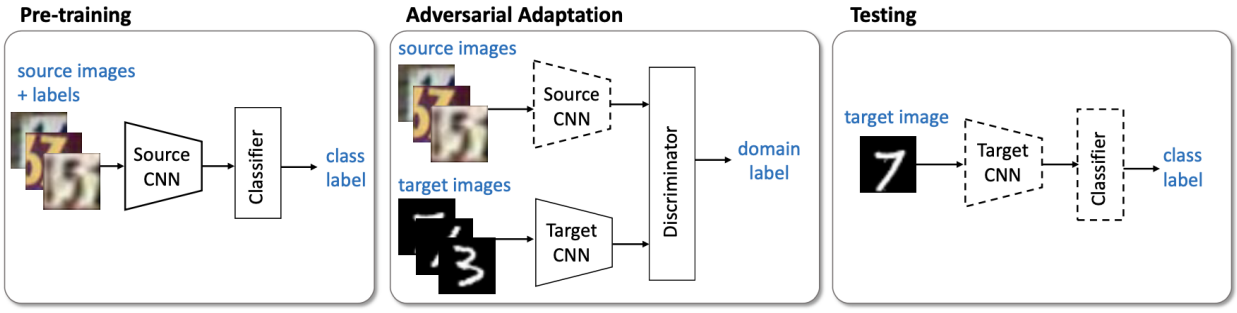


Figure 5. Overview of Adversarial Discriminative Domain Adaptation (ADDA).

The ADDA model assumes that the target domain is unlabeled, so it does not optimize for the *class labels* of the target domain. Instead, it introduces a discriminator model, which is a series of fully connected layers, that tries to classify the *domain label* of each image. It receives both source encoder and target encoder's encoded features of each image, and tries to classify which one belongs to which domain. The goal is to confuse the discriminator so that it will not be able to tell if the encoded feature is from source domain or target domain. This way, the target encoder can be trained to map target images to a similar distribution with the source images.

Lastly, during the test stage, we test the model with target images, using the target encoder and source classifier, to get the class labels for the target images.

Modifications. Few modifications are made to the original ADDA model. Instead of two separate encoders for the source and target domains, a single encoder is used. This eliminates the pre-training stage of the source encoder, and instead trains a single encoder with both source and target images together. Secondly, the original ADDA model trained the target encoder with domain classification loss only, since the source encoder weights were frozen. In the modified model, the single encoder is trained with source classification loss (only using the source images), and is simultaneously trained with domain classification loss with both source and target images. This way, classification and domain adaptation tasks can be done at the same time.

Multi-label Approach

Most Xray images contain many more objects than the object of interest, as shown in Figure 1 and 2, while many Google images have a distinctive, isolated presence of the object of interest. This presence of many other objects in Xray images can be quite confusing to the model that was trained only with Google images. To account for this, instead of making the model to classify the image as one of the three classes, a more flexible approach was taken by making the model predict the probability of the image belonging to each of the three classes.

If using a single-label approach, the label of a single image would be a single number (i.e. 0, 1, 2) for each class. For a multi-label model, a label for an image is composed of three binary numbers (0 and 1), where 1 means the label is present in the image and 0 means it is not.

It is also found to be beneficial if a *soft label* was used for benign class, due to the model's tendency to always predict the benign class with more confidence. Thus, the label for only the benign class is changed to 0.5, while the other ones are left as 1.

All Google images' labels are re-formulated to distinguish between images that are with or without benign objects. For example, in Figure 6, image (a) was given a label of [0, 0, 1] for [isBenign, isGun, isKnife] with only isKnife = 1. Image (b) with a knife and benign objects both present were given a label of [0.5, 0, 1] with isBenign = 0.5 (soft label) and isKnife = 1.

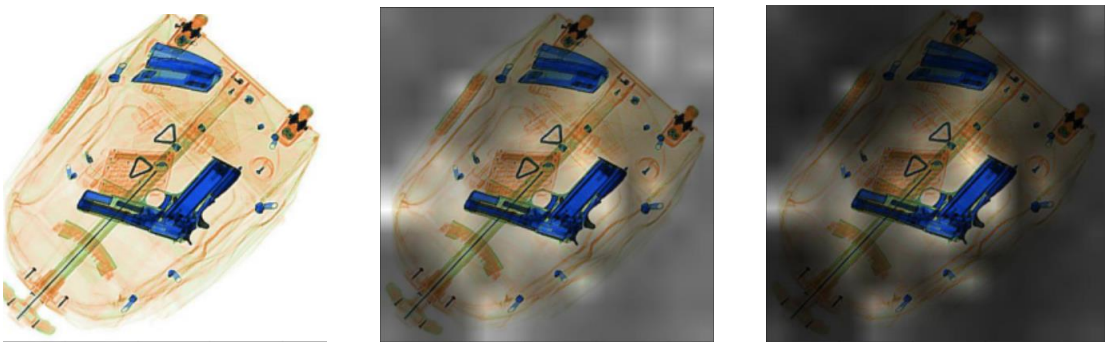
(a) Google image with knife only *(b) Google image with knife and benign objects*

Figure 6. ***Google images of knife class.***

Iterative Erasing Using Visual Attention

When a human operator suspects that there is a gun in an image, she would look closer and focus on the area of the image that she thinks contains the gun and pay less attention to the surrounding areas. Intuitively motivated by this, we experiment with the iterative "erasing"

method during the prediction stage, using the intermediate attention weights available by the Vision Transformer [7] model. The interpolated attention weight matrix, which has the same dimension with the image (e.g. 224 by 224), contains a value between 0 and 1 for each pixel of how much the model attended to the pixel while making the class prediction for the image, with 0 being not at all and 1 being highly attended. When we mask the original image using this attention matrix as the mask, the desired output is having the area containing the object of interest untouched, while the surrounding and background areas darkened. Such an example is shown in Figure 7, with three iterations of erasing. The label P(gun) for each image which indicates the model's predicted probability of the image containing a gun. It can be seen that with each iterative erasing, the model predicts that there is a gun in the image with higher and higher probability. This shows the effectiveness of the iterative erasing method in correctly identifying the object of interest in the image.



(a) P(gun) = 0.023 (b) P(gun) = 0.188 (c) P(gun) = 0.588

Figure 7. ***Three Iterations of "Erasing" the Image Using Attention Weights***

Experiments

For ResNext-50 backbone, we use the architecture and pre-trained weights from ImageNet competition provided in Pytorch. For ViT backbone, we use the architecture using 16x16 patches and weights pre-trained with ImageNet-21K dataset. When training ADDA, the adversarial discriminator consists of 2 fully connected layers with 1024 and 2048 hidden nodes. Each layer uses a ReLU activation function. Optimization uses Adam optimizer for 50 epochs with a learning rate of 2e-6, and a batch size of 20 images. All training images are rescaled to 224x224 pixels. We use mean square error (MSE) loss for class classification by the encoder (regarding the soft label for the benign class), and cross-entropy loss for domain classification by the discriminator. ADDA was trained using ResNext-50 backbone only, because using ViT backbone showed a non-sensical result of all images being predicted as a single class, with the single-predicted class alternating between different classes at each training iteration. We are currently looking into possible causes of this odd problem.

Results

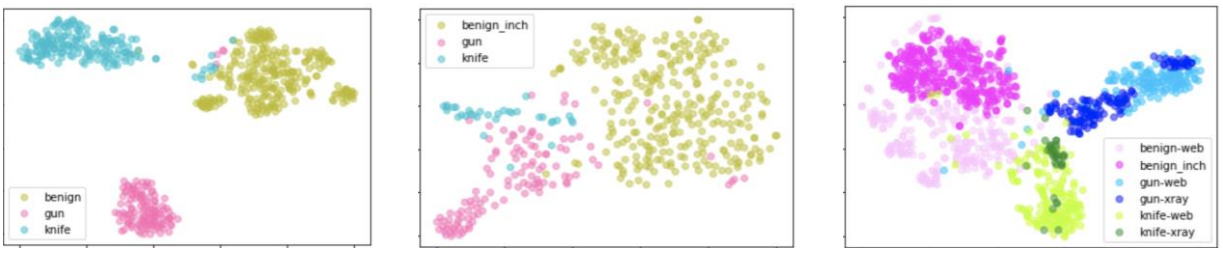
Summary. The summary of the experiments is presented in Table 2, which shows that using a multi-label approach with ADDA using ResNext-50 backbone achieved the best result with both benign and gun classes. It is also noted that the multi-label ViT model trained with google

images only reaches gun and knife recalls comparable with ADDA when predicting with the iterative erasing strategy.

Table 2. Recall for Each Class Using Different Models.

Method	**Benign Recall**	**Gun Recall**	**Knife Recall**
ResNext-50, Source-only, single-label	0.997	0.137	0.072
ResNext-50, Source-only, multi-label	0.997	0.436	0.145
ViT, Source-only, single-label	1.000	0.359	0.545
ViT, Source-only, multi-label (without iterative erasing prediction strategy)	1.000	0.436	0.424
ViT, Source-only, multi-label (with iterative erasing prediction strategy)	1.000	0.709	0.609
ADDA (ResNext-50), single-label	**1.000**	0.479	0.084
ADDA (ResNext-50**), multi-label**	**1.000**	**0.778**	**0.697**

t-SNE plots for class labels. For *class-wise* qualitative analysis, Figure 8 shows t-SNE [6] plots that visualize the feature distributions of the multi-label ADDA model for different class labels (shown in different colours). Plot (a) for Google image features show quite well-seperated three clusters for the three class labels, while plot (b) for Xray image features show much unclear boundaries between the three classes, which explain the lower recalls for gun and knife images for Xray domain. Plot (c) shows overlaid features of both domains, using the same colour for the same class label while lighter for Google domain and darker for Xray domain. It shows that the features with the same class label across different domains are located near each other (e.g. skyblue and darker blue points representing Google gun images and Xray gun images, respectively, are near each other). This semantically equivalent distribution of image features from the two different domains is called *semantic alignment* [9].



(a) Google image features (b) Xray image features (c) Google + Xray overlaid

Figure 8. ***t-SNE of class labels.*** *t-SNE visualizations of ADDA features of Google and Xray images of three different labels.*

t-SNE plots for domain labels. For *domain-wise* qualitative analysis, Figure 9 visualizes with t-SNE [6] the feature representations of the ADDA model with the two different domain labels (Google as target domain- green and Xray as source domain- pink). Near the beginning at epoch 5 of ADDA training, the two domains look quite separated. As training proceeds; however, the two become closer. At epoch 25, clusters of the two domains appear much closer to each other.

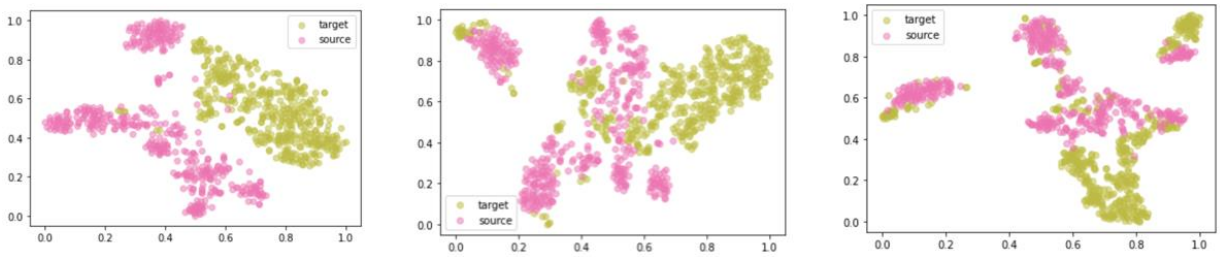


Figure 9. ***t-SNE of domain labels.*** *t-SNE visualizations of ADDA features of Google and Xray images at epoch 5, 15, and 25 of ADDA training.*

Conclusion

The ADDA model with multi-label approach was shown to be a simple yet effective method for domain adaptation between Google images and scanned Xray images, despite large domain shifts. The effectiveness of the proposed method was demonstrated through higher recall for both gun and knife classes than using a single-label, source-only model without domain adaptation. Since this is still an initial model development stage of the research, more effective and intuitive approaches will be considered in the future.

