Ref. Ares(2021)5961456 - 30/09/2021

European Commission

ECOLE

Experience-based Computation:
Learning to Optimise

Project Number: 766186
Project Acronym: ECOLE
Project title: Experienced-based Computation: Learning to Optimise

Deliverable D4.3

Research and Personal Skill Report, Final

Authors:
Shuo Wang, Leandro Minku, Xin Yao - University of Birmingham
Thomas Bäck - Leiden University
Stefan Menzel, Bernhard Sendhoff – HRI-EU
Zhao Xu – NEC Laboratories Europe

Project Coordinator: Professor Xin Yao, University of Birmingham
Beneficiaries: Leiden University, Honda Research Institute Europe, NEC Laboratories Europe

H2020 MSCA-ITN
Date of the report: 30.09.2021

Table of Contents

Introduction **3**
Research Progress **3**
2.1 Publications 3
2.2 PhD Progress and Process 8
Work Packages **9**
Thesis Plan and Career Development **17**
4.1 Training Plan 17
4.2 Outreach 18
4.3 Career Development Plans 19
4.3.1 Thiago Rios 19
4.3.2 Sneha Saha 20
4.3.3 Sibghat Ullah 22
4.3.4 Duc Anh Nguyen 23
4.3.5 Jiawen Kong 24
4.3.6 Stephen Friess 26
4.3.7 Gan Ruan 27
4.3.8 Giuseppe Serra 29
4.4 ECOLE Summer schools and workshops 30
Professional Skill Development **33**
ECOLE Management **34**
6.1 Covid 34
6.2 Updates secondment plan 34
Appendix A- PhD Thesis Outlines **37**
1 Thiago Rios (ESR1): Learning-based Representations of High-dimensional CAE Models for Automotive Design Optimization 37
2 Sneha Saha (ESR2): Multi-Criteria Preference-Aware Design Optimization of 3D Designs 38
3 Sibghat Ullah (ESR3): Model-Assisted Robust Optimisation for Continuous Black-Box Problems 39
4 Duc Anh Nguyen (ESR4): Automatic algorithm configuration for parameter tuning of modelling and optimization algorithms 40
5 Jiawen Kong (ESR5): Class imbalance classification through semi-supervised and active learning for experience-based optimisation 41
6 Stephen Friess (ESR6): Predicting Inductive Biases for Search-based Optimization 43
7 Gan Ruan (ESR7): Knowledge transfer in evolutionary dynamic multi-objective optimization 44
8 Giuseppe Serra (ESR8): Explainable models for preference learning with user-generated data 45

1. Introduction

This “Research and Personal Skill Report” provides a comprehensive summary of the different activities implemented and executed covering the ECOLE project time from 04/18 until 09/21 (42 months). On the one hand, we report on our project achievements for the individual work packages related to research along with our consortium publications and outreach activities. On the other hand, we reflect on provided courses, workshops, secondments and company events, which educated and supported each ESR to advance their career by carrying out excellent research and building their paths towards successful PhDs, even under the current challenges of Covid.

2. Research Progress

2.1 Publications

The management team have defined a publication procedure for approval of manuscripts prior to the submission of the journal. This allows all beneficiaries to comment on published work in addition to ensuring that any potential intellectual property is not disclosed prior to protection. The detailed publication process and example of consent for publication were reported previously in our mid-term report and will not repeated here.

A number of research papers have been published or submitted from the project team of ESRs and supervisors. The published papers have appeared at a wide range of venues, including conferences/journals in evolutionary computation, neural networks, computational intelligence, machine learning, data mining and application areas. These papers have shown good potentials in generating significant impact on the international research community. For example, two ESRs’ papers, by Gan Ruan and Stephen Friess, respectively, were short-listed for the best student paper awards at the 2020 IEEE International Conference on Evolutionary Computation (CEC’2020). One paper by Thiago Rios (ESR) has already been published by the flagship journal in the field of evolutionary computation --- IEEE Transactions on Evolutionary Computation. More details about the list of our publications can be found in Table 2 below and in the deliverable D5.4.

Table 1 – Publications

ESR Number	Research Output
1	Published: [1] T. Rios, P. Wollstadt, B. van Stein, T. Bäck, Z. Xu, B. Sendhoff and S. Menzel, "Scalability of Learning Tasks on 3D CAE Models Using Point Cloud Autoencoders," *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1367-1374, 2019. [2] T. Rios, B. Sendhoff, S. Menzel, T. Bäck and B. van Stein, "On the Efficiency of a Point Cloud Autoencoder as a Geometric Representation for Shape Optimization," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 791-798, 2019. [3] T. Rios, B. van Stein, S. Menzel, T. Back, B. Sendhoff and P. Wollstadt, "Feature Visualization for 3D Point Cloud Autoencoders," in *International Joint Conference on Neural Networks (IJCNN)*, pp. 1-9, 2020. [4] T. Rios, J. Kong, B. van Stein, T. Bäck, P. Wollstadt, B. Sendhoff and S. Menzel, "Back to Meshes: Optimal Simulation-ready Mesh Prototypes for Autoencoder-based 3D Car Point Clouds," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 942-949, 2020. [5] T. Rios, B. van Stein, P. Wollstadt, T. Bäck, B. Sendhoff and S. Menzel, "Exploiting Local Geometric Features in Vehicle Design Optimization with 3D Point Cloud Autoencoders," in *IEEE Congress on Evolutionary Computation (CEC)*, pp. 514-521, 2021. [6] T. Rios, B. van Stein, T. Bäck, B. Sendhoff and S. Menzel, "Multi-Task Shape Optimization Using a 3D Point Cloud Autoencoder as Unified Representation," in IEEE Transactions on Evolutionary Computation, 2021 (Early Access). Submitted: [7] T. Rios, B. van Stein, T. Bäck, B. Sendhoff and S. Menzel, "Point2FFD: Learning Shape Representations of Simulation-ready 3D Models for Engineering Design Optimization," in *International Conference on 3D Vision*, 2021.
2	Published: [1] S. Saha, T. Rios, L. L. Minku, X. Yao, Z. Xu, B. Sendhoff, and S. Menzel, "Optimal Evolutionary Optimization Hyper-parameters to Mimic Human User Behavior", in *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 858-866, 2019.

	[2] S. Saha, T. Rios, S. Menzel, B. Sendhoff, T. Bäck, X. Yao and P. Wollstadt, "Learning Time-Series Data of Industrial Design Optimization using Recurrent Neural Networks", in *International Conference on Data Mining Workshops (ICDMW)*, pp. 785-792, 2019. [3] S. Saha, S. Menzel, L.L. Minku, X. Yao, B. Sendhoff, and P. Wollstadt, "Quantifying the Generative Capabilities of Variational Autoencoders for 3D Car Point Clouds," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1469-1477, 2020. [4] S. Saha, L. L. Minku, X. Yao, B. Sendhoff, and S. Menzel, "Exploiting Linear Interpolation of Variational Autoencoders for Satisfying Preferences in Evolutionary Design Optimization," in *IEEE Congress on Evolutionary Computation (CEC)*, pp. 1767-1776, 2021. Submitted: [5] S. Saha, T. Rios, L. L. Minku, B. v. Stein, P. Wollstadt, X. Yao, T. Baeck, B. Sendhoff and S. Menzel, "Exploiting Generative Models for Performance Predictions of 3D Car Designs," in *IEEE Symposium Series on Computational Intelligence (SSCI), 2021*.
3	Published: [1] S. Ullah, H. Wang, S. Menzel, B. Sendhoff and T. Bäck, "An Empirical Comparison of Meta-Modeling Techniques for Robust Design Optimization", in *IEEE Symposium Series on Computational Intelligence (SSCI)*, Xiamen, China, 6-9 December 2019. [2] S. Ullah, Z. Xu, H. Wang, S. Menzel, B. Sendhoff and T. Bäck, "Exploring Clinical Time Series Forecasting with Meta-Features in Variational Recurrent Models," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, pp.19-24, July 2020. [3] S. Ullah, D.A. Nguyen, H. Wang, S. Menzel, B. Sendhoff and T. Bäck, "Exploring Dimensionality Reduction Techniques for Efficient Surrogate-Assisted Optimization," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, Australia, pp.1-4, December 2020. [4] S. Ullah, H. Wang, S. Menzel, B. Sendhoff and T. Bäck, "A New Acquisition Function for Robust Bayesian Optimization of Unconstrained Problems," In *Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion)*, July 10–14, Lille, France, 2021.
4	Published: [1] D.A. Nguyen, J. Kong, H. Wang, S. Menzel, B. Sendhoff, A.V. Kononova and T. Bäck, "Improved Automated CASH Optimization with Tree Parzen Estimators for Class Imbalance Problems," in *the 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2021. Submitted:

	[2] D.A. Nguyen, A.V. Kononova, S. Menzel, B. Sendhoff and T. Bäck, "Efficient AutoML via combinational sampling," *IEEE Symposium Series on Computational Intelligence (IEEE SSCI)*, 2021
5	Published: [1] J. Kong, W. Kowalczyk, D.A. Nguyen, T. Bäck and S. Menzel, "Hyperparameter optimisation for improving classification under class imbalance," In *IEEE symposium series on computational intelligence (SSCI)*, pp. 3072-3078, 2019. [2] J. Kong, T. Rios, W. Kowalczyk, S. Menzel and T. Bäck, "On the performance of oversampling techniques for class imbalance problems," *Advances in Knowledge Discovery and Data Mining*, 12085, pp.84-96, 2020 [3] J. Kong, W. Kowalczyk, S. Menzel and T. Bäck, "Improving Imbalanced Classification by Anomaly Detection," In *International Conference on Parallel Problem Solving from Nature*, pp. 512-523, Springer, Cham, 2020. Submitted: [4] J. Kong, T. Rios, W. Kowalczyk, S. Menzel and T. Bäck, "On the performance of oversampling techniques for class imbalance problems," *International Journal of Data Science and Analytics*, 2021.
6	Published: [1] S. Friess, P. Tiňo, Z. Xu, S. Menzel, B. Sendhoff and X. Yao, "Artificial Neural Networks as Feature Extractors in Continuous Evolutionary Optimization," In *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2021. [2] S. Friess, P. Tiňo, S. Menzel, B. Sendhoff and X. Yao, "Improving sampling in evolution strategies through mixture-based distributions built from past problem instances," In *International Conference on Parallel Problem Solving from Nature (pp. 583-596),* Springer, Cham, 2020. [3] S. Friess, P. Tiňo, S. Menzel, B. Sendhoff and X. Yao, "Representing experience in continuous evolutionary optimisation through problem-tailored search operators," In *2020 IEEE Congress on Evolutionary Computation (CEC) (pp. 1-7)*, 2020. [4] S. Friess, P. Tiňo, S. Menzel, B. Sendhoff and X. Yao, "Learning transferable variation operators in a continuous genetic algorithm," In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 2027-2033, 2019. Submitted: [5] S. Friess, P. Tiňo, S. Menzel, B. Sendhoff and X. Yao, "Preliminary: Improving Evolutionary Optimization through Prediction of Inductive Biases with Applications to Shape Optimization," In *2021 In IEEE Symposium Series on Computational Intelligence (SSCI)* [submitted], IEEE.

7	Published: [1] G. Ruan, L. L. Minku, S. Menzel, B. Sendhoff, and X. Yao, "When and how to transfer knowledge in dynamic multi-objective optimization," in *IEEE Symposium Series on Computational Intelligence*, Xiamen, 2019. [2] G. Ruan, L. L. Minku, S. Menzel, B. Sendhoff, and X. Yao, "Computational Study on Effectiveness of Knowledge Transfer in Dynamic Multi-objective Optimization," in *IEEE Congress on Evolutionary Computation*, Glasgow, 2020.
8	Published: [1] G. Serra, Z. Xu, M. Niepert, C. Lawrence, P. Tino and X.Yao, "Interpreting Node Embedding with Text-labeled Graphs," *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021. [2] Z. Xu, D. Onoro-Rubio, G. Serra, M. Niepert, "Learning Sparsity of Representations with Discrete Latent Variables," *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021. Submitted: [3] G. Serra, P. Tino, Z. Xu, and X.Yao, "Product Rating Prediction through Interpretable Latent Class Modeling of User Reviews," *IEEE Trans. Neural Networks and Learning Systems*, 2021.

2.2 PhD Progress and Process

The progress of the ESRs' advances towards their PhDs has been closely assessed and monitored throughout the whole project time. In monthly management meetings the progress is evaluated and checked on potential obstacles. Besides the weekly supervision meetings with each ESR, there are further (formal) confirmation meetings in place at the different beneficiaries. In addition, all ESRs are integrated into the HRI-EU European Graduate Network (HRI-EU EGN), which fosters the exchange of research ideas and discussions between all PhD-students within this network and organizes every 6 months formal meetings where each ESR reports the current project state and agrees together with the responsible supervisors on the plans towards the successful completion of their PhD.
With respect to the finalization of the ESRs' theses, each one has been asked to provide a draft of the thesis outline starting End of 2020. This draft has been reviewed by ECOLE supervisors and feedback has been given to the ESRs for improvement. Revised versions have been completed by the ESRs, which currently serve as the basis for the discussions on the thesis' content and which are constantly updated during the process of writing up.

From a formal perspective, ESRs need to follow these processes at their host university for successfully receiving their PhD degrees:

Leiden University:
The PhD candidate conducts independent research and writes a manuscript during the PhD research time period. The manuscript is, after finishing, finally submitted to the supervisors.
After the submission of the manuscript, supervisors have six weeks to decide whether they approve the dissertation and will notify the Dean whether they approve. After approval, the PhD candidate will submit a maximum of twelve propositions to the supervisor.
Within three weeks after approval, a Doctorate Committee is established. The Doctorate Committee will receive the dissertation and each member will assess within six weeks after receiving the dissertation whether the PhD candidate is allowed to defend the dissertation. The individual assessments and suggestions are send back to the supervisor, who may advise the PhD candidate to adopt those suggestions.
After the decision of the Doctorate Committee that the PhD candidate may defend the dissertation, the PhD candidate applies to the beadle for a public defence. The defence lasts for 45 minutes, after which the Examining Committee withdraws to decide whether the doctorate is awarded.
The full details of the PhD regulations are provided online:
https://www.organisatiegids.universiteitleiden.nl/binaries/content/assets/ul2staff/onderzoek/promoveren/phd-regulations-2021-definitive.pdf

University of Birmingham:
All ESRs registered at the University of Birmingham (UoB) for their PhDs follow the UoB regulations for training PhD students. UoB has a thesis group set up for every PhD candidate. The thesis group members include all supervisors and at least one non-supervisor. The group meets roughly every six months with the student, when the student has to submit a formal written progress report as well as an oral presentation of his/her work. The student is quizzed about his/her research, research progress, research plan, skill training and future plan. After each thesis group meeting, a formal report form will be generated and submitted to the departmental RSMG (Research Student Monitoring Group), which will examine and discuss each report and take appropriate actions according to the feedback in the form. In addition to thesis group meetings, each student is required to meet his/her supervisor at least once a month, although almost all supervisors have weekly meetings with their students. One of these

meetings needs to be documented electronically every month to monitor the student's progress and discover any potential problems early.
A formal and complete description of the degree regulation can be found from:
https://intranet.birmingham.ac.uk/as/studentservices/graduateschool/rsa/progressreviews.aspx.
All current thesis outlines are attached as Appendix A.

3. Work packages

Table 2 - Update on Work Packages

WP1 – Experience-guided Optimisation in Automotive Product Design **Honda Research Institute Europe**
In WP1, we develop techniques to accumulate, explore and exploit experience across a series of automotive-like product development cycles. We envision a cooperative design system (CDS) that assists designers and engineers in the informed exploration of design ideas in the early development stage by enriching the information space with engineering performance measures. Specific questions are as follows: Firstly, with the design of efficient design representations, which we research and propose based on the latent parameters of a geometric deep learning framework. Here, current machine learning algorithms offer a promising approach to be part of a CDS, namely by using deep-learning models like autoencoders (AEs) and variational autoencoder (VAEs) to compress input data into low-dimensional latent representations in an unsupervised fashion. Secondly, we extract and utilize user design preferences in a multi-criteria optimization framework, where the criteria relate e.g. to aesthetics and aerodynamics. As a starting point for our research, we implemented a design modification technique based on free-form deformation as a state-of-the-art method used in industry for baseline comparisons and an aerodynamic simulation set-up for vehicle geometries given as triangulated meshes using OpenFOAM (https://www.openfoam.com). Both implementations have been made available to the other researchers of the ECOLE team. In real world, virtual prototyping enables designers to generate feasible solutions for a given task through optimization, which combines a search strategy for design parameter modifications with the simulation of the 3D shape's performance. However, 3D design optimization requires a flexible, yet compact design parameterization to create reasonable shape variation without losing algorithmic efficiency. Thus, we strive for supporting the engineering process to generate efficient representations by uncovering experience hidden in large historic design data sets. This is important since engineers traditionally handcraft design features for shape optimization problems based on their acquired experience only. In WP1, we advanced current geometric deep learning techniques, in particular 3D point cloud (variational) autoencoders for vehicle design data, to learn novel data-driven representations, the so-called latent space, directly from computer aided engineering (CAE) data, which potentially embeds design expertise from various users in a unified database. Furthermore, since the proposed autoencoders learn exclusively on geometric data, the features in the latent space are domain-agnostic and enable the transfer of design features between solutions assigned to different optimization problems. The methods utilized in our research exploit engineering experience both in online (during actual optimization) and offline (from past optimizations) fashions. Here, our proposed point cloud autoencoders (PC-AE) learn design representations from static sets of benchmark CAE data, which reflect the practical use case where large design data has been generated by designers in previous product development processes. Hence, the shape-generative capability of the networks,

which we see as a notion of design experience, is related to the diversity of geometric features in the data [Rios, SSCI2019a]. Yet, compared to other suitable data-driven representations of geometric data, autoencoders provide a more diverse set of degrees of freedom for design optimization and an information-richer design space [Rios, CEC2021]. To improve our understanding on the information learned by the network, we proposed a novel technique to visualize the features of 3D PC-AEs [Rios, IJCNN2020], which also allowed us to select specific degrees of freedom to modify and optimize 3D shapes [Rios, TEC2021]. For exploring experience in an online fashion, we utilized the latent features of our proposed PC-AE as search space in multi-task design optimization problems. We demonstrated that the task-agnosticism of the autoencoder features allows us to describe populations of solutions assigned to different tasks in a single space, where we can transfer knowledge between individuals during the optimization and accelerate simultaneously the solution of multiple tasks [Rios, TEC2021]. With respect to the utilization of shape generative models, we further developed and proposed a point cloud variational autoencoder (PC-VAE) for vehicle design data, which has a different notion of the latent space to foster the generative character of the model. We evaluated and compared our proposed network with other popular existing generative models and verified its capability to generate diverse and realistic car shapes for engineering design assistance [Saha, SSCI2020]. This method allowed easy interpolation and extrapolation of new designs, which otherwise is challenging to perform by user manipulation of the 3D shapes.