References

[1] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *CVPR*, 2017.

[2] Seungmin Lee, Dongwan Kim, Namil Kim, and Seong-Gyun Jeong. Drop to Adapt: Learning Discriminative Features for Unsupervised Domain Adaptation. In *IEEE CVPR*, 2019.

[3] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated Residual Transformations for Deep Neural Networks. In *CVPR*, 2017.

[4] Hui Tang and Kui Jia. Discriminative Adversarial Domain Adaptation. In AAAI, 2020.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[6] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. Journal of Machine Learning Research, 2008.

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *ICLR*, 2021.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

[9] Saeid Motiian, Marco Piccirilli, Donald A. Adjeroh, Gianfranco Doretto. Unified Deep Supervised Domain Adaptation and Generalization. In IEEE ICCV, 2017.

Detecting TTC Power Rail Anomaly from Infrared Recordings

Tushar Aggarwal

Introduction

Toronto Transit Commission (TTC) Line 3 in Scarborough, Ontario is a 6.4km long light rail transit (LRT) railway line which runs from Kennedy to McCowan; the train connects 6 stations in its path. In May 2017, during operation, a train sustained some damages caused due to an anomaly in the power rail which resulted in evacuation and line 3 closure for several hours. Anomalies are high temperature spots along the power rail which could cause the metal to bend and deform.

To prevent another incident in line 3, TTC actively monitors and does preventative maintenance along the track. One of the key methods used by TTC is actively scanning the track using FLIR infrared cameras. The thermal camera is mounted on the front of the train which scans and records the track as the train runs throughout the day. This allows power rail anomalies to be identified from the scanned thermal images as high temperature spots show up on track and then fixed by maintenance team to prevent any damage or unexpected closure of line 3. Though, this solution is effective it has some major drawback; the recorded thermal video of the track is manually checked and observed for anomaly with each video spanning 15-20 minutes. This requires long span attention and require high amount of worker time which makes such task to be a more suitable application for computer vision applications.

The goal of this project is to create an anomaly detector which detects railway track anomalies in video frames recorded by the FLIR thermal camera.

In 2018, Tuocheng Liu from C-MORE Labs at University of Toronto, worked on this project using two algorithms: Single Shot Detector (SSD) and Faster R-CNN. Both Algorithms have shown remarkable results in the field of object detection. However, for detecting anomalies a major requirement for the project is to have high recall value of nearly 1 (100%) and high precision above 85%. While both algorithm implementation by Liu showed high precision values, the recall values were only 75%. This project builds on the work of Liu in exploring new models and algorithms.

Background

For this project following methods were used:

1. Object Detection
2. Classification
3. Anomaly Detection Algorithms

Object Detection
Object detection is one of the key applications of computer vision where algorithms are trained to detect an instance of a particular class in each image/video. Classification and localization are the two main working principle of an object detector. The true positive frame shows a bounding box around the detected object along with the detection confidence score and class label of the object.

YoloV4 model was selected to detect anomalies (high temperature spots) in power rails as one class.

Classification
Classification is one of the key methods in machine learning specially for image classification. For this project convolutional neural networks were used to classify images into two classes namely: Anomalous frames, and Normal Frame. The dataset was split in 80% training images and 20% validation images with equal number of anomalous and normal images in both classes for training and testing.

Convolutional neural network learns and extract the details in the image while then a fully connected neural network is used for classification.

In addition to the classification in two classes, a single class convolutional neural network (CNN) was also trained using normal images. This was to test if the single class CNN can detect anomalous images from normal images.

Anomaly detection algorithms
Anomaly detection algorithms are used for detecting anomalies in several different types of datasets. One Class Support Vector Machine (OC-SVM) is one of the selected algorithms which is a modified version of SVM. Isolation forest is the second algorithm used for detecting anomalies.

Anomaly detector algorithms are trained on normal data points; This allows the algorithm to fit on normal data and when an outlier is present which does not conform to normal dataset it is detected as an anomaly. Since images are very high dimensional data, both anomaly detection algorithms were used in conjunction with a feature extractor model such as VGG-16 for dimensionality reduction.

Metrics
For the two classes (anomalous and normal frames/data point) a confusion matrix is used which can help derive the evaluation matrix.

	Predicted Conditions	
	Positive	Negative

True Condition	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Figure 1: Confusion Matrix

Where true positives are the predicted frames that are positive predictions and are positive. That is an anomalous frame is predicted as anomalous.

False Negatives are frame that are positive for a particular class but are detected as negative. An example of this is an anomalous frame detected as normal frame.

False positive are frames that are negative frames but are classified as positive frames. An example of this is anomalous frames being classified as normal frame.

$\underline{\text{Recall}} = \frac{TP}{TP+FN}$

The proportion of actual positives that are correctly detected. It tells how much of the frames with human observed anomalies are identified by the model.

$\underline{\text{Precision}} = \frac{TP}{TP+FP}$

The proportion of true positives to all detected positives. It tells how much of the detections are correct in all inference results.

Tuning and optimization
Methods like grid search were used to find the best hyper parameters for different algorithms. This helps to optimize the algorithm to give best results.

Dataset and configuration

TTC provided two sets of videos scanning Line 3. One set contains 4 videos recorded in December 2017 (winter set) with each video containing approximately 16000 frames, the second set contains 8 videos in May 2018 (summer set) with each video containing approximate 10,000 frames. Each set contains videos of Northbound as well as Southbound videos of the train; however, since the camera was mounted on the front of the train, some of the videos are not useful as the camera is not looking at the power rail. This gives us only three videos useful in the winter dataset and only 2 videos useful in the summer set.

An expert from TTC manually observed the videos to identify several anomalies in the winter data set. A total of 45 anomalies were identified with 9 anomalies found on the power rail on the other side of the track. Since the two datasets are recorded at different time of the year, they cannot be used together without domain adaptation which is beyond the scope of this project. Since, only winter dataset was annotated by an expert. Only the winter dataset was used for training and testing. However, summer dataset was reviewed to find anomalies, a total of 3 anomalies were identified. Which shall be used for future work.

It is important to note that the videos provided are not visual data stored as RGB/Grayscale channel(s), but the intensity values of radiation recorded by the thermal camera as a single

channel matrix. The frames can be visualized in multiple ways, however, for the winter videos temperature values setting were selected before extracting images. All the frames were extracted as '.jpg 'image files without using any colour palette; this was done to help algorithms identify anomalies specially in scenarios where the train was running in the tunnels. A special software from FLIR was used for frames manipulation called ResearchIR before the images were extracted.

For this project four set of image datasets were created from the same winter set video frames. The first dataset was 640x512 images that were provided with the project (LIU's document). The anomaly markings and bounding boxes for YoloV4 were created for this dataset and were used in the YoloV4 algorithm. A total of 779 Anomalies were labeled. It is important to note that this dataset was a 3 channel RGB dataset where a colour palette was applied on the images using ResearchIR software. For winter dataset, pascal VOC bounding boxes information was also available from previous student which was used to create new bounding box information for YOLOV4 algorithm using a python script. For YoloV4 algorithm's implementation, it was ensured that all the frames for an anomaly only appeared either in the training set or validation/test set.

Datasets 2 and 3 were created with the help of ResearchIR software, dataset 2 was with high temperature threshold values which only showed high temperature spots (between 11^{o}C to 21^{o}C) as shown in figure 11 and dataset 3 was created without any threshold values as shown in figure 10. These 35000+ images in both datasets were then reviewed to extract anomalies from original datasets and to remove all the files with noise such as high temperature spots due to wires, etc. Datasets 2 and 3 were then combined to increase the dataset images for training and testing for some experiments. However, anomaly detection and classification algorithms were also tested on dataset 3 (original images) alone. Images for datasets 2 and 3 were both grayscale images to reduce complexity for the convolutional neural network. The final dataset had 779 anomalies.

Dataset 4 was created from dataset 3 where the images were cropped to bottom left corner to keep most of the track in the center of the image as shown in figure 12. This dataset was also used for experimentation with anomaly detection and classification algorithms. These images were further reviewed for to separate anomalies from normal images as some of the anomalies were cropped. This brought total number of anomalies from 779 to 668.

All the anomalies were handpicked from the datasets and moved to different folders. Additionally, a python script using OpenCV library was used to resize the images based on algorithm input requirements. Apart from the provided data, some extra anomalies were created using photoshop that were experimented during model training.

In addition to our dataset, data augmentation technique was used in TensorFlow to increase the size of our dataset while training.

Figure 2 : Image in Dataset 3

Figure 3: Image in Dataset 2

Figure 4: Dataset 4 - Before and After Cropping

Experiment and Results

Anomaly detector development requires a lot of experimentations. Different algorithms and combinations were used to find the best results. For the first part of the experiment, an object detection algorithm was used.

The project previously showed some good results using the object detection models such as Single Shot Detector (SSD) and Faster RCNN. However, with newer advancements better detectors have been developed since then. As a results YOLO V4 was selected for an updated algorithm, to detect anomalies in the winter dataset.

YoloV4

For YoloV4 Dataset 1 was used which contained images in RGB channel and the model was trained only on the anomalies. Out of the 45 anomalies the model was trained using 40 anomalies (705 frames) and the 5 anomalies in test set (74 Images). The model was trained for 3000 iterations and 6000 iterations. The training loss for the model is shown in figure 6 where

the model loss descends to a very low value. The results for the YoloV4 model are shown in table 1 below.

Table 1: YoloV4 result summary

	3000 iterations	6000 iterations
Average Precision (ap)	93.74%	83.88%
True Positive (TP)	88	88
False Positive (FP)	6	10
False Negative (FN)	11	11
Precision	0.94 @C =0.25	0.90 @C =0.25
Recall	0.89	0.89
Average IoU	0.91	0.89

Tracking the key metric of recall and precision it can be inferred from table that the model performed best with 3000 iterations showing high precision of 93.7% and a recall value of 89%. However, when the number of iterations are increased, the model is overfitting on the training dataset as the model's precision reduces significantly. While the YoloV4 algorithm shows promising results, it failed to provide a very high recall value which is required for the project as any missed True positive frames would require manual inspection of the videos which would defeat the purpose. Figure 5 shows some of the correct classification by the Yolo Algorithm.

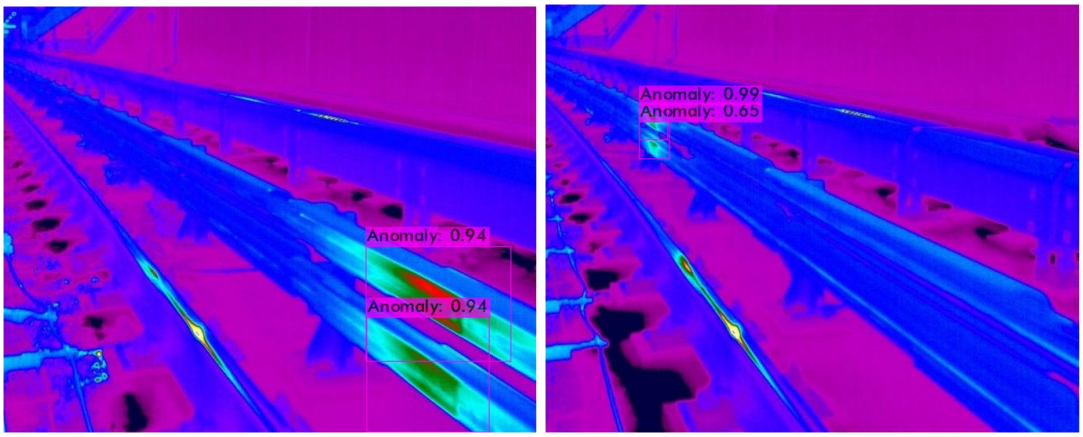


Figure 5: YOLO V4 results examples

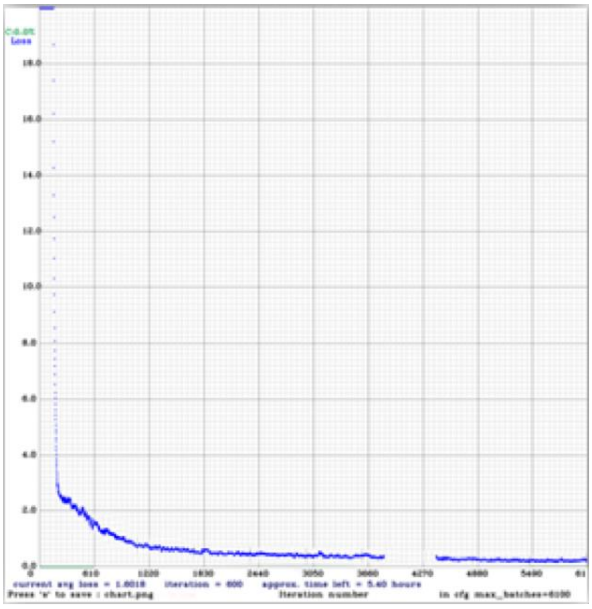


Figure 6: YoloV4 training loss graph

Classification

Deep Convolution Neural Network.
While YoloV4 was able to show promising results, it did not show high recall value which led to exploring other alternatives such as using state of the art deep convolutional neural networks (CNN) such as VGG-16 for the purpose of classifications.

In the next experiment a modified VGG-16 model was used for image classification, where 1400 images of anomalies and normal frames of each were selected from the combination of dataset 2 and dataset 3. The model was trained for 100 epochs as shown in figure 7 below.

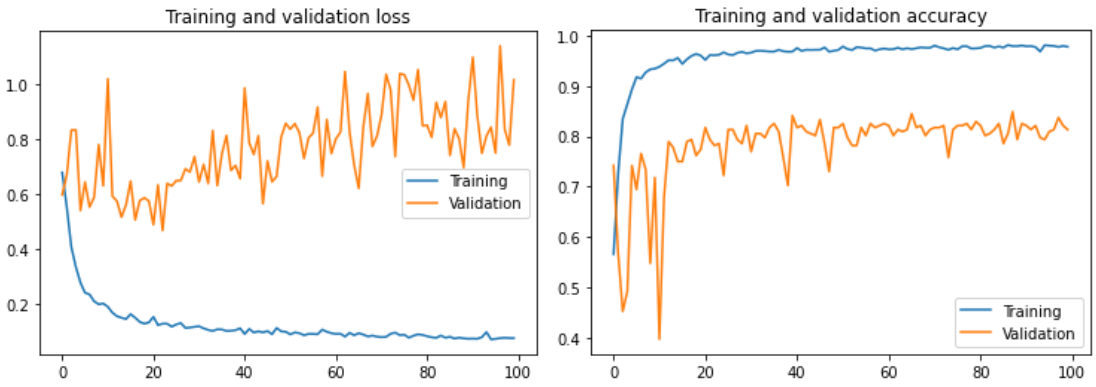


Figure 7: VGG-16 Training/Validation accuracy and loss graphs

It can be observed that the training loss for training set decreased while the validation loss for the model increased and kept fluctuating. Additionally, the training accuracy of the model was close to 1 for the training set while the model's validation set accuracy is not very high.

However, testing the model on the test dataset the model misclassified all the anomalies as normal frames. Additional experiments were used with regularization methods such as using dropout for neural network nodes and batch normalization for convolutional neural network which also failed to resolve the problem.

Additionally, looking at the GradCAM++ visualization of the VGG-16 trained detector shown in figure 16 below, it could be observed that the model is able to identify the anomalies, However, due to deep convolutions, the details of anomalies might be lost or other bright spots due to background noise are contributing misclassification of anomalies.

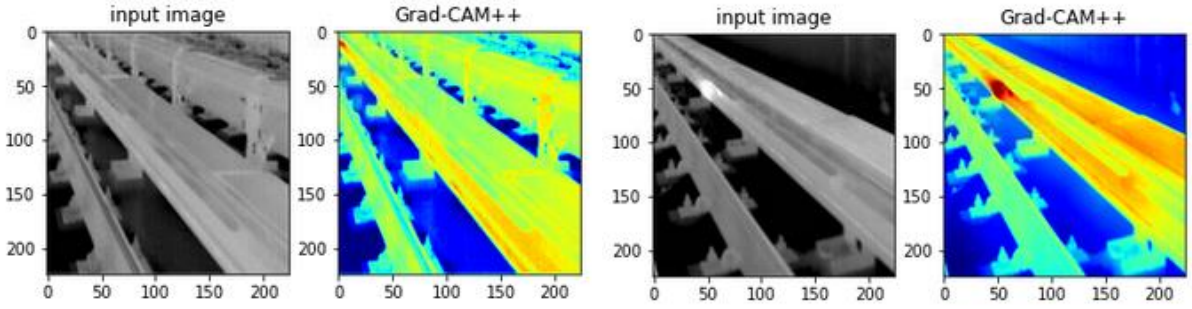


Figure 8: GradCAM++ Visualization of VGG-16 network. Left: normal frame; right: Anomaly frame

One Class Convolution Neural Network

In addition to using classification network experiment tried above one class convolution network for anomaly detection was used as described in [2]. In this experiment the model is trained only on normal frames (35,000+frames). The model as described in literature review adds gaussian noise to the feature extractor's output and concatenate the noise images with the feature extractor output. However, this method was not able to learn and classify the anomalous images correctly leading to very low precision and recall value, hence the algorithm was not tested any further. The feature extractor used for this algorithm is VGG-16.

AutoEncoder

Autoencoders have shown great potential in anomaly detection especially detecting anomalies in still camera surveillance, therefore, normal convolutional autoencoders and robust convolutional autoencoders are explored as part of the experiment in building an anomaly detector.

Normal AutoEncoder

Normal autoencoder is trained with all the normal frames to learn all the normal features of the power rail. The concept behind the autoencoder is that the reconstruction loss of normal frames will be significantly lower than the anomalous frames. Shallow and deep convolutional auto encoders were also explored for this project. As shown in the figures 9 the original and reconstruction images for shallow and deep convolutional autoencoders.

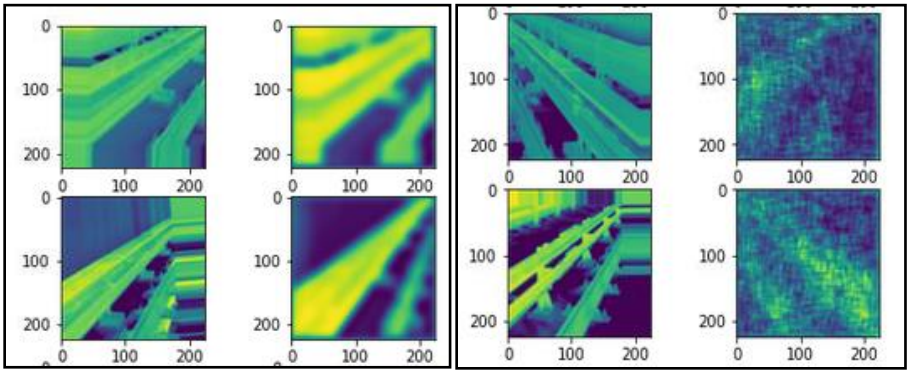


Figure 9: AutoEncoder output visualization left: Shallow & right: Deep

The shallow autoencoder was trained for 100 epochs and deeper autoencoder was trained for 50 epochs with normal frames from dataset 3. However, shallow autoencoders showed the best results. Looking at the evaluation metrics in table 2 below for shallow encoder.

Table 2: Shallow AutoEncoder Result Metrics

Precision	Recall
47%	76%

The encoder is able to recall several anomalous frames, however, it is not able to differentiate between several normal frames from anomalous frames giving us very low precision. Therefore, autoencoder was not selected as a viable model for anomaly detection for power rails alone.

Robust Convolutional AutoEncoder

For the second part of the experiment, a Robust Convolutional Autoencoder [4] with neural networks in the middle was used for anomaly detection. The model was trained for 100 epochs with all normal frames from dataset 3, the results are tabulated in table 3.

Table 3: Robust AutoEncoder Result Metrics

Precision	Recall
51%	88%

Comparing robust autoencoder results to normal autoencoder results it could be observed that recall value and precision of the model increased for robust autoencoder. Recall values increased significantly while precision is still around 50% which is less than what is acceptable for the project.

While auto encoders show better results for recall they fail in precision, therefore, should not be used individually as an anomaly detection model.

Anomaly Algorithm

Isolation forest and one class SVM are two algorithms that are trained on normal images. This allows the algorithm to detect anomalous datapoint as discussed in literature review. To work with high dimensional image dataset, all the images were passed through VGG-16 model for

feature extractors through the convolution layers with the last convolution layer being global average pooling. This allows the data to be reduced to lower dimension from 224x224x1 to 512x1x1). Please note the images are only passed through convolution layers and the neural network portion of the model is not used.

Two different pretrained VGG-16 feature extractors were used with Isolation Forest and One Class SVM for experimentation purposes. The first VGG-16 network was pretrained on ImageNet dataset and transfer learning was used to extract features on the TTC power rail dataset. The second VGG-16 network was trained on TTC power rail dataset from the classification task (discussed above). The results for both cases are tabulated and discussed below.

Isolation Forest

Table 4: Isolation Forest Result Metrics table

	Pretrained on ImageNet dataset		Pretrained on TTC Dataset	
Iteration	**Precision**	**Recall**	**Precision**	**Recall**
1	51%	95%	52%	100%
2	76%	49%	**64%**	**94%**
3			71%	81%

From table 4 it can be inferred that the VGG-16 model pipeline that was trained on TTC image dataset showed best results with a recall value of 94% and a precision value of 64%. Though the recall value is significantly higher than the autoencoders that were discussed, it can be observed that model has a very low precision in classifying normal frames and anomalies correctly. The hyperparameters selected for the algorithms to give best results were found using grid search.

One Class Support Vector Machine (OC-SVM)

Table 5: One Class SVM Result Metrics Table

	Pretrained on ImageNet dataset		Pretrained on TTC Dataset	
Iteration	**Precision**	**Recall**	**Precision**	**Recall**
1	44%	99%	53%	100%
2	70%	68%	61%	93%

Similar to Isolation Forest it can be seen from table 5 that OC-SVM works best with VGG-16 model that was pretrained using TTC power rail dataset. While the OC-SVM is able to provide a recall value of 93%, the lower precision value shows that the model is not able to identify normal data points correctly. This could be due to the extra information in an image present around the power rail tracks and anomalies that are contributing to the prediction as most of the features in a normal and anomalous frame are same. Hence, making it harder for the detector to classify.