As a second aspect in WP1.2, we applied our proposed methods in industrial optimization scenarios. We first assessed the scalability of our proposed PC-AE to higher-dimensional CAE models, where we ensure the compatibility of the methods to typical sizes of engineering models [Rios, SSCI2019b]. Based on experiments with target shape matching optimization [Rios, SSCI2019a] and simplified vehicle aerodynamic optimization experiments [Rios, CEC2021], [Rios, TEC2021], we proposed a novel deep-generative model called Point2FFD [Rios, ICV2021], which generates geometric models by deforming simulation-ready mesh templates with free-form deformation (FFD). Hence, besides learning design features from benchmark data, the network exploits the engineering expertise embedded in the mesh templates that were manually parameterized. We show that Point2FFD generates meshes with higher quality than post-processed autoencoder-based 3D point clouds and improves the quality of real-world like vehicle aerodynamic optimization results [Rios, ICV2021].

Further, for WP1.2 which focuses on online/offline adaptation of problem representation utilizing experience, we applied our trained PC-VAE for offline adaptation of prior knowledge for optimization and surrogate modeling. In the automotive domain, vehicle aerodynamic or structural simulations are often very time-consuming, leading to the need for compact design representations to learn surrogates to replace the costly function evaluations. In our research, we evaluated the feasibility of the low-dimensional latent space of PC-VAEs for multi-objective optimization and utilized an interpolation strategy on the learned latent space to propose an improved optimization framework [Saha, CEC2021].

Further, in a vehicle aerodynamic optimization scenario, we verified the property of the PC-(V)AEs to build effective surrogate models to replace costly optimization. Here, we exploited the generative capability of the PC-VAEs not only to generate new realistic designs but also to address the data limitation in the real-world surrogate building. In a next step, we will explore the coupling of our user modeling strategies [Saha, ICDMW2019], [Saha, SSCI2019] in the PC-(V)AEs based optimization frameworks. In summary, we proposed deep learning-based PC-(V)AEs in the context of vehicle development and demonstrated their capabilities for optimization and surrogate modelling to support the automotive design process.

Our software is available online at https://github.com/ECOLE-ITN/GDL4DesignApps for the global research community.

Update on deliverables: In WP1, three deliverables have been successfully achieved. In D1.1, we reported on the generation of the shape deformation-based method which usually are applied in simulation-based industrial design optimizations. The deliverable contains information on the implemented shape morphing method, the aerodynamic simulation set-up for vehicle designs and collected data sets available for class imbalance problems. All parts have been made available to the ECOLE project for other ESRs. In D1.2, we focused on the point cloud (variational) autoencoders for generating representations for vehicle data. The deliverable provides details on the characteristics and advantages of these methods as well as the implementations. D1.3 focuses on our proposed frameworks for vehicle multi-objective/multi-task design optimization based on our autoencoders to demonstrate their effectiveness. In addition, it provides details on the implementations so that other researchers fully utilize our online available source codes.

WP2 – Learning in Robust and Efficient Multi-Objective Optimisation
Leiden

In WP2, the aim is to develop, analyse and evaluate methods to solve expensive-to-evaluate black-box problems with high dimensionality, multiple conflicting objectives and uncertainties and noise. These are frequently-encountered obstacles in real-world applications of continuous optimisation, e.g., automobile manufacturing, building construction and steel production. To gain efficiency in optimisation, this work package emphasizes on the so-called “Surrogate-Modeling” (SM) approach which substitutes the evaluation of expensive-to-evaluate function by the prediction of statistical/machine learning based models such as Kriging and Polynomials. Based on historical mainstays and recent developments in black-box optimisation, we deal with four principal issues in this WP, namely Robust Optimisation (RO) – Optimisation in the face of uncertainty and noise, Dynamic Multi-Objective Optimisation (DMOO), High-dimensional Optimisation, and Hyperparameter Optimisation (HPO) for machine learning algorithms.

Since surrogate models were initially proposed to find the nominal solution of a black-box problem, their validity to find a robust solution – a solution whose performance is not greatly affected by the uncertainty in the optimisation setup – needs further empirical evidence. To this end, a comprehensive study was conducted in WP2 which evaluates and compares the performance of surrogate models based on six modeling techniques, three noise levels, two robustness criteria, and three distinct settings of dimensionality [Ullah, SSCI 2019]. The total number of test cases in the study is 1440, and the surrogate models are appraised based on the criteria of modeling accuracy and quality of the robust solution. The key findings from this study indicate the suitability of Kriging and Polynomials for model-assisted RO.

To address the issue of uncertainty, the famous Bayesian Optimisation (BO) algorithm has been adapted to the robust scenario. The performance of the BO algorithm is determined by the so-called “Infill Criterion” or “Acquisition Function”, which balances the trade-off of exploration and exploitation in the search to find the global optima. The so-called “Moment-Generating Function of the Improvement” (MGFI) has been proposed as an Infill Criterion for the nominal BO algorithm, and is extended to the robust scenario in WP2. The MGFI linearly combines all the moments of the improvement, and is compared with the baseline, namely the “Expected Improvement Criterion” (EIC) on 12 test cases [Ullah, GECCO 2021]. The results indicate the promising nature of the extended MGFI as it yields better objective function values in half of the test cases considered.

Similar to RO, DMOO problems are a special class of multi-objective problems subject to uncertainty, i.e., the objectives and/or constraints change over time, making them difficult to optimise directly. To efficiently solve such problems, the notion of "similarity" is employed in WP2. The intuition for this is that in real-world applications, problems exhibit similarity. Therefore, the knowledge extracted by solving one problem can be utilized to help solving another related problem. This technique of transferring knowledge to efficiently solve related optimisation problems is known as "Transfer Learning", and is at the core of efficiently solving DMOO problems in WP2. The key challenges in transfer learning are associated with when and how – In which situation transferring knowledge is beneficial, and how can it be implemented effectively? A comprehensive empirical study to answer these fundamental research questions was conducted in WP2 [Ruan, SSCI 2019]. The results demonstrate that in the presence of Gaussian kernels, the performance of transfer learning is significantly decreased. Therefore, the study proposes to employ Linear kernels in lieu of Gaussian kernels for knowledge transfer. A related question about the effectiveness of transfer learning with regards to computational efficiency was addressed with another empirical investigation [Ruan, CEC2020].

Constructing surrogate models of high-dimensional optimisation problems is challenging due to the computational complexity involved. Various methodologies have been proposed to deal with the issue of high dimensionality in SM including divide-and-conquer, variable screening, and mapping the data to a lower dimensional space using dimensionality reduction techniques (DRTs). The DRTs can encapsulate the high dimensional search space into compact representations. In WP2, the practicality of some of the widely adapted DRTs is examined by evaluating and comparing the low-dimensional surrogate models based on these DRTs [Ullah, SSCI 2020]. The DRTs considered are Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Autoencoders (AEs) and Variational Autoencoders (VAEs) respectively. The low-dimensional surrogate models are compared against each other based on the criteria of modeling accuracy and quality of the solution. The experimental setup in this study focuses on a total of 720 test cases. The key findings indicate the superiority of AEs and PCA for efficiently constructing the low dimensional surrogate models.

Model selection and HPO of machine learning algorithms is also challenging due to the computational complexity involved. In the domain of imbalanced learning, the problem can be exacerbated due to a large portfolio of resampling and classification techniques, which give rise to a great number of hyperparameters in total. Sequentially performing model selection and HPO in imbalanced learning is therefore costlier, and efficient methods are needed which can solve these problems in a timely manner. In WP2, instead of sequentially solving these problems, we represent both problems as a single one, which is commonly referred to as the "Combined Algorithm Selection and Hyperparameter Optimization" (CASH) problem [Nguyen, DSAA 2021]. The approach to address CASH problem is based on a special type of BO, namely the "Tree-structured Parzen Estimators" (TPE). This approach results in significant improvement in classification accuracy in imbalanced learning domain when compared with state-of-the-art techniques.

The research invested in WP2 extends the MGFI to the robust scenario, and provides a novel perspective on the applicability of SM in the context of RO. The issue of high dimensionality is addressed with the help of DRTs such as AEs and PCA. In addition, the problem of model selection and HPO for an arbitrary imbalanced data set is tackled based on TPE. The extension of the existing work to the real-world engineering case studies is left for future. Other ideas on potential future research emphasize on quantifying the so-called “Computational Cost of Robustness” (CCOR) – the need for the additional computational resources to find the robust instead of a nominal solution. Providing a new perspective on the CCOR is crucial since practitioners often overlook the intrinsic computational cost associated with the chosen robustness formulation.