Conclusion and Future Work

From different anomaly detection models tested in this project it can be observed that YOLOV4 performed the best in providing a balance between recall (89%) and precision (93.7%) and should be explored more for improvement. The YoloV4 has shown improvement compared to predecessor experiment of using Faster RCNN and SSD networks. Therefore, YoloV4 network should be explored with cross validation and hyperparameter tuning.

Isolation Forest and One Class SVM have shown promising results in recalling all the anomalies, however, they lack in precision. An issue encountered here was the loss of information in deeper convolution of VGG network, the effects of which were also seen in classification experiment where all anomalies were misclassified. A way to improve the Isolation forest and OC-SVM algorithm is by using a shallower feature extractor with a way of passing information of anomalies to the prediction stage. Newer more complex networks can be developed along with Isolation Forest or OC-SVM to identify anomalies in the Images.

Additionally, autoencoders showed promising results, however, due to moving camera along the track it was hard for the autoencoder to recreate a lot of the normal frames properly. Increasing training time and experimenting with different network design with shallow encoders plus decoders could help resolve this issue.

While the end goal of the project is to have a recall value of 1, however, since each anomaly stays in video for 10-20 frames, the recall value goal can be set slightly lower if a fixed number of minimum frames per anomaly are identified by the model. Having a perfect recall with high precision requires an excellent model which is trained and optimized for large and consistent datasets. Line 3 has most of the track elevated, different weather conditions cause the images to change; sun being at different angles reflect from power rail and shines into the camera which generates noise in the images. Additionally, different light settings effect image quality as well as different time of day can cause track to heat up or cool down. Designing a model with the limited dataset is a challenge and getting more data is currently out of scope of the project. A final model for deployment should allow for changing environment during the day and season.

The summer dataset though reviewed and labelled, was not used during experimentation due to different weather conditions resulting in a dataset that is different from the winter dataset. To use different datasets (summer set), domain adaptation techniques should be applied to the model to incorporate summer videos.

A recommendation for the future is to explore YoloV4 by modifying the underlying architecture of the model to improve recall and accuracy along with cross validation. Additional anomalies created using photoshop can also be incorporated for training to increase the dataset size and improving the detector. Changing filters and hyperparameters for YoloV4 shall also be explored for improved performance and pushing precision and recall to above 95%.

An ensemble of the currently trained models should also be tested using bagging where the final classification result is detected based on average classification of each frame. This will allow the predictions to be averaged across all models and improve frame classification.

Noise in many frames was a key issue found while reviewing many frames. Developing an algorithm that reviews classification of 2 frames before and after an anomaly is detected could solve this issue. While noise stays for few frames, anomalies stay for 10-25 frames. If 3 to 5 consecutive frames are classified as anomalies, then the image can be classified as an anomaly in

those frames otherwise it can be classified as noise therefore a normal frame. This method could also help remove any misclassification and improve precision of the models.

Several videos in the dataset were not pointing at the power rail or the power rail was not in the center of the frame. A mounting solution and a standardized method should be proposed to TTC to ensure the data is consistent in the future which could help improve developing better models and improve their performance metrics.

References

[1] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, YOloV4: Optimal Speed and Accuracy of Object Detection, April 23, 2020

[2] Poojan Oza, Vishal M. Patel, One Class Convolutional Neural Network, January 24, 2019

[3] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, Vineeth N Balasubramanian, Grad-CAM++: Improved Visual Explanations for deep convolutional Networks, October 20, 2017

[4] Raghavendra Chalapathy , Aditya Krishna Menon , and Sanjay Chawla: Robust, Deep and Inductive Anomaly Detection, June 30, 2017

[5] Raghavendra Chalapathy , Aditya Krishna Menon , and Sanjay ChawlaAnoma, Group Anomaly Detection using deep generative models, April 13, 2018

[6] Karen Simonyan, Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition, September 4, 2014

[7] Image: https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

[8] image: https://towardsdatascience.com/convolutional-autoencoders-for-image-noise-reduction-32fce9fc1763

[9] website: scikit-learn.org

[10] website: https://www.researchgate.net/figure/Fig-A1-The-standard-VGG-16-network-architecture-as-proposed-in-32-Note-that-only_fig3_322512435

Predictive Maintenance of Hydroforming Equipment using Failure Analysis

Sahil Nagpal

Background

The hydroforming process is a cold working forming process, which uses hydro pressure to shape metals into complex shapes. Hydroforming is often utilized, when the desired shape of a component is too complex to accurately manufacture using other forming operations. One example of such complex components are the side rails for engine subframes for the Automotive Industry. The subframe is an integral part of the chassis of the vehicle, as it has numerous attachment points to various other components, such as control arms. Therefore, it is paramount, that there are no defects present on the frame, as it can compromise the safety of the entire chassis.

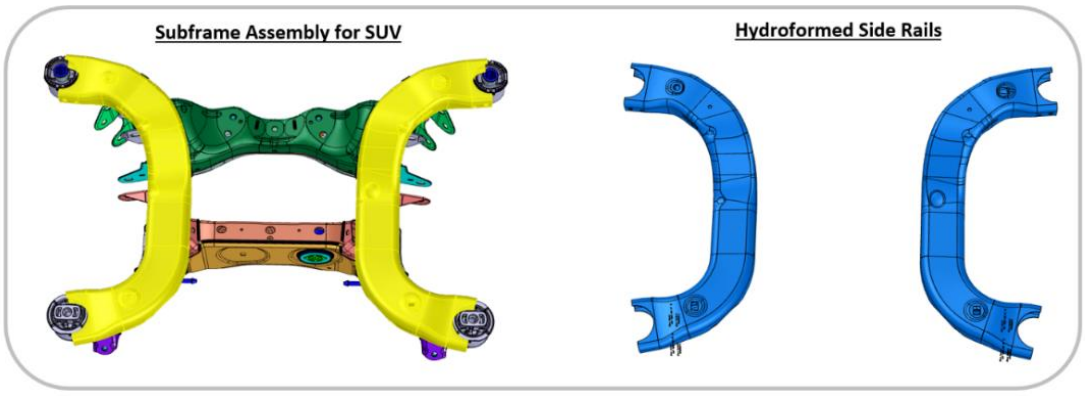


Figure 1: *Hydroformed side rails of a subframe.*

There are certain components within the Hydroform presses, which if not replaced or repaired at regular intervals, have a direct correlation to the scrap being generated. Such components include, check valves, and piston seals. If any of these components fail, the Hydroform press will lose pressure, and begin to produce deformed parts.

The main reason these components malfunction is due to wear, and thus the thesis of this project is to use Failure Analysis to predict the likely timeframe that each component will need to be repaired or replaced.

Analysis

Currently, failure data for each component is available and stored in a MySQL database. The majority of the data is not censored, as the components are only replaced once absolute failure occurs.

A framework was written in Python to access and manipulate the data stored in the MySQL database for any component. As an example, all data pertaining to the component "Check valve" was tabulated using this framework. This data can then be used in numerous Failure analysis techniques.

Machine ID	comment	machine_id	start_time	end_time	duration	downtime_code_id	Name_2	Failure_times(hours)
HYD2 Press 2		12	05:02.1	16:17.0	674.635	217	check valve	22.49570278
		12	29:11.1	41:12.0	720.606	217	check valve	27.95312222
		12	31:28.1	05:03.1	2014.65	217	check valve	66.99730556
		12	28:45.1	46:20.0	1054.615	1094	hydro- press problem intensifier check valve issue	104.7240639
		12	04:05.0	51:47.0	2861.99	1094	hydro- press problem intensifier check valve issue	243.5240389
		12	08:43.0	30:00.0	1276.577	217	check valve	284.4019917
		12	26:52.0	07:42.1	2450.407	217	check valve	389.4732083
	mechanic replace the check valve	12	40:00.0	47:02.0	422.299	330	maintenance	449.9827583
		12	09:02.0	30:00.0	1257.905	217	check valve	488.0730417
	changed check valve	12	20:53.0	51:23.0	1829.994	355	hydro- press problem unknown issue	512.3499028
		12	36:44.1	00:12.1	1408.441	1094	hydro- press problem intensifier check valve issue	880.707275
	change check valve	12	28:47.1	36:55.0	487.589	217	check valve	1313.468539
		12	11:17.1	56:32.0	2714.161	217	check valve	1363.494572
		12	30:00.0	55:27.1	1527.857	1094	hydro- press problem intensifier check valve issue	

Figure 2: *Data extraction using Python*

Weibull Distribution

As a first attempt, Weibull Distribution was used to determine if it can adequately be used to model the time to failure for each component. The steps taken to obtain the distribution parameters are shown below. *Note, that this procedure was done for all individual components.*

1. Time to failures were sorted in ascending order.

2. As the sample size is quite small, Bernard's Approximation was used to calculate the Median Ranks of the failure data. [1]

$$\boldsymbol{Bernard's\ Approximation}: \boldsymbol{F(t_i) = \frac{i-0.3}{n+0.4}}$$

3. The Reliability function of the two parameter Weibull Distribution is the following [2]:

$$\boldsymbol{R(T) = e^{-\left(\frac{T}{\eta}\right)^{\beta}}}$$

Consequently, the CDF or Failure function of the two parameter Weibull Distribution is the following [3]:

$$\boldsymbol{F(T) = 1 - e^{-\left(\frac{T}{\eta}\right)^{\beta}}}$$

The Failure function was linearized, by taking logs:

$$\ln\left(ln\left(\frac{1}{1-F(T)}\right)\right) = \beta(\ln(T)) - \beta\ln(\eta)$$

$$y = \ln\left(ln\left(\frac{1}{1-F(T)}\right)\right)$$

$$x = \ln(T)$$

4. The x, and y values were calculated using the above, and graphed. As can be seen from the below figure, the values can be modeled using a linear line, which indicates that the Weibull distribution is a good fit.

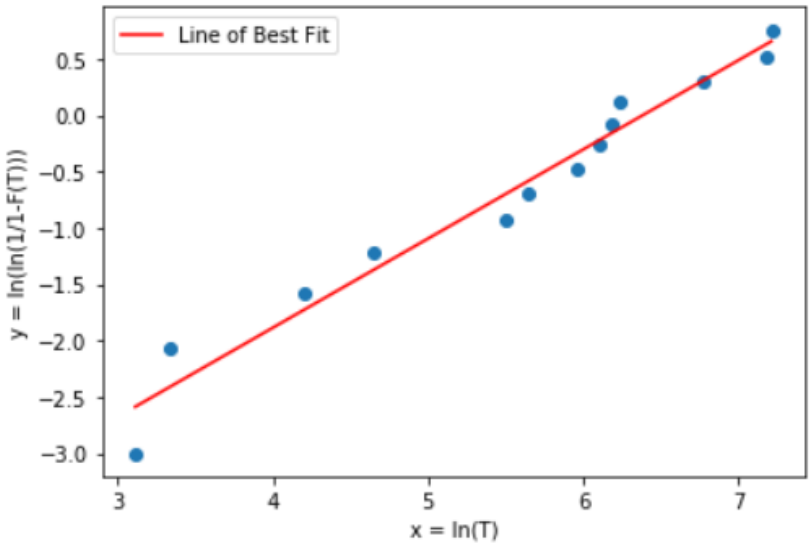


Figure 3: *Line was fit to the transformed data, indicating that a Weibull distribution is a good fit.*

The equation of the best fit line was determined to be: $y = 0.79081\,x + -5.04854$

5. The parameters of the Weibull distribution η *and* β were calculated using the following equations:

$$\beta = slope\ of\ the\ best\ fit\ line$$

$$\eta = e^{\frac{y_{intercept}}{\beta}}$$

Figure below shows the Weibull curve overlaid the original failure data.

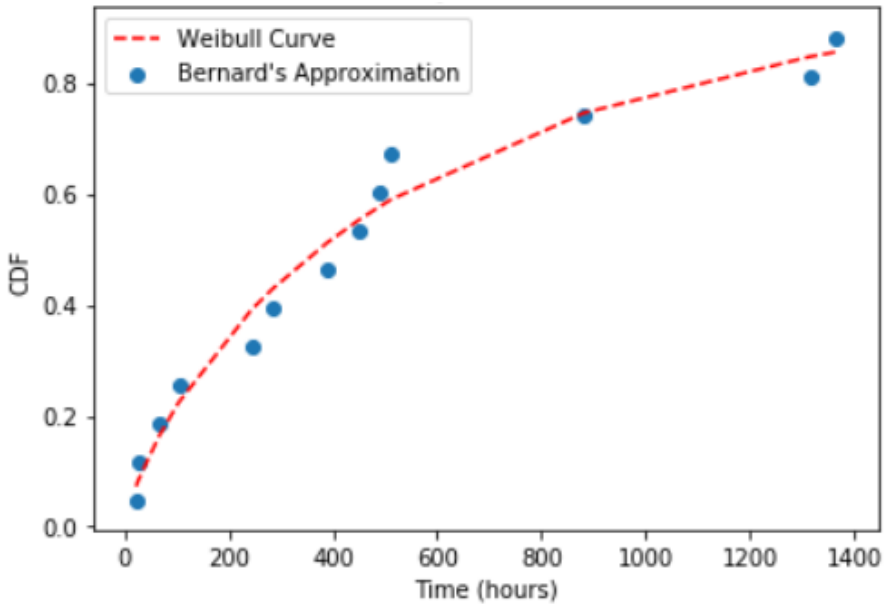


Figure 4: *Weibull curve generated using the calculated Weibull Parameters.*

6. Finally, the probability of failure was calculated for each component over the next 2 weeks (240 hours). This was done using conditional probability:

$$P(\ Operational\ Time\ \leq T \leq Operational\ Time + 240\ Hours \mid T > Operational\ Time) = \frac{F(Operational\ Time + 240\ Hours) - F(Operational\ Time)}{R(Operational\ Time)}$$

Communication of Results

Another important aspect of this project is to effectively communicate the results to the maintenance department. The process of communicating these results is as follows:

1. With the use of python, data is exported to a csv file in the desired format, which includes all calculations, such as the Weibull parameters. This was set-up to be done automatically every 8 hours or end of every shift.

2. Power BI is used to display the results in a dashboard (*see figure 5*). The key metrics which are chosen for display on the dashboard are the "Current Operational Time" and "Probability of Failure for the next 2 weeks".

3. Maintenance personnel, can then quickly get a high-level overview of which components are at a high risk of failure. This actionable insight can then be used schedule maintenance activities to replace the components before they wear, and cause downtime, or quality defects.

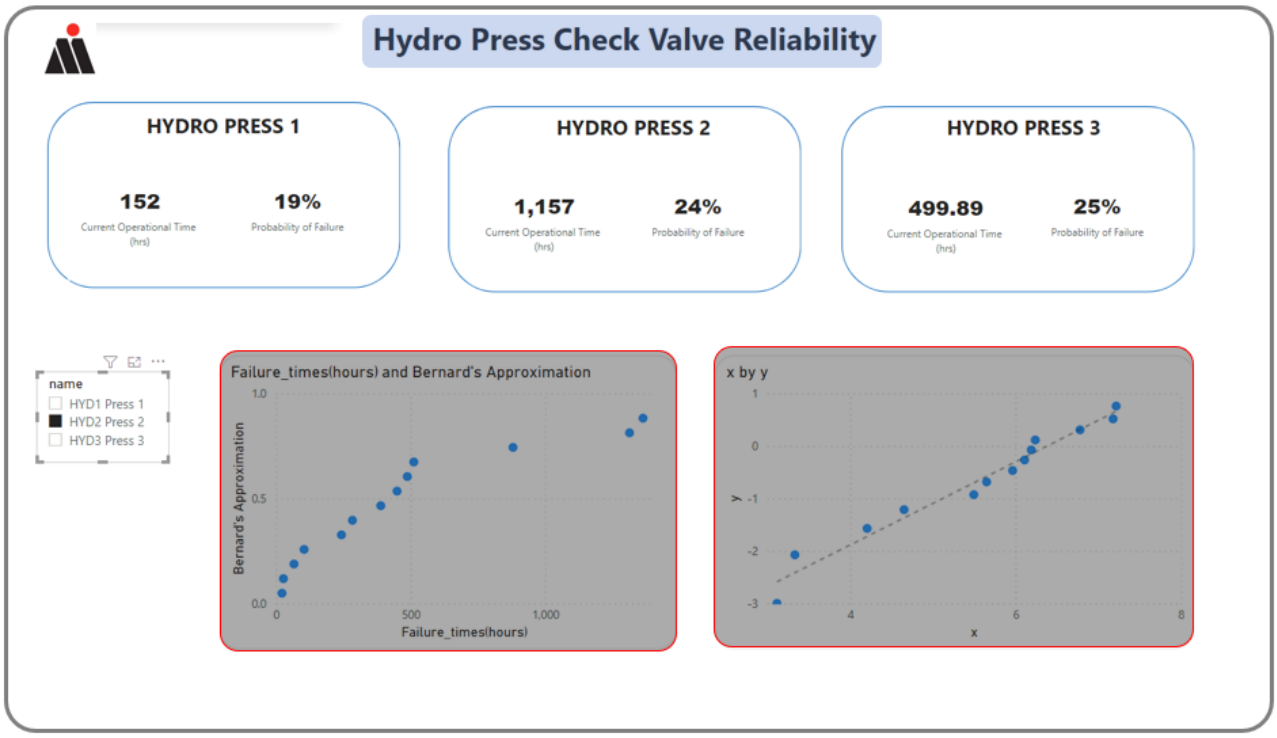


Figure 5: *Power BI Dashboard.*

Next Steps

In addition to using previous failure data, there is also opportunity to use real time sensor data, such as Intensifier Pressure, and Fluid volume, to determine the Remaining Useful life of the components. The proposed methodology is to model the remaining useful life of the check valves by monitoring the fluctuations in pressure and volume of fluid used for each hydroforming cycle over time. The hypothesis here is that as the check valves begin to wear out, there will be certain trends in the pressure and volume signatures which can provide insight into the remaining useful life of the valve. This investigation is ongoing, as sensors are still being installed, and the data pipeline is in the process of being built. As an added bonus of collecting this data, the pressure signatures, and volume could also be used to train a classification machine learning model, to predict whether the current part created may contain defects or not, which is also another concern.

Conclusion

In summary, the Weibull Distribution is doing an adequate job in modelling the failure times of Hydroforming components. With this insight, maintenance personnel can more strategically allocate their man power and replace components before they begin to cause defects or equipment downtime. The next step of the project is to investigate the use of Pressure and Volume as explanatory variables to aid in RUL estimations.

References

[1] Dragran Banjevic, 2021, *MIE1721 – Introduction to Weibull Analysis From MIE 469, slide # 6.*

[2] Dr John Bentley, 1999, *Introduction to Reliability and Quality Engineering, 2nd Edition, Pearson Education Publishing, Pg 37-38.*

[3] Dr John Bentley, *Introduction to Reliability and Quality Engineering, 2nd Edition, Pearson Education Publishing,* 1999, *Pg 36.*

STNG Engine RUL Prediction

Jeong Cheol Seok

Problem Statement

Data-driven techniques, especially artificially intelligence (AI) have been getting huge amount of attention in the manufacturing and production sector, due to their ability to model highly nonlinear, complex, and multi-dimensional systems. Tremendous researches of DL techniques have been started to be applied in machine health monitoring and on the Remaining Useful Life (RUL) prediction. Precise RUL prediction can significantly improve the reliability and operational safety, avoid fatal breakdown, and reduce cost overall. Thus, Cummins and its client, STNG, proposed a project on investigating the remaining useful life and predict failures in the N – timeframe of the engines that are used in the mining site. The project explores data-driven techniques for developing a RUL prediction model.

Related Work

There have been many state-of-the-art applying data driven approaches to the C-MAPSS Dataset for prediction of RUL [1].

Huang et al. [2] implemented an MLP approach for modeling the remaining useful life of the laboratory-tested bearings and received a superior performance compared to the reliability-based approaches. The ANN model developed took measurement values at present and previous as inputs and got the equipment life as percentage as output.

Recurrent neural networks (RNN) are commonly used for problems involving time series data, because of their ability to process and pass information over time. The authors of [3] used multi-layer LSTM followed by a feed forward neural network to map the features from LSTM to the predicted RUL.

Convolutional neural networks are often used for dealing time series data due to their ability to model correlations in a temporal window. [4] proposed a deep learning CNN architecture that contain multiple CNN layers followed by a fully connector layers to improve the performance.

To combine the advantages of LSTM and CNN, a hybrid architecture is proposed in [5], developing CNN layer and deep LSTM and a followed by a fully connected layer.

Classical deep learning algorithms tend to encounter the vanishing/exploding gradient problem found in ANN with gradient-based learning methods and backpropagation. To overcome the

problem of vanishing gradient problem, a new residual CNN (ResCNN) model is proposed for RUL estimation in [6]

Data

The provided Dataset from STNG contains four different categories which are as follows:

Table 7 - Provided Dataset Breakdowns

Description	Count
In-Operation Engines	105
Failed Engines	37
Passed Engines	9
Infant Mortality	7

The in-operation engines represent engines that have not yet failed and are in-operation. Hence, these engines cannot be used to predict the RUL as the target variables are currently unknown. The failed engines represent engines that have failed before the service life of 18000 hours, and passed engines represent engines that have passed or operated over the service life of 18000 hours. Lastly, the infant morality represents engines that have failed before the operational hours of 5000.

For model development, the engines that could be used are “Failed Engines” as the target variable of RUL is known only for ones that have failed.

Each sub-dataset contains total of 221 readings / features, however, most of the sensors and features such as bit signals, do not provide useful information for RUL prediction. The 24 features are selected amongst the 221 readings. The selection methods include analyzing the correlation respect to time, feature importance using classifiers such as Random Forest and creating a threshold percentage of empty values.

Proposed Methodology

The methodology followed for the project is discussed using the different aspects. The overall approach is summarized below:

a) Data Preparation

The measurement data is complex and high dimensional. There are many features that have low variance which mean that they are ineffective against the performance. Features with a dataset variance lower than threshold have been removed.

The measurement data showed great outliers such as spikes and noise. After analyzing the data, the spikes occurred randomly and does not have any relationship with the failure. Therefore, the filtering method using Z score threshold is used to eliminate the noises.

The data is then normalized as this can help the model accelerate the convergence rate. The min-max normalization is used to convert the sensor data scale within the range of [0,1].

b) Target Variable

Instead of predicting for Remaining Useful Life (RUL) directly, the model instead predicts for health index (HI) which is an RUL normalized into scale [0,1]. Since all engines show different

levels of degradation even at same remaining useful life, the health index or percentage reflects the actual state of the engine more than remaining useful life in hours. This can be formulated as follows:

$$HI_i(t) = 1 - \frac{x_i}{\max(x_i)} \quad (1)$$

c) Metrics

For evaluation of the proposed model, the commonly used, root mean squared error (RMSE) is used; however, in real life situation, the most important predictions are when the machines are at a critical state. Instead of evaluating the entire interval, the RMSE of prediction less than 0.25 and actual target variable less than 0.25 is used as evaluation. This evaluates the recall and precision for the machines at critical state of 25% health level.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(HI_{real}(t) - HI_{pred}(t))^2} \quad for\ HI_{real} < 0.25 \ \ or\ HI_{pred} < 0.25 \quad (2)$$

d) Data Split for Training

For development of the model, the dataset is initially split between train and test per engine. Afterwards, k-fold partition is applied on the train set and cross validation is applied to get the train and validation set for training.

e) Ensemble Model Method

Since the number of engine samples are relatively low, the variance error tends to be high. Depending on the test, train, and validation split, the model performance tends to fluctuate heavily. Hence, the k-fold ensemble is applied to combine the k homogeneous DL models, which help to improve prediction accuracy and reduce variance error. The ensemble is formulated as $EnsM = \frac{1}{k}\sum m_i$.

f) Model Training

For neural network algorithm, the learning rate of 0.0001, epochs of 1000 with patience of 100 is applied. Usually, due to small amount of data, the model learns before its full epoch and early stops.