Update on deliverables: WP2 encompasses four deliverables, namely D2.1-2.4, all of which have been successfully achieved. In particular, D2.1 provides an overview on the research invested for efficiently solving high dimensional optimisation problems. D2.2 reports the scientific achievements in the field of hyperparameter tuning and algorithm configuration for learning and mining in the presence of imbalance data sets. The issue of uncertainty and noise in continuous optimisation is addressed in D2.3 with the help of surrogate models. Lastly, D2.4 provides an overview on the practical implementations and reproducibility of the research in all preceding deliverables. This deliverable shares the publicly available code, as well as the technical details and instruction manual for utilizing this code.

WP3 – Big Data Analytics and Optimisation using Novel Machine Learning Approaches
NEC

WP3 aims to develop, analyze and evaluate novel machine learning algorithms for more effective and efficient optimization in the context of system engineering and big data analytics. Driven by the real industrial problems and applications, the strengths of machine learning and optimization methods are fully explored and integrated to work together towards distinguished performance with practical impact. Some of the work has been published in top-tier conferences on computational intelligence. Based on the work plan and the emerging research problems in the areas of telecommunication and automotive industry, the research focuses on such as data imbalance, compact representation of big data, and explanation of machine learning. The main achievements are reported as follows:

Class imbalance is an important but difficult problem within many optimization processes that degrades performance of traditional algorithms significantly. For example, one may find that a vehicle design optimization targeting aerodynamic performance offers more solutions which violate parameter constraints rather than being valid solutions, though the optimization process often converges properly. As a consequence, identifying and evaluating these valid solutions saves a lot of computational time. To this end, oversampling techniques are studied to enhance class imbalance problems with successful application to vehicle data available from deliverable D1.1 and can improve the performance by around 10%. This research has explored the efficiency of six powerful oversampling approaches in the imbalanced learning domain. According to the empirical analysis, oversampling techniques that consider the minority class distribution perform better in most cases and RACOG gives the best performance. Moreover, we have developed another novel method to further improve imbalanced classification by borrowing strength from anomaly detection. The proposed method learns to craft additional attributes, which are sensitive for properly modeling the overlapping region of the two classes (majority and minority classes). The proposed method provides an orthogonal vision to solve the data imbalance problem, and can integrate with the proposed oversampling techniques to work together for enhancement of the imbalanced learning performance. The two works have been published in: the 24th Pacific-Asia Conference on Knowledge Discovery and

Data Mining (Kong et al., PAKDD 2020), and the 16th International Conference on Parallel Problem Solving from Nature (Kong et al., PPSN 2020).

We have also tailored and extended the optimization technologies to address the emerging problems in new generation of telecommunication networks. Resource optimization is an important task for telecommunication network operators, as it is directly related to service quality and operation expenditure. With the application of new network architecture, e.g. ORAN (Open Radio Access Network), the problem is becoming more challenging due to the increasing complexity of the networks. Inspired by limitations of existing resource optimization methods, a new EO based solution is proposed for O-RAN. Specifically, the major limitation is the resource allocation scheme with more delay is likely selected as these methods do not clearly distinguish under- and over-provisioning cases in their optimization. In addition, the criteria of peak distribution used in the existing methods is often redundant. To solve the limitations, an EO based optimization method together with traffic load forecasting is introduced to find a resource allocation scheme that proactively and dynamically deploys the optimal amount of computing resource for processing upcoming communication traffic. Experimental studies carried out on multiple real-world datasets have demonstrated the effectiveness of the EO based approach. In addition, we investigate the impact of the parameter settings on the performance of the proposed method for model selection issue.

The hybridization of optimization with machine learning techniques has recently attracted notable interest (Friess et al., CEC2019, CEC2020, PPSN2020). For example, exploiting evolutionary optimization (EO) in deep neural networks (DNNs) has achieved significant results in AutoML. A research question naturally arises: could we improve performance of evolutionary optimization by integrating DNNs? We analyze the optimization process of the EO methods, and develop an approach that employs DNNs to learn the structure of the meta data generated during the optimization process. The proposed approach is able to significantly distinguish structures within the meta-data from different black-box EO settings. The results have been published in the IEEE International Joint Conference on Neural Networks (Friess et al., IJCNN 2021). An extension of the work is under review at the IEEE Symposium Series on Computational Intelligence (Friess et al., SSCI 2021), in which we apply the approach to car shape optimization, which is critical in automotive product design.

In real applications, the data is often massive. So another focus of WP3 is to explore robust and compact representation learning for big data, such as massive amounts of complex data, such as time series, generated by an industrial system when monitoring the system status. To meet the challenge, variational recurrent neural networks are studied to learn the representations of time series. We have enhanced the learning method by integrating high level meta information. Such information is often instructive and collected from the domain expert experience and the prior knowledge of the system, which generally reveals latent correlations and links the set of time series into a network. The information and the learned patterns can thus propagate over the network to improve forecasting of the connected time series. The proposed method has been validated on a popular benchmark MIMIC for multi-step-ahead prediction with superior performance. The work has been published in IEEE international joint conference on neural networks (Ullah et al., IJCNN 2020).

With success of ML/AI in a variety of applications, they have been pervasive in diverse vertical sectors. Trustworthiness of ML/AI is thus an emerging concern, and the corresponding techniques are being investigated to give users the confidence in embracing AI-based solutions. To engender trust in AI

systems, they should firstly be explainable. We have studied explainable AI (XAI) approaches to solve the problem in the context of personalized optimization for product design. In particular, interpretable graph neural networks (iGNN) is proposed to learn understandable customer profiling for their preference. We formulated the problem as node (customer) embedding and meanwhile learning text explanations of the embedding vectors. The iGNN models the patterns among the embedding vectors, the corresponding text explanations, and the texts associated with links between nodes via extra node cluster embedding. The proposed method was validated on user generated product review data and showed promising explanation about users' preference for product feature optimization. Furthermore, we have extended the XAI approach to provide human-understandable explanation of latent structure learning for user/product segmentation. The commonly used embedding techniques model each entity by employing dense and complicated representations. In the proposed method, we model classes of users and products with latent vectors that are compact, interpretable and rather discrete. In contrast with previous works where the interpretability is provided by post-hoc evaluations, the proposed probabilistic modeling framework allows us to explicitly learn the interpretability of the latent codes. In addition, the topographic organization of the inter-connected data provides additional benefits that help us understand the relationships among different latent classes. The experiments demonstrate that, the proposed approach learns compact discrete representations, and in the meanwhile achieves competitive results in user preference learning. More importantly, it learns interpretation of the latent representations for user and product segmentation. The two works have been published in IEEE international joint conference on neural networks (Serra et al., IJCNN 2021), and submitted to IEEE Transactions on Neural Networks and Learning Systems (under review).
Update on deliverables: In WP3, there are five deliverables that have been successfully submitted on time. D3.1 reports the proposed methods for class imbalance problems with applications in vehicle design optimization, where oversampling and anomaly detection techniques are exploited to improve the performance. In D3.2, the developed deep learning methods are reported for engineering and ICT data. The correlations of sensor data are studied. The performance on real applications, e.g. healthcare, is validated. D3.3 presents the proposed explainable AI methods to model the graph data with user-generated texts. The learned user preference facilitates product feature optimization. D3.4 introduces novel methods for proactive dynamic and robust optimization. D3.5 reports the integrated software environment and provides a detailed manual for potential users of the developed techniques.
WP4 - Core Knowledge and Transferable Skills Training **Birmingham**
Update on work packages: All training modules for the ECOLE project have been completed successfully by ESRs. In particular, ESRs have completed the core academic modules and interdisciplinary modules offered by the Universities of Birmingham and Leiden. At the end of each module, ESRs were viva-ed by senior researchers from some of the four beneficiaries of the ECOLE project. Formal certificates were given after the successful passing the vivas. For full details of the modules completed, please refer to Deliverable D4.1. In addition to academic modules, a large number of transferable skill training courses and activities were organized and provided to ESRs. The training covers speed reading skills, communications skills, entrepreneurial skills, IPR issues, conference presentation skills, project management skills, career development (in academia, industry or consulting firms), etc. Some of these training courses were

specifically designed and tailored to our ESRs. For full details of these training activities, please refer to Table 4 in Section 4 of this document.
Update on deliverables: All deliverables of the ECOLE project have been completed and submitted on time, with all the targets met.
WP5 - Project Coordination and Management **Birmingham**
Update on work packages: COVID-19 has posed a big challenge to our project because ESR's mobility has been affected significantly due to travel restrictions. Some of the physical secondments could not be implemented as originally planned. Fortunately, our beneficiaries were able to be flexible and allowed for virtual online secondments. Our research has not been affected in a major way by COVID-19 because most of ESR's work could be done virtually online. This can be seen from the large number of publications generated by ESRs. As shown by the description of WP1-3 above, we have met all the objectives of work packages. In terms of project coordination and management, the ECOLE project management team has been meeting online monthly since the outbreak of COVID-19. These are in addition to our six-monthly workshops and meetings for everyone in the project. There has been no problems or conflicts in any way so far for our project coordination and management.
Update on Deliverables: We have completed and submitted all required Deliverables for this work package and for the entire project.
WP6 – Communication and Dissemination **Birmingham**
Update on Work packages: Communication and dissemination have been carried out at different levels and through different channels. First, we have communicated and disseminated our research output through publications at conferences and journals. ESRs have presented their work personally at various international conference and seminars. Supervisors have given invited keynote and plenary talks at major international conferences or workshops. Second, we have communicated and disseminated our work through the project website, social media, email lists, etc. Third, have communicated and disseminated our project work through summer schools, student workshops, industrial workshops and seminars. All four project beneficiaries organize summer schools or workshops in turn every six months, participated by both our project ESRs as well as other researchers. Even during COVID-19, we still maintain such a six-month schedule (online of course). Fourth, we have our ESRs and supervisors outreached to broader audiences outside the traditional academic communities. For example, one of our ESR attended a public science event in Germany and gave a talk. One of our supervisors also gave a public talk in New Zealand as part of the outreach event of IEEE CEC'2019. In order to support and train our ESRs in communication and dissemination, some of our Summer School activities were purposely designed to include talks on the translation of research and a hands on statistics workshop. ESRs developed their outreach and communication strategy, with reps being selected to have oversight of contents on the channels such as website and social media.
Update on Deliverables: We have completed and submitted all required Deliverables for this work package and for the entire project.