Model Result and Summary

The custom RMSE of the developed ensembled models are summarized in Table 2. Most of the proposed architecture of the neural networks have been used from previous literatures.

Table 8 - Result for each proposed model

Description	RMSE
Random Forest	0.18730
Support Vector Regression	0.17230
Neural Nets [2]	0.16550
Deep Neural Nets / MLP [2]	0.15826
Stacked LSTM [3]	0.16590
DCNN [4]	0.15550
CNN – LSTM [5]	0.16190
CNN ResNET [6]	0.16506
MLP – LSTM – MLP [7]	**0.14899**

The best performing architecture is the MLP-LSTM-MLP. The LSTM networks address the issue of gradient vanishing / exploding problems; however, they do not have the capacity to handle complex feature processing. Conversely, MLP are well fitted to perform such a task. The idea of this architecture is to feed the raw inputs to the MLP before the LSTM layers. This allows the initial MLP layers to process and learn a good representation, while LSTM captures the dependencies in the time sequences. Afterwards, final MLP layer is added to predict the RUL from these temporally smoothed representations.

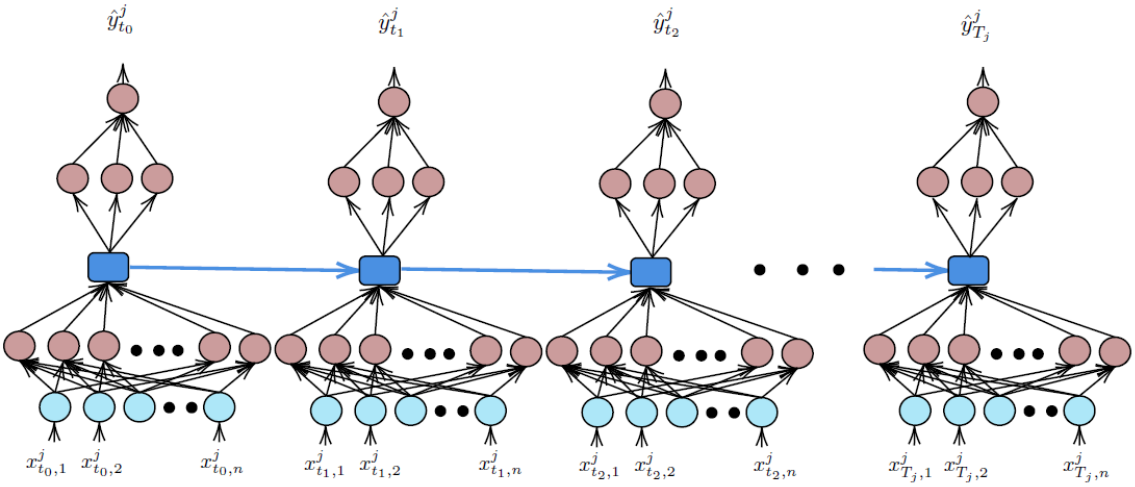


Figure 12 - Architecture of the best performing model

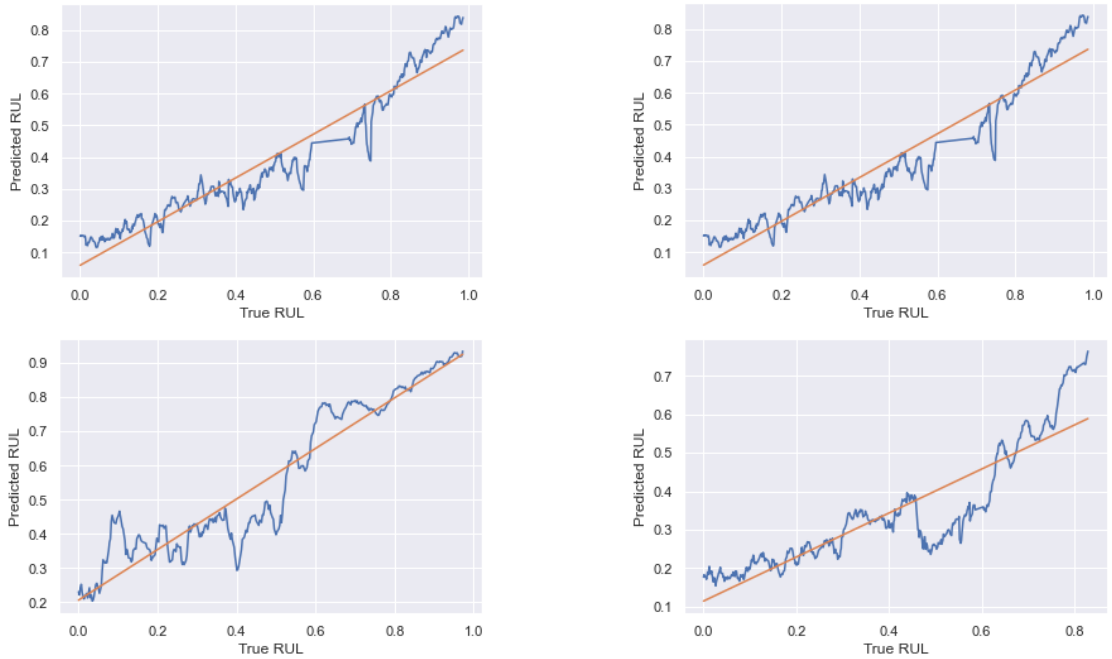


Figure 13 - Examples of RUL prediction for engine units from testing samples

Next Steps

The next step is to develop ways to interpret the predictions and provide preventive plans. The current approach that is being worked on is defining the Health index threshold where the engine is determined to be in the critical stage as illustrated in Figure 3. When the model starts predicting the engines below the critical stage, the operators are suggested to have a thorough maintenance to analyze whether the machine is at critical stage and make preventive plans accordingly.

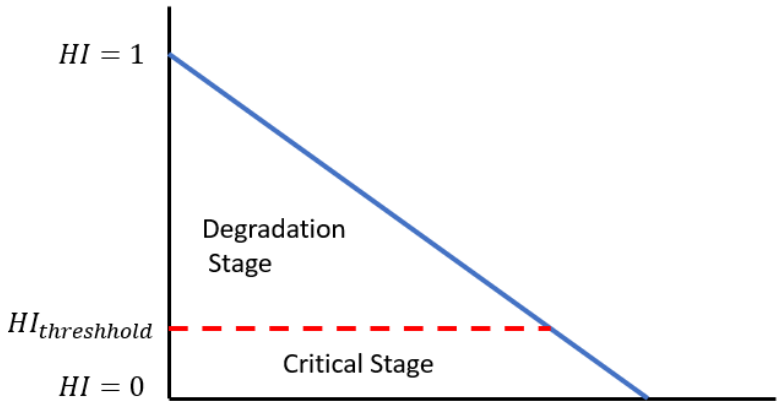


Figure 14 - Diagram of the engine state in the cycle

Ideally, we would like the HI threshold is closer to zero while achieving good performance and prediction. The steps for determining the HI thresholds are as follows:

1. Iterate through failure thresholds from 0 – 1
2. Convert the numerical health index (HI) to discrete classes using the threshold
    a. If $HI(t) \geq HI_{threshhold}$ predict $HI_{class} = 0$
    b. If $HI(t) < HI_{threshhold}$ predict $HI_{class} = 1$
3. Measure the average precision, recall, and F1 score for different critical threshold
4. E.g., $\overline{Precision} = \frac{\sum Precision_i}{N}$

References

[1] Abhinav Saxena, Kai Goebel, Don Simon, and Neil Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In 2008 international conference on prognostics and health management, pages 1–9. IEEE, 2008

[2] R. Huang, L. Xi, X. Li, C. R. Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based

on self-organizing map and back propagation neural network methods, Mechanical Systems and

Signal Processing 21 (1) (2007) 193–207

[3] Shuai Zheng, Kosta Ristovski, Ahmed Farahat, and Chetan Gupta. Long short-term memory network for remaining useful life estimation. In 2017 IEEE international conference on prognostics and health management (ICPHM), pages 88–95. IEEE, 2017

[4] Li, Xiang & Ding, Qian & Sun, Jian-Qiao, 2018. "Remaining useful life estimation in prognostics using deep convolution neural networks," Reliability Engineering and System Safety, Elsevier, vol. 172(C), pages 1-11

[5] Han Li, Wei Zhao, Yuxi Zhang, and Enrico Zio. Remaining useful life prediction using multi-scale deep
convolutional neural network. Applied Soft Computing, 89:106113, 2020.

[6] Wen L, Dong Y, Gao L. A new ensemble residual convolutional neural network for remaining useful life estimation. Math Biosci Eng. 2019 Jan 28;16(2):862-880. doi: 10.3934/mbe.2019040. PMID: 30861669.

[7] Chaoub, Alaaeddine & Voisin, Alexandre & Cerisara, Christophe & Iung, Benoît. (2021). Learning representations with end-to-end models for improved remaining useful life prognostics.

NRC combustion signature analysis for freight fires

Sophie Tian

Introduction

The Fire Safety Unit within the National Research Council Canada (NRC) is responsible for the research on increasing resilience to outdoor and infrastructure fires, developing fire safety technology, and applying computational technologies in fire safety. Currently, in response to freight transportation fire incidents, first responders survey the scene of the accident to identify potential hazardous materials present, and follows the Emergency Response Guidebook (ERG), which is a 400-page document, to determine the most appropriate mitigation strategy. Since first responders routinely confront unknown hazards that might explode or release toxins in response to ordinary fire extinguishers, the Fire Safety Unit in NRC is interested in developing an Artificial Intelligence (AI) enabled tool to provide decision making support for when an unlabeled freight is burning and the material it contains is unknown. In particular, this tool will assist first responders when they survey the scene of the accident by identifying the characteristics of the materials burning and determining the most the most appropriate mitigation strategy.

In this work, we focus on the design of machine learning (ML) models to predict what type of material is burning based on combustion signatures collected from a Fourier-transform infrared spectroscopy (FTIR). The designed model will be employed as the decision-making module within the AI tool and will provide guidance to first responders in the appropriate mitigation strategies to take in a freight fire.

Dataset

Combustion gas signatures reveal what type of material is burning and how the combustion progresses. NRC has conducted 30 tests on 20 distinct materials to collect the combustion signature data. For each test, the material is burned within a cone calorimeter and the FTIR is used to detect the chemicals emitted. The FTIR is capable of detecting the concentration profiles of up to 100 different chemicals over time, creating time series profiles. For each test, the NRC provided binary hazard labels in five categories: flammability, toxicity, water extinguishable, corrosiveness, and oxidizing. Given this dataset, the goal of the designed ML is to correctly predict the five hazard labels, taking into consideration the small-sized dataset.

Related Work

This section will outline previous works that leveraged the FTIR either for the analyses of combustion processes or for machine learning, highlighting the fact that machine learning techniques have not been applied in combustion sciences. I will then discuss the state-of-the-art algorithms in the realm of time series classification.