4. Thesis Plan and Career Development

4.1 Training Plan

All ESRs have completed the Academic Modules (AMs), the Research Skill Training modules (RST), and The Personal and Career Skill Training modules (PCST) as required by the original grant proposal, as indicated in Table 4 below. The full details of these training activities are documented in Deliverable D4.1 and summarized below.

Table 3 - Training Activities

Module	Learning Objective/description	**ESR Number**							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
AM1	Nature Inspired Search and Optimisation: A comprehensive introduction to the field of natural inspired optimisation, covering theories, algorithms and applications								
AM2	Intelligent Data Analysis: A comprehensive introduction to statistical pattern analysis, high-dimensional data mining, and text mining.								
AM3	Machine Learning: Advanced topics in machine learning, covering several forms of supervised, semi-supervised and unsupervised learning, in both theories and applications.								
AM4	Multiple-Criteria Optimisation and Decision Analysis: Theoretical foundations, algorithms, and application techniques of multi-objective optimisation								
AM5	Advances in Data Mining: Recent developments in data mining for classification, regression and clustering and beyond, dealing with massive data sets. Techniques for distributed data mining (e.g., Hadoop).								
AM6	Evolutionary Algorithms: State-of-the-art in evolutionary computation; including efficient optimisation techniques (i.e., small number of function evaluations)								
RST1	Research Skills: Including literature review, academic writing, referencing, LATEX, presentations, etc.								
RST2	Reading Groups: In reading group meetings, ESRs will present their work to the rest of the group on a rotating basis and strengthen their communication skills and confidence in presenting their work								
RST3	Research Seminars: In seminars, primarily given by external speakers, ESRs will network with leading professionals and extend their research insights.								
PCST1	Outreach training courses: Half-day course focusing on the effective communication of research to wider public and academic community								

PCST2	Time management course: Half-day course discusses planning techniques and resources for time management, helping students to prioritise to-do lists.								
PCST3	Speed reading workshop: Half-day workshop aiming at improving personal skills for more effective handling of reading material, master the techniques of scanning and skimming and improve the retention of written documents.								
PCST4	Talent pool: A unique five-day programme for ESRs, providing customised professional development training and access to practical, 'real-life' opportunities to develop the next generation of academics, business consultants and entrepreneurs								
PCST5	Enterprise training programme (Medici): A training programme with flexible time, helping ESRs realise the commercial potential of their research, covering business strategies, marketing, finance and business planning, funding opportunities, sale and negotiation skills and networking skills.								
PCST6	Open science programme: Focus on publishing and communicating scientific knowledge, particularly open access and open data strategies and techniques.								
PCST7	Intellectual Property Rights: Introduction to the formal aspects of intellectual properties, including guidelines on how to summarise research results in a way that they can be turned into powerful IPR by IPR professionals.								
PCST8	Technology Evaluation and Transfer: Introduction to the steps of turning research results into technology, technology into innovations and innovations into products and services with a sustainable business value.								
PCST9	Project Management in Research: Planning, Quality and Risks: Introduction to different standard methods for project management, including risk management.								

4.2 Outreach

The ESRs discussed plans for outreach at Summer School. The plan is to maximize content on existing channels with ESRs social media reps selected and dedicated to posting content on Twitter and Instagram. Twitter will have alternative week output. Instagram will contain a collection of pictures from everyday life of ESRs and publish when required. The audience that has been selected is public and specific scientific/industrial audience and content will be tailored as per. The ECOLE leaflets are finalized and will be circulated to appropriate audiences once they have been received.

<u>BIS Conference – 26th June, 2018, Seville</u>

The 22nd International Conference on Business Information Systems (BIS) took place in June. ESR5 presented her research about class imbalance in the "Doctoral Consortium" to an audience of PhD students and mentors from other institutions. She gave a short introduction of ECOLE to the audience.

Z2X Festival – 31st August, 2018, Berlin

ESR6 recently took part in the Z2X festival organised by the German newspaper "Die Zeit" and participated in a 10 minute talk delivered to young adults active in science, activism and arts. The aim of the event is to encourage exchange of interesting and novel ideas. ESR6 presented a talk on AI with reference to the ECOLE project. The talk titled "Nature Inspired Artificial Intelligence" was received well by the audience and encouraged questions from other participants. *"I have been invited by a group later in this evening who enjoyed my talk to join them for further discussion and beers. Overall I really enjoyed the experience and would likely do a popular talk again when given the opportunity."* ESR6

Further outreach activities by ESRs are detailed below and reported in Deliverable 5.4.

ESR Number	Outreach Activities
1	Attended conferences: a. With paper presentation • IEEE Symposium Series on Computational Intelligence (SSCI) 2019, 6-9 December 2019, Xiamen, China • IEEE World Congress on Computational Intelligence (WCCI) 2020, 19-24 July 2020, Glasgow, UK (online) • IEEE Symposium Series on Computational Intelligence (SSCI) 2020, 1-4 December 2020, Canberra, Australia (online) • IEEE Congress on Evolutionary Computation (CEC) 2021, 28 June – 1 July, Kraków, Poland (online) b. Without paper presentation • Symposium on Geometry Processing 2020, 4-8 July 2020, Utrecht, The Netherlands (online) • Symposium on Geometry Processing 2021, 10-14 July 2021, Toronto Ontario, Canada (online) Other activities: • 3rd International Summer School on Deep Learning, 22-26 July, Warsaw, Poland • Talk in the "Career planning" course of the Mechanical Engineering Department at the Federal University of Santa Catarina (UFSC), Brazil, 19 August 2021 (online)
2	Attended conferences: a. With paper presentation

<table>
<tr><td></td><td>

International Conference on Data Mining Workshops (ICDMW) 2019, 8-11 November 2019, Beijing, China
IEEE Symposium Series on Computational Intelligence (SSCI) 2019, 6-9 December 2019, Xiamen, China
IEEE Symposium Series on Computational Intelligence (SSCI) 2020, 1-4 December 2020, Canberra, Australia (online)
IEEE Congress on Evolutionary Computation (CEC) 2021, 28 June – 1 July, Kraków, Poland (online)

b. Without paper presentation

Symposium on Geometry Processing 2021, 10-14 July 2021, Toronto Ontario, Canada (online)
Summer school on Data Driven Artificial/Computational Intelligence, 23-26 August, UK (online)

<u>Other activities</u>:

3rd International Summer School on Deep Learning, 22-26 July, Warsaw, Poland

</td></tr>
<tr><td>3</td><td>
<u>Attended conferences</u>:

IEEE Symposium Series on Computational Intelligence (SSCI), 2019.
IEEE World Congress on Computational Intelligence (WCCI), 2020.
Sixteenth International Conference on Parallel Problem Solving from Nature, 2020.
IEEE Symposium Series on Computational Intelligence (SSCI), 2020.
Genetic and Evolutionary Computation Conference (GECCO), 2021.

<u>Other activities</u>:

2nd Dortmund-Bielefeld Summer School on Time Series Analysis, 2019.
Malvern Science in the Park, 2020.
IEEE Summer School on Data-Driven Optimization, 2021.

</td></tr>
<tr><td>4</td><td>
<u>Other activities</u>:

The COST Action CA 15140 training, 25th-29th, November 2019 in Coimbra, Portugal (attended training classes)
2021 IEEE CIS Summer School on Data Driven Artificial/Computational Intelligence: Theory and Applications, 23-26 August 2021, Virtual Event (attended training classes).