Fourier-transform infrared spectroscopy (FTIR) is a technique that collects the infrared absorption spectrum of solid, liquid, or gas sample. In Speitel (2002), the FTIR analyzer was studied on its ability to accurately quantify a variety of gases during material combustion, and a reasonable agreement between combustion gases measured using a FTIR analyzer and the cone nondispersive infrared (NDIR) gas analyzers was found. This demonstrates that FTIR has the capability to accurately quantify the gases evolved from burning materials.

FTIR spectroscopy has become one of the most common methods for the analysis of combustion data, among other methods such as mass spectrometry (MS), gas chromatography (GC), and thermogravimetric analysis (TGA). Sanchez et al. designed an online FTIR monitoring methodology to quantify the evolution profiles for NO,N2O and CO2 from char combustion, and found that it is especially advantageous to use FTIR for online monitoring of combustion since its running time for the evolution of the target gas was much lower than the duration of the combustion experiments (Sanchez et al., 2010). In addition to being used in isolation, FTIR and TGA have often been coupled together, named TG-FTIR, to analyze the characteristics of materials under pyrolysis and combustion. The advantage of TG-FTIR is that it allows the gases released during pyrolysis and before combustion to be analyzed (Sonnier et al., 2014; Bassilakis et al., 2001). Oudghiri et al. (2016) demonstrated that valuable qualitative information can be extracted using TG-FTIR on gaseous volatile species by analyzing marine sediment under pyrolysis and combustion. Similarly, Jiang et al. (2010) used TG-FTIR to qualitatively analyze urea-formaldehyde resin residue during pyrolysis and combustion. On the other hand, FTIR coupled with cone calorimetry is the preferred method to analyze the gaseous products evolved during a combustion process. The advantage of this method is that the cone calorimeter allows the heat release rate to be monitored in a controlled environment, and the FTIR analyzer allows the gas species emitted to be collected simultaneously. Kallonen (1990) used FTIR-cone calorimetry to measure seven gases evolved from burning PUR, PVC and wool carpet, and concluded that FTIR has comparable precision with online gas analyzers and may even outperform online specific gas analyzers due to overlapping impurities in the evolved effluents. Puente et al. (2016) leveraged this coupled technique to analyze how two types of fireproof pavements contribute to fire growth. Similarly, studies have been conducted using FTIR-cone calorimetry to analyze the thermal degradation behavior of fir wood (Batiot et al., 2014) and two types of plywood (Fateh et al., 2014).

As demonstrated by the abundance of previous works, FTIR is capable of accurately detecting and quantifying a large number of the products evolved during material combustion and has proved to be a useful tool for analyzing combustion data. In addition, since the spectral data obtained from an FTIR analyzer produces unique fingerprints of the sample materials (Titus et al., 2019), FTIR is well-suited to collect data for machine learning models where patterns and structures are expected to exist within the data. However, previous works were mainly focused on the qualitative assessment of materials under combustion, and to the best of our knowledge, only one previous work has attempted to analyze combustion data collected using FTIR in a data-driven approach. This work by Chen et al. used data from an FTIR analyzer capable of detecting the concentrations of 18 gas species to train an artificial neural network (ANN) to

classify the input data as either flaming fire, smoldering fire, or nuisance. The training data was composed of repeated tests on eight different materials or chemical compounds, and the validation data came from fire and nonfire cases of materials not observed in the training set (Chen et al., 2000). Using this dataset, the trained three-layer ANN correctly predicted 96% of the 248 test cases, demonstrating the feasibility of using FTIR data to detect fires (Chen et al., 2000). Aside from this work, there has not been any other machine learning algorithm developed to make classification decisions using FTIR analyzers on combustion processes.

Although machine learning has yet to be adopted as a candidate method in the analyses of combustion processes, it has been gradually introduced into and increasingly used in the field of fire sciences, especially in fire detection. In a comprehensive review of chemical-based indoor fire detection systems and associated algorithms, Fonollosa et al. identified that chemical gas sensors can improve fire sensitivity and early detection, but they suffer from a high rate of false alarms. They suggested that pattern recognition algorithms is the only path to improve false alarm immunity and cited numerous works that utilized machine learning techniques such as ANNs, K-nearest-neighbors (KNN) and decision trees to decrease the chances of false alarms (Fonollosa et al., 2018). However, in their discussion of gas sensors for combustion products, Fonollosa et al. only discussed the use of chemical sensor components, namely electrochemical cells, metal oxide sensors (MOX), and non-dispersive infrared cells (NDIR), for fire detection, and explicitly excluded the FTIR analyzer from this discussion without providing rationale for this choice. This suggests that there is a gap in our understanding of how FTIR analyzers compare to chemical sensor components in fire detection. These chemical gas sensors are similar to FTIR analyzers since they are both used to detect and to quantify target gases, and both take measurements over time. However, the FTIR analyzer has the advantage of being able to detect and collect the signatures of a variety of gases simultaneously, while typically an array of chemical sensor components need to be manually selected and assembled together to measure the target gases. Nevertheless, both tools produce multivariate, time series data. As a result, machine learning models that have been demonstrated to work well for gas sensor data should also work well for FTIR data. In the field of fire detection using images, Park et al. proposed a fire detection system incorporating a deep neural network for time series sensor data analysis as well as a convolutional neural network (CNN) for fire detection in images among other multi-functional components. Overall, machine learning techniques have become increasingly recognized in the area of fire detection, using either time series data or image data.

In a broader context, machine learning techniques have been widely adopted to the classification of gases using data collected from gas sensors. One particular type of gas sensor named the electronic nose (E-nose), typically developed using metal oxide semiconductor (MOS) sensors or MOX sensors, has utilized machine learning techniques quite extensively. In a review of smart gas sensing technologies, Feng et al. outlined and compared various smart gas sensor arrays, signal processing methods and gas pattern recognition algorithms including SVM, KNN and ANNs. It was highlighted that machine learning now plays a key role not only in making gas classification decisions but also in its ability to adapt for sensor drifts, which alleviates the need to manually manipulate the signals captured by the gas sensors (Feng et al., 2019). Earlier works on gas classification using E-Nose data mostly leveraged traditional machine learning techniques such as support vector machines (SVM), for example, to discriminate between two pure gases and their mixture gases (Khalaf et al., 2008), and to classify six indoor air contaminants using a hybrid approach combining an SVM and fisher linear discrimination analysis (FLDA) (Zhang et al., 2012). In recent years, as neural network proved to be universal function approximators, gas classification also began to adopt this

approach, replacing traditional approaches. For example, Raja Kumar et al. applied a five-layer ANN to detect pollutant gases measured using an E-nose, and demonstrated that its performance in terms of accuracy, sensitivity and specificity were all better than a trained SVM and a Naive Bayes classifier. However, since gas classification uses time series as input data, a more intuitive model to use is a CNN, which employs convolutional kernels to reduce the number of trainable parameters in the model and also to detect the same feature at different time points within a time series. In (Peng et al., 2018), a deep-CNN model named GasNet with up to 38 layers was proposed to classify four types of gases using data from eight MOS sensors, and this model significantly outperformed a trained SVM and ANN. In (Zhao et al., 2019), a one-dimensional deep CNN named 1D-DCNN was proposed to classify three pure gases and two binary mixture gases. The 1D-DCNN model was able to automatically extract important features from the data through convolutional kernels, and it significantly outperformed the benchmarks which are a SVM, a KNN, an ANN and a random forest (RF) (Zhao et al., 2019). Another idea is to apply existing CNN architectures such as VGG and ResNet to gas classification as demonstrated in (Han et al., 2019), however such deep architectures would require a large amount of training data, which is often difficult to obtain in gas classification. Since the field of gas classification using E-nose data gradually favored CNN architectures, it is natural to approach the classification task using combustion data measured by an FTIR via CNNs as well.

As previously discussed, gas classification typically deals with multivariate time series data. A simple way to account for the temporal information is to leverage a CNN-based architecture as demonstrated by Peng et al. (2018); Zhao et al. (2019); Han et al. (2019). This type of approach is analogous to using 2dimensional CNNs to capture the spatial relationship or objects in computer vision tasks and is much more efficient than training traditional feed-forward ANNs due to their weight sharing property. Another research area for consideration is research on the time series classification task itself, which focuses on developing algorithms leveraging the temporal information, and much work has gone into algorithmic development and comparison. A recent paper by Ruiz et al. compared state-of-the-art multivariate time series classification (MTSC) methods on 26 equal-length time series datasets, and cited a model named ROCKET, developed by Dempster et al., as the top performing method and the recommended starting point for benchmarking in future research. ROCKET leverages a large number of randomly generated convolutional kernels to capture the temporal features within the time series, and trains a linear classifier using the feature maps obtained from the random kernels to make the classification decisions Dempster et al. (2019). This method is extremely scalable and at the same time offers the same level of performance compared to other state-of-the-art methods, as shown in (Ruiz et al., 2021). Even though this is not a CNN-based method, which learns the weights of the convolutional kernels, it is evident that there is a theme of using convolutional kernels to perform classification on multivariate time series data. As a result, a CNN-based method or ROCKET could be a great starting point for our classification task of using FTIR combustion data to classify the burning materials.

Anticipated outcomes

The proposed tool addresses the safety of freight transportation by providing a solution that can be deployed easily to freight fires and by allowing data-driven decisions to be made given real-time information. It reduces the risks to first responders by identifying possible hazardous materials present based on combustion signatures collected at the scene and by providing decision-making support to first responders in determining the most effective mitigation measures.

Conclusions and future work

This report introduces the combustion signature analysis project in collaboration with the Fire Safety Unit of NRC, including the background and objectives of the project, and a literature review on related fields. Future work includes designing ML models to classify the combustion signature data and incorporating the model into the AI decision making module.

References

Bassilakis, R, R.M Carangelo, M.A Wojtowicz. 2001. Tg-ftir analysis of biomass pyrolysis. Fuel 80 1765–1786. doi:https://doi.org/10.1016/S0016-2361(01)00061-8. URL https://www.sciencedirect.com/science/article/pii/S0016236101000618.

Batiot, Benjamin, J. Luche, T. Rogaume. 2014. Thermal and chemical analysis of flammability and combustibility of fir wood in cone calorimeter coupled to ftir apparatus. Fire and Materials 38 418–431.

Chen, Y., Michael Serio, Sandeep Sathyamoorthy. 2000. Development of a fire detection system using ft-ir spectroscopy and artificial neural networks. Fire Safety Science 6 791–802. doi:10.3801/IAFSS.FSS.6-791.

Dempster, Angus, Franc¸ois Petitjean, Geoffrey I. Webb. 2019. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. CoRR abs/1910.13051. URL http://arxiv.org/abs/1910.13051.

Fateh, Talal, Thomas Rogaume, Jocelyn Luche, Franck Richard, Florent Jabouille. 2014. Characterization of the thermal decomposition of two kinds of plywood with a cone calorimeter – ftir apparatus. Journal of Analytical and Applied Pyrolysis 107 87–100. doi:https://doi.org/10.1016/j.jaap.2014.02.008. URL https://www.sciencedirect.com/science/article/pii/S0165237014000321.

Feng, Shaobin, Fadi Farha, Qingjuan Li, Yueliang Wan, Yang Xu, Tao Zhang, Huansheng Ning. 2019. Review on smart gas sensing technology. Sensors 19. doi:10.3390/s19173760. URL https://www.mdpi.com/1424-8220/19/17/3760.

Fonollosa, Jordi, Ana Sol´orzano, Santiago Marco. 2018. Chemical sensor systems and associated algorithms for fire detection: A review. Sensors 18. doi:10.3390/s18020553. URL https://www.mdpi.com/1424-8220/18/2/553.

Han, Lu, Chongchong Yu, Kaitai Xiao, Xia Zhao. 2019. A new method of mixed gas identification based on a convolutional neural network for time series classification. Sensors 19. doi:10.3390/s19091960. URL https://www.mdpi.com/1424-8220/19/9/1960.

Jiang, Xuguang, Chunyu Li, Yong Chi, Jianhua Yan. 2010. Tg-ftir study on ureaformaldehyde resin residue during pyrolysis and combustion. Journal of Hazardous Materials 173 205–210. doi:https://doi.org/10.1016/j.jhazmat.2009.08.070. URL https://www.sciencedirect.com/science/article/pii/S0304389409013508.

Kallonen, Raija. 1990. Smoke gas analysis by ftir method. preliminary investigation. Journal of Fire Sciences 8 343–360. doi:10.1177/073490419000800503. URL https://doi.org/10.1177/073490419000800503.

Khalaf, Walaa, Calogero Pace, Manlio Gaudioso. 2008. Gas detection via machine learning .

Oudghiri, Fatiha, Nabil Allali, Jos´e Mar´ıa Quiroga, Mar´ıa Roc´ıo Rodr´ıguez-Barroso. 2016. Tg–ftir analysis on pyrolysis and combustion of marine sediment. Infrared Physics Technology 78 268–274. doi:https://doi.org/10.1016/j.infrared.2016.08.015. URL https://www.sciencedirect.com/science/article/pii/S1350449516302870.

Park, Jun Hong, Seunggi Lee, Seongjin Yun, Hanjin Kim, Won-Tae Kim. 2019. Dependable fire detection system with multifunctional artificial intelligence framework. Sensors 19. doi:10.3390/s19092025. URL https://www.mdpi.com/1424-8220/19/9/2025.

Peng, Pai, Xiaojin Zhao, Xiaofang Pan, Wenbin Ye. 2018. Gas classification using deep convolutional neural networks. Sensors 18. doi:10.3390/s18010157. URL https://www.mdpi.com/1424-8220/18/1/157.

Puente, E., D. L´azaro, D. Alvear. 2016. Study of tunnel pavements behaviour in fire by using coupled cone calorimeter – ftir analysis. Fire Safety Journal 81 1–7. doi:https://doi.org/10.1016/j.firesaf.2016.01.010.

Raja Kumar, Jambi Ratna, Rahul K. Pandey, Biplab K. Sarkar. 2019. Pollutant gases detection using the machine learning on benchmark research datasets. Procedia Computer Science 152 360–366. doi:https://doi.org/10.1016/j.procs.2019.05.005. URL https://www.sciencedirect.com/science/article/pii/S1877050919306568. International Conference on Pervasive Computing Advances and Applications- PerCAA 2019.

Ruiz, Alejandro Pasos, Michael Flynn, James Large, Matthew Middlehurst, Anthony Bagnall. 2021. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery 35 401–449.

Sanchez, Astrid, Eric Eddings, Fanor Mondragon. 2010. Fourier transform infrared (ftir) online monitoring of no, n2o, and co2 during oxygen-enriched combustion of carbonaceous materials. Energy Fuels - ENERG FUEL 24. doi:10.1021/ef100575v.

Sonnier, Rodolphe, Ga¨elle Dorez, Henri Vahabi, Claire Longuet, Laurent Ferry. 2014. Ftir–pcfc coupling: A new method for studying the combustion of polymers. Combustion and Flame 161 1398–1407. doi:https://doi.org/10.1016/j.combustflame.2013.11.011. URL https://www.sciencedirect.com/science/article/pii/S0010218013004240.

Speitel, Louise C. 2002. Fourier transform infrared analysis of combustion gases. Journal of Fire Sciences 20 349–371. doi:10.1177/0734904102020005484. URL https://doi.org/10.1177/0734904102020005484.

Titus, Deena, E. James Jebaseelan Samuel, Selvaraj Mohana Roopan. 2019. Chapter 12 nanoparticle characterization techniques. Ashutosh Kumar Shukla, Siavash Iravani, eds., Green Synthesis, Characterization and Applications of Nanoparticles. Micro and Nano Technologies, Elsevier, 303–319. doi:https://doi.org/10.1016/B978-0-08-102579-6.00012-5. URL https://www.sciencedirect.com/science/article/pii/B9780081025796000125.

Zhang, Lei, Fengchun Tian, Hong Nie, Lijun Dang, Guorui Li, Qi Ye, Chaibou Kadri. 2012. Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine. Sensors and Actuators B: Chemical 174 114–125. doi:https://doi.org/10.1016/j.snb.2012.07.021. URL https://www.sciencedirect.com/science/article/pii/S0925400512007010.

Zhao, X., Z. Wen, X. Pan, W. Ye, A. Bermak. 2019. Mixture gases classification based on multi-label one-dimensional deep convolutional neural network. IEEE Access 7 12630–12637. doi:

10.1109/ACCESS.2019.2892754.

Machine Learning for Process Monitoring and Control in Additive Manufacturing

Katie Xu

Introduction

Additive manufacturing, also known as 3D printing, are fabrication processes where 3D objects are built up layer-by-layer. Fused deposition modeling (FDM) is a type of additive manufacturing where layers are formed by melting a thermoplastic material and depositing it in the desired locations where it solidifies and becomes part of the object. The goal of this project is to develop a closed loop system to monitor and control the quality of parts made using a FDM 3D printer. This project is in collaboration with Professor Zou and his students from the Material Science and Engineering department.

Additive manufacturing processes have important advantages for applications in industrial manufacturing. In particular, the ability to make objects with complex geometries which may be difficult or impossible to make with other processes, reduced waste material compared to traditional machining, and lower tooling costs compared to casting and moulding processes. These properties are advantageous for creating complex or custom parts that are needed in relatively small quantities. Unfortunately, there remain challenges which limit the current practicality of additive manufacturing in industry. Parameter search is time consuming because there is a large number of parameters to be found and it is generally done through trial and error. Moreover, optimal parameter values depend on factors such as the material being used and the part geometry, so in the worst case the parameter search process must be repeated for each unique part. Additionally, even with a good set of parameters, external disturbances can result in inconsistent part quality and high rejection rates [1]. Thus, there is a need for closed-loop process monitoring and parameter adjustment. This document summarizes recent work towards developing an approach to building such a system.

System overview

The overall goal is to build a closed-loop FDM printing system as outlined in Figure 15. To summarize, the proposed system would function as follows. First, the user would initiate the process by setting the initial parameters. Then, after printing each layer an image of the layer would be captured to observe the progress of the process. Porosity and surface roughness of the newly printed layer would be estimated the images, and used to predict the tensile strength of

the final part. Finally, process parameters would be adjusted based on the observed porosity/surface roughness and predicted strength.

Porosity and surface roughness were chosen as indicators of tensile strength because porosity measures the existence of voids within a layer and surface roughness is related to the formation of voids between layers. The existence of such internal voids can be expected to weaken the printed part because it reduces the load bearing area and creates areas of high stress concentration.

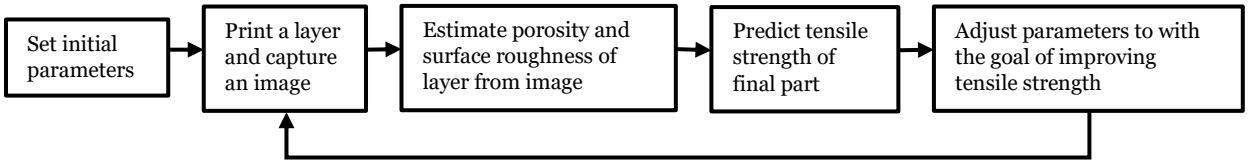


Figure 15: System Overview

Based on this outline, the three main modules of the system which need to be developed (excluding the physical system) are:

1. Image analysis
    a. From an image, estimate the porosity of the layer
    b. From an image, estimate the surface roughness of the layer
2. Tensile strength model
    a. From porosity and surface roughness estimates, predict the tensile strength of the final part
3. Controller
    a. Based on porosity, surface roughness, and predicted tensile strength, choose appropriate parameter adjustments to improve tensile strength of the final part

Image analysis

Porosity

Porosity in the context of this project can be defined as the percentage area of a layer which is taken up by pores (unintentional voids or gaps between adjacent lines of extruded filament). Given a top-down image of a layer, it is possible to see such voids. Quantifying the porosity then can be done by segmenting the image into sections which are pores and sections which are not, and then computing the ratio of the areas. Thus, porosity estimation can be addressed as a semantic segmentation problem.

UNet and DeepLab v3 are two state of the art convolutional neural network (CNN) architectures designed for semantic segmentation. UNet [2] is characterized by convolutional layers which reduce then increase the size of the feature map, and skip connections between earlier and later layers to preserve fine-grained feature information. DeepLab v3 [3] is the third version of their model which is characterized by the use of atrous convolutions to explicitly control the scale of the features being extracted.

CNNs like UNet and DeepLab v3 generate pixel-wise predictions without explicit spatial constraints, and thus the predictions can be noisy as a result. Conditional random fields (CRF) are undirected graphs which can be used to encode spatial relationships and long-range interactions between pixels. They can be applied on top a CNN or other classifier to improve the final predictions. This was used in previous versions of the DeepLab model for improved results

[4, 5]. However, a major drawback which limits the practicality of CRFs is inefficient inference. This must be done iteratively, unlike the straight-forward forward pass of a CNN.

Texture analysis
Surface roughness can be quantified measuring the profile of the surface and computing statistics such as average deviation from the mean or a similar quantity. However, measuring the surface profile can be difficult and requires specialized equipment. As a result, it may not be practical to collect enough labelled data to successfully train a deep CNN on raw image data. Instead, a possible approach is to compute low-level texture features from images to use as proxies for surface roughness. For example, the variance or entropy of pixel intensities present in the image provides information on the variability of pixel intensities. Statistics computed from a grey level co-occurrence matrix can provide information on positional relationships between pixels of similar grey values. Other features can provide information on the frequency content or geometric properties. [6]

Tensile strength model
This is the topic of an ongoing project with two undergraduate summer students. The goal is to characterize the relationship between the porosity/surface roughness of individual layers and the tensile strength of the final part. That is, we want to build a predictive model from porosity/surface roughness to tensile strength, and to answer the following research questions:

- How strong is the correlation to tensile strength for each of porosity and surface roughness?
- In order to predict tensile strength, it necessary to observe both? Is it sufficient?
- Are porosity and/or roughness of certain layers more influential than others?

Controller
The planned approach for developing a controller to close the loop is as follows:

1. Build a virtual model of the 3D printer
2. Use virtual model to train an agent using reinforcement learning
3. Transfer to physical system

The specific details of how to accomplish these steps will be addressed moving forward.

References

[1] J. Fox, F. Lopez, B. Lane, H. Yeung, and S. Grantham, "On the requirements for model-based thermal control of melt pool geometry in laser powder bed fusion additive manufacturing," in *Proceedings of the 2016 Material Science & Technology Conference*, 10 2016. [Online]. Available: https://www.nist.gov/publications/requirements-model-based-thermal-control-melt-pool-geometry-laser-powder-bed-fusion

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[3] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017. [Online]. Available: https://arxiv.org/abs/1706.055872

[4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," 2016. [Online]. Available: https://arxiv.org/abs/1412.7062

[5] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018.

[6] L. Armi and S. Fekri-Ershad, "Texture image analysis and texture classification methods - a review,"2019. [Online]. Available: https://arxiv.org/abs/1904.065542