</td></tr>
<tr><td>5</td><td>
<u>Attended conferences</u>:

2019 IEEE symposium series on computational intelligence (SSCI) [as an author and presenter]

</td></tr>
</table>

	• The 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) [as an author and presenter] • The 16th International Conference on Parallel Problem Solving from Nature (PPSN) [as an author and presenter and online chair] Other activities: • 22nd International Conference on Business Information Systems, 26-28 June 2019, Venue: ETSI Informática, University of Seville [give a presentation] • The COST Action CA15140 training, 25th-29th November 2019 in Coimbra, Portugal [attending training classes] • 2021 IEEE CIS Summer School on Data Driven Artificial/Computational Intelligence: Theory and Applications, 23 - 26 August 2021, Virtual Event [attending training classes]
6	Attended conferences: • IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, 2019 • EvoStar 2020 • IEEE World Congress on Computational Intelligence (WCCI), 2020 • International Conference on Parallel Problem Solving from Nature (PPSN), 2020 • IEEE IJCNN, virtual, 17-22 July 2021 Other activities: • Session Chair for "Optimisation and Signal Processing" at IEEE IJCNN 2021. • Google Virtual Workshop on Conceptual Understanding of DL • IPAM Workshop on Deep Learning and Combinatorial Optimization • Talk at Z2X19 Popular Science Session Knowledge-to-Go on "Nature-inspired Artificial Intelligence • 3rd International Summer School on Deep Learning (DeepLearn 2019) • The COST Action CA 15140 training, 25th-29th, November 2019 in Coimbra, Portugal (attended training classes)
7	Attended conferences: • IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, 2019 • IEEE World Congress on Computational Intelligence (WCCI), 2020 Other activities: • The COST Action CA 15140 training, 25th-29th, November 2019 in Coimbra, Portugal (attended training classes) • 2021 IEEE CIS Summer School on Data Driven Artificial/Computational Intelligence: Theory and Applications, 23-26 August 2021, Virtual Event (attended training classes).

8	<u>Attended conferences</u>: • IJCNN 2021, virtual, 17-22 July 2021 <u>Other activities</u>: • PC member at ECML-PKDD 2020 (July 2020) • EGN Symposium at HRI (September 2020) • EGN Seminar Talk at HRI (November 2020) • BAI-ML Meetup at NLE (March 2021) • DYNAmore Workshop (June 2021) • Tutorial : Deep Learning for Graphs (instructor: Davide Bacciu) @ IJCNN2021 (July 2021) • Summer School on Data-Driven Artificial/Computational Intelligence: Theory and Applications (August 2021)

4.3 Career Development Plans

In the following subsections, we present the career development plans for each ESR. Please note that all ESRs have successfully taken part in all events organized by the ECOLE project (as time of writing Sep 2021) and successfully passed all provided courses. Please see the other subsections in this report to refer to more detailed information on these workshops and summer schools as well as course lists. For more details on the individuals PhD-thesis please refer to the outlines in Appendix A.

4.3.1 Thiago Rios

Thiago Rios (ESR1)
Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:
Learning-based Representations of High-dimensional CAE Models for Automotive Design Optimization
Thesis Objectives. Please state the aims of your research and timeline for achieving them:
In design optimization problems, engineers typically handcraft design representations based on personal expertise, which leaves a fingerprint of the user experience in the optimization data. Thus, learning this notion of experience as transferrable design features has potential to improve the performance of similar, yet more challenging, design optimization problems. In a first step, we developed a point cloud autoencoder for 3D vehicle designs and evaluated its performance towards scalability and usability in optimization (2 peer-reviewed conference publications, 2018/19). For a deeper understanding of the autoencoders as well as their applicability for simulation-based design optimization, we proposed a novel feature visualization method and suggested a framework for automated vehicle surface mesh generation (2 peer-reviewed conference publications, 2019/20). We then further exploited local geometric features for improved design optimization, suggested a novel deep neural network architecture for improved surface mesh generation and demonstrated experience transfer in multi-task optimization (1 peer-reviewed conference and 1 peer-reviewed journal publication, 1 further conference submission currently under review, 2020/21). Further contributions were made to research carried out by other ESRs for learning user time-series data and class imbalanced classification (3 peer-reviewed conference publications as

co-author). All software code is available for other researchers at https://github.com/ECOLE-ITN/GDL4DesignApps. In addition, a shape deformation set-up for simulation-based vehicle optimization was realized together with ESR2 and made available to the ECOLE project.
Training Objectives. Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:
I improved soft and hard skills by participating in trainings on technical reading, scientific writing, and workshops on research-related topics, e.g., International Summer School on Deep Learning.
Publications. What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe: 1. novel engineering methods for faster product development time and new high-quality products 2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks 3. novel machine learning methods for product feature optimisation and big data analytics. Indicate if your research falls into these categories *(please use the publication consent form to notify beneficiaries of publication 1 month prior to submission)*.
The publication strategy was successfully realized with a high number of publications in high-ranking peer-reviewed IEEE conferences and journals targeting (1.) novel engineering methods for faster product development and by (2.) utilizing learning methods to learn from similar design tasks
Outreach Activities. Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):
I presented my research at high-ranking international conferences (SSCI 2019 and 2020, IJCNN 2020, CEC 2021) and participated in conferences and events related to the topic of my research (SGP 2020 and 2021, Ansys Simulation World 2021), where I discussed with several researchers and specialists. Finally, I also gave scientific talks at HRI-EU and in ECOLE workshops, and gave a guest lecture at the Federal University of Santa Catarina about my experience as former student at the university and currently as Early-Stage Researcher in the ECOLE project.
Conferences. Please describe your (upcoming) participation in conferences:
International Conference on 3D Vision, 2021 (tentative)

Discussed with (Insert PI Signature):	
Date:	Sep 2021
Declaration:	
I, Thiago Rios, have discussed my career development plan with my supervisor and we have agreed on methods for the continuation of my personal development that are satisfactory to all parties.	

4.3.2 Sneha Saha

Sneha Saha (ESR2)
Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:
Learning-based Representations of High-dimensional CAE Models for Automotive Design Optimization

Thesis Objectives. Please state the aims of your research and timeline for achieving them:
In the early design phase of automotive digital development, engineers exploit their experience to address the challenges in design ideations and optimization problems to develop a new prototype. In our research, we imagine a cooperative design system in the automotive domain that can learn from these digital data and from engineer's expertise to provide assistance to the engineers for several tasks in the design process. First, we proposed methods to capture the engineer's expertise and provide guidance during the design process (2 peer-reviewed conference publications, 2018/19). Second, we proposed a deep-generative variational autoencoder model for vehicle data that learns low-dimensional representations of the design data in an unsupervised fashion. We evaluated this data-driven representation for generating realistic novel shapes, surrogate modelling for performance predictions, and multi-criteria optimization tasks (1 peer-reviewed conference publication 2020, 1 further conference submission currently under review 2021). We also explored the feasibility of this learned representation for multi-objective optimization and exploited the learned knowledge to achieve faster convergence to optimal solutions and to better address the engineer's preference (1 peer-reviewed conference publication 2021). All software code is available for other researchers at https://github.com/ECOLE-ITN/GDL4DesignApps. In addition, a shape deformation set-up for simulation-based vehicle optimization was realized together with ESR1 and made available to the ECOLE project.
Training Objectives. Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:
I improved soft and hard skills by participating in trainings on technical reading, scientific writing, and workshops on research-related topics, e.g., International Summer School on Deep Learning.
Publications. What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe: 1. novel engineering methods for faster product development time and new high-quality products 2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks 3. novel machine learning methods for product feature optimisation and big data analytics. Indicate if your research falls into these categories *(please use the publication consent form to notify beneficiaries of publication 1 month prior to submission)*.
The publication strategy was successfully realized by a high number of publications in high-ranking peer-reviewed IEEE conferences targeting (1.) novel engineering methods for faster product development by (2.) utilizing learning methods to learn from similar design tasks.
Outreach Activities. Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):
I presented at high-ranking international conferences (LMID 2019, SSCI 2019 and 2020, CEC 2021) and participated in conferences and events related to the topic of my research (SGP 2021). Finally, I also gave scientific talks at HRI-EU and in ECOLE workshops.
Conferences. Please describe your (upcoming) participation in conferences:

IEEE Symposium Series on Computational Intelligence 2021 (tentative)	
Discussed with (Insert PI Signature):	
Date:	Sep 2021
Declaration:	
I, Sneha Saha, have discussed my career development plan with my supervisor and we have agreed on methods for the continuation of my personal development that are satisfactory to all parties.	

4.3.3 Sibghat Ullah

Sibghat Ullah (ESR3)
Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:
Model-Assisted Robust Optimization for Continuous Black-Box Problems
Thesis Objectives. Please state the aims of your research and timeline for achieving them:
While solving real-world optimization problems, a frequently-encountered obstacle is the presence of the uncertainties and noise within the system, or model of the system, for which optima are sought. Due to various reasons, various types of uncertainties and noise can emerge in optimization problems. Due to the uncertainties and the black-box assumption on the optimization setup, traditional optimization methods are rendered inapplicable for real-world scenarios. The objective of this thesis involves developing, analysing and comparing methods to efficiently solve noisy black-box problems. In the first step, we evaluate the practicality of the so-called “Surrogate-Modeling” (SM) to efficiently solving expensive-to-evaluate black-box problems in the face of uncertainty and noise (one peer-reviewed publication 2018-19). Since high dimensionality can affect the computational effort required to find the optimal solution with surrogate models, we explore the suitability of some of the widely adapted dimensionality reduction techniques (one peer-reviewed publication, 2019-20). The dimensionality reduction techniques discussed in the thesis are Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Autoencoders (AEs) and Variational Autoencoders (VAEs) respectively. Our findings provide a novel perspective on the issue of high dimensionality in SM as they are based on a comprehensive empirical study with 720 test cases. The famous “Bayesian Optimization” (BO) algorithm has been adapted to the robust scenario to tackle expensive-to-evaluate noisy black-box problems. As an important contribution, we extend the so-called “Moment-Generating Function of the Improvement” (MGFI) as an acquisition function for the BO algorithm (one peer-reviewed publication, 2020-21).
Training Objectives. Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:
To improve the technical and interpersonal skills, several training activities have been completed. The training activities include participation in summer schools, academic modules, and outreach activities.
Publications. What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe: 1. novel engineering methods for faster product development time and new high-quality products 2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks 3. novel machine learning methods for product feature optimisation and big data analytics.

Indicate if your research falls into these categories *(please use the publication consent form to notify beneficiaries of publication 1 month prior to submission)*.
Publication strategy: publications. The research falls in the category no. 2.
Outreach Activities. Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):
I have presented the key findings of my research on the surrogate modeling in IEEE Symposium Series on Computational Intelligence (SSCI), 2019. Furthermore, the research on the issue of robust time series forecasting has been presented in International Joint Conference on Neural Networks (IJCNN), 2020. The key results originating from the empirical study on the issue of high dimensionality were shared in the IEEE SSCI, 2020. The latest research on extending the MGFI to the robust scenario was presented in Genetic and Evolutionary Computation Conference (GECCO), 2021. Apart from the presentations on the conference venues, an online session introducing the younger audience to Artificial Intelligence was also held.
Conferences. Please describe your (upcoming) participation in conferences:
IEEE Symposium Series on Computational Intelligence 2021

Discussed with (Insert PI Signature):	
Date:	Sep 2021

Declaration:
I, Sibghat Ullah, have discussed my career development plan with my supervisor and we have agreed methods for the continuation of my personal development that are satisfactory to all parties.

4.3.4 Duc Anh Nguyen

Duc Anh Nguyen (ESR4)
Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:
Automatic algorithm configuration for parameter tuning of modelling and optimization algorithms
Thesis Objectives. Please state the aims of your research and timeline for achieving them:
A plethora of algorithmic strands exists for applying machine learning to real-world problems. However, for efficient implementation of these algorithms, the practitioners have to make several high-level decisions such as choosing which approach to use on a given data set, whether and how to preprocess data and tuning hyperparameters. Therefore, in order to efficiently apply the machine learning techniques in solving real-world problems, we focus on the research topic of how to automatically select the optimally suited algorithm and its optimized hyperparameter settings for an arbitrary real-world problem from a given portfolio of the existing works and achievements with less human effort. In other words, we take care of the final step of every research line to make the research achievements accessible to the non-expert. As a first step, we perform algorithm selection and hyperparameter tuning for imbalanced learning (one peer-reviewed publication, 2021). The problem is tackled with the help of Tree-structured Parzen Estimators (TPE), a special class of Bayesian Optimization (BO) algorithms. The code to reproduce scientific findings in this direction is available at https://github.com/ECOLE-ITN/CASH4IMBALANCE. This research line is further extended to solve a real-world imbalance problem at TATA Steel (The Netherlands), finalized in September

<table>
<tr><td colspan="2">2021. Furthermore, we utilize the TPE for performing dimensionality reduction in high dimensional model-assisted optimization.
In addition, we propose a novel sampling approach to improve BO used in Automated Machine Learning (AutoML) optimization, together with a new BO variant and a new AutoML framework. The key findings is submitted to the IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2021). These libraries and source code will be available at https://github.com/ECOLE-ITN/BO4ML.</td></tr>
<tr><td colspan="2">Training Objectives.
Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:</td></tr>
<tr><td colspan="2">I improved soft and hard skills by participating in training on technical reading, scientific writing, and workshops on research-related topics, e.g., the COST Action Training School (2019) and the SS-DDACI summer school (2021).</td></tr>
<tr><td colspan="2">Publications.
What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe:
1. novel engineering methods for faster product development time and new high-quality products 2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks
3. novel machine learning methods for product feature optimisation and big data analytics.

Indicate if your research falls into these categories (please use the publication consent form to notify beneficiaries of publication 1 month prior to submission).</td></tr>
<tr><td colspan="2">The publication strategy focuses on publications, which was successfully achieved by publications at high-ranking peer-reviewed IEEE conferences and journal targeting (1.) novel engineering methods for faster product development by (3.) novel machine learning methods for product feature optimisation and big data analytics.</td></tr>
<tr><td colspan="2">Outreach Activities.
Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):</td></tr>
<tr><td colspan="2">I will present the key findings of my study on imbalanced learning and hyperparameter tuning at the IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2021. Furthermore, I have participated in events related to the topic of my research (COST Action Training School-2019), where I discussed the latest research with various specialists.</td></tr>
<tr><td colspan="2">Conferences.
Please describe your (upcoming) participation in conferences:</td></tr>
<tr><td colspan="2">DSAA 2021 (presenter, Oct 2021)
IEEE SSCI 2021 (tentative)</td></tr>
<tr><td>Discussed with (Insert PI Signature):</td><td></td></tr>
<tr><td>Date:</td><td>Sep 2021</td></tr>
<tr><td colspan="2">Declaration:</td></tr>
<tr><td colspan="2">I, Duc Anh Nguyen, have discussed my career development plan with my supervisor and we have agreed methods for the continuation of my personal development that are satisfactory to all parties.</td></tr>
</table>

4.3.5 Jiawen Kong

Jiawen Kong (ESR5)

Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:
Class imbalance classification through semi-supervised and active learning for experience-based optimisation
Thesis Objectives. Please state the aims of your research and timeline for achieving them:
In real-world classification problems, there are many applications that suffer from the class-imbalance problem, for instance, fault diagnosis, anomaly detection, medical diagnosis and etc. In these problems, it is much more important to correctly identify the minority samples. The price of misclassifying the minority samples would be a huge loss of money in fault diagnosis, an unqualified product in anomaly detection and a person's life in medical diagnosis. Hence, it is significant to improve the class-imbalance classification. Within the imbalance research in ECOLE project, data complexity in the imbalanced datasets is studied. The particular focus was on whether significant performance improvement can be achieved on any given imbalanced datasets through performing hyperparameter optimisation techniques (1 peer-reviewed conference paper). After that, we further studied the performance of several resampling techniques and investigated the relationship between data complexity measures and different resampling techniques, based on both benchmark datasets and a real-world inspired vehicle mesh dataset (1 peer-reviewed conference paper and 1 journal submission currently under review). Moreover, we also proposed to improve imbalanced classification by introducing additional attributes, which gives significant improvement on imbalanced classification performance and is simple to implement and can be combined with resampling techniques and other algorithmic-level approaches in the imbalanced learning domain (1 peer-reviewed conference paper).
Training Objectives. Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:
The domain knowledge and intellectual abilities as well as the personal skills have been improved by attending various summer schools and workshops, e.g. Doctoral Consortium in BIS 2019, COST action (CA15140), Summer school on Data Driven Artificial/Computational Intelligence and etc.
Publications. What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe: 1. novel engineering methods for faster product development time and new high-quality products 2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks 3. novel machine learning methods for product feature optimisation and big data analytics. Indicate if your research falls into these categories *(please use the publication consent form to notify beneficiaries of publication 1 month prior to submission)*.
Publication strategy: publications. The research falls in the category No. 3.
Outreach Activities. Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):
I presented my research and our ECOLE project at various international conferences (BIS 2019, SSCI 2019, PAKDD 2020 and PPSN 2020), where I communicated with other researchers and had nice

discussions. As an Early-Stage Researcher in ECOLE, I also presented my work to our colleagues in HRI-EU, NEC and TATA Steel, which provides us opportunities to corporate with each other.
Conferences. Please describe your (upcoming) participation in conferences:
PAKDD2022 (tentative)

Discussed with (Insert PI Signature):	
Date:	Sep 2021

Declaration:
I, Jiawen Kong, have discussed my career development plan with my supervisor and we have agreed methods for the continuation of my personal development that are satisfactory to all parties.

4.3.6 Stephen Friess

Name:	Stephen Friess
ESR No:	6

Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:
Preliminary: Predicting Inductive Biases for Search-based Optimization
Thesis Objectives. Please state the aims of your research and timeline for achieving them:
Within recent years, knowledge transfer research in population-based optimization has gained significant traction. Notable work takes inspiration from research on transfer learning as a way to exploit complementarities between problems. However, these approaches suffer from a lack of understanding on what constitutes problem similarity in the first place. Thus, it is elusive in regards to what contributes to performance improvements in a knowledge transfer scenario. Uncontrolled, it has the capability to jeopardize algorithm performance in what is dubbed as 'negative transfer'. The objectives of our research are therefore the following: First of all, we want to clarify on what can be considered to be knowledge within population-based algorithms in the first place. And second, we want to understand how it can be consolidated and harnessed in a domain-dependent manner. We find within our research, that essentially optimization algorithms need to form inductive biases. These can be for instance represented through search operators which can be problem-tailored by incorporating information about the problem-structure. We proposed this idea within a paper submitted IEEE CEC 2020. Secondly, we need a way of arbitration between operators. Such can be done through methodology from algorithm behaviour studies. We investigated such an approach and showed that graph structures can particularly realize performance gains in a paper accepted at IEEE IJCNN 2021. In principle, for application scenarios constructing both, operators and a predictor is problematic for application scenarios. However, one can reframe it as a problem of predicting operator configurations. We demonstrated the efficacy so in a paper submitted to IEEE SSCI 2021 with an application to a shape optimization scenario. For the remaining time of the ECOLE project, we intend to attempt a journal submission to Elsevier AIJ based upon the ideas proposed within our conference papers .
Training Objectives.

Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:
I received training in the foundations of Natural Computing and Complex Adaptive Systems through ECOLE training in Genetics and Neuroscience, as well as attended external seminars such as the Virtual Google Workshop on Conceptual Understanding of Deep Learning as well as IPAM Workshop on Deep Learning in Combinatorial Optimization at UCLA (virtual).
Publications. What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe: 1. novel engineering methods for faster product development time and new high-quality products 2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks 3. novel machine learning methods for product feature optimisation and big data analytics. Indicate if your research falls into these categories *(please use the publication consent form to notify beneficiaries of publication 1 month prior to submission)*.
The research falls in the category No. 3.
Outreach Activities. Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):
• Accepted and presented a conference paper at IEEE IJCNN 2021. • Volunteered as Session Chair for “Optimisation and Signal Processing” at IEEE IJCNN 2021. • Submitted conference paper at IEEE SSCI 2021 (tentative)
Conferences. Please describe your (upcoming) participation in conferences:
• IEEE SSCI 2021 (tentative)

Discussed with (Insert PI Signature):	
Date:	16.09.2021
Declaration:	
I, Stephen Friess, have discussed my career development plan with my supervisors and we have agreed methods for the continuation of my personal development that are satisfactory to all parties.	

4.3.7 Gan Ruan

Name:	Gan Ruan
ESR No:	7
Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:	
Knowledge transfer in evolutionary dynamic multi-objective optimization	

Thesis Objectives.
Please state the aims of your research and timeline for achieving them:

There are many real-world dynamic multi-objective optimization problems (DMOPs) where multiple conflicting objectives may change over time. It is therefore significant to study how to solve them. However, most existing evolutionary algorithms for solving DMOPs neglects the useful experience or knowledge obtained when solving past problems. Recently, people firstly apply transfer learning for solving DMOPs, named as transfer learning-based dynamic multi-objective evolutionary algorithm (Tr-DMOEA). However, it has been found that transfer learning does not always help in Tr-DMOEA and the used Gaussian kernel function is not ideal. Therefore, we try to answer the research question of when and how to transfer knowledge in Tr-DMOEA and proposed an improved version of Tr-DMOEA in a paper accepted by IEEE SSCI 2019. In addition, it has been also found that Tr-DMOEA is extremely computational cost. We did computational study on how to improve the efficiency and effectiveness of Tr-DMOEA in a paper accepted by IEEE CEC 2020. In DMOPs, the number of objectives may also vary over time. They are called DMOPs with a changing number of objectives. The existing algorithm for solving DMOPs with a changing number of objectives is unable to provide enough diversity for DMOPs with a changing number of objective where the problem property is complex. Therefore, a knowledge transfer-based technique was proposed to improve the diversity while maintain the convergence of a population. This work is under preparation with experimental studies. Through literature review, it has also been found that most existing DMOEAs using transfer learning algorithms always regard solutions found in the past problem as the only knowledge to be transferred. However, the experience of search for optimal solutions for the problem has not been explored and taken into consideration. Therefore, in the rest time of PhD period, we intend to study on what experience to be extracted from the search process, when and how to transfer the search experience for solving DMOPs.

Training Objectives.
Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:

Take the training classes in the 2021 IEEE CIS Summer School on Data Driven Artificial/Computational Intelligence: Theory and Applications, 23-26 August 2021, Virtual Event.

Publications.
What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe:
1. novel engineering methods for faster product development time and new high-quality products
2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks
3. novel machine learning methods for product feature optimisation and big data analytics.

Indicate if your research falls into these categories *(please use the publication consent form to notify beneficiaries of publication 1 month prior to submission)*.

Publication strategies: study how to solve a research problem; select an appropriate journal or conference to submit a research paper after solving the research problem; prepare the draft paper according to the requirement of the journal or conference regarding format of article, references, ect and revise it before submitting it; revise the paper according to reviews' feedback and resubmit it until it is accepted.
My research falls into the second category which develops novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks.

Outreach Activities. Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):	
I have been involved into 2021 IEEE CIS online Summer School on Data Driven Artificial/Computational Intelligence: Theory and Applications, 23-26 August 2021. I would like to attend some workshops or conferences to present my research to wider academic community.	
Conferences. Please describe your (upcoming) participation in conferences:	
Past participation: onsite IEEE Symposium Series on Computational Intelligence 2019 and online IEEE World Congress on Computational Intelligence 2020 conferences.	
Discussed with (Insert PI Signature):	
Date:	Sep. 2021
Declaration:	
I, Gan Ruan, have discussed my career development plan with my supervisor and we have agreed methods for the continuation of my personal development that are satisfactory to all parties.	

4.3.8 Giuseppe Serra

Giuseppe Serra (ESR8)
Thesis Title. Please state the name of your thesis title, this can be amended if necessary in future:
Explainable Models for Preference Learning with User-generated Data
Thesis Objectives. Please state the aims of your research and timeline for achieving them:
AI-based systems have been widely used in a variety of applications. However they are often viewed as black boxes. The users do not know how the AI methods reach the results, which largely influence their confidence in embracing these solutions. To engender trust in AI systems, they should firstly be explainable. The thesis aims to study explainable AI methods with user generated data, such as texts and user behavior. During the first period, the interpretable node embedding method is explored with text-labelled graph data. Node embedding is important in graph learning. The embedded vectors are used in many down-stream tasks, however the high dimensional vectors are not understandable for humans. A graph neural network method is proposed to learn interpretable embeddings. Applications with real world data demonstrate the performance of the proposed method. The results have been published in IJCNN 2021. Next, the interpretable latent structure learning is explored, which learn explanation of the clusters of the inter-connected entities. The work is used for interpretable preference learning of users. The current results have been submitted to IEEE Trans. Neural Networks and Learning Systems. Now, further investigations are performed to understand possible new directions in this area, e.g. measurement of interpretability quality. In particular, along with the development on new interpretable approaches for graph neural networks, we will investigate in depth how to clearly quantify interpretability and trust. A new quantitative criterion of interpretability will be introduced. The results will tentatively be reported to IJCAI 2022.
Training Objectives.

Please state plans for training including transferrable skills training (e.g. public engagement, communication, entrepreneurship, technology transfer, IP). See page 2 of this plan for full details of courses stated in the proposal and indicate if they are completed:
Successfully completed all the training activities provided by ECOLE, including the related courses of UoB and Leiden, the summer schools and workshops on the advances of machine learning and optimization, as well as the training about soft skills, such as technology transfer, IPR and management skills. Actively served in PC of multiple conferences to train paper writing and presentation.
Publications. What is the publication strategy for your research (publication, software, patent)? The proposal states anticipated publications will describe: 1. novel engineering methods for faster product development time and new high-quality products 2. novel, multi-objective optimisation methods utilizing learning methods to learn from similar tasks 3. novel machine learning methods for product feature optimisation and big data analytics. Indicate if your research falls into these categories *(please use the publication consent form to notify beneficiaries of publication 1 month prior to submission)*.
Publication strategy: publications. The research falls in the category no. 3.
Outreach Activities. Please state what activities you have been/would like to be involved in. Please include any information on dissemination of your research to a lay audience (links if possible):
Disseminated research results in IJCNN 2021 (two accepted papers).
Conferences. Please describe your (upcoming) participation in conferences:
IJCAI2022 (tentative)

Discussed with (Insert PI Signature):	
Date:	02/09/2021
Declaration:	
I, Giuseppe Serra, have discussed my career development plan with my supervisors and we have agreed methods for the continuation of my personal development that are satisfactory to all parties.	

4.4 ECOLE Summer schools and workshops

Each beneficiary briefly reports about their summer school and workshop events according to table 1.2b

University of Birmingham:
The University of Birmingham hosted a workshop on the 20th and 21st of May, 2019. It was held at the Edgbaston Park Hotel and Conference Centre, Birmingham. In the first day, there were five talks given by the supervisors and researchers in the related fields, including Dr. Leandro L. Minku (University of Birmingham), Prof. Peter Tino (University of Birmingham), Prof. Ata Kaban (University of Birmingham), Dr. Per Kristian Lehre (University of Birmingham) and Dr. Zhao Xu (NEC). The talks covered statistical core knowledge, statistical machine learning and emerging topics in evolutionary computation and deep learning models. In the second day, all the ESRs gave presentations on their research with feedback given from the audiences. There were 21 participants during the workshop, including all the ESRs.

The University of Birmingham co-led the organization of the international workshop on Learning with Imbalanced Domains, in conjunction with ECML/PKDD 2021. It was held on the 17th of September, 2021. The workshop focused on providing a significant contribution to the problems of learning with imbalanced domains, aiming to increase the interest and the contributions to solving its challenges. There were 11 submissions, and 8 were accepted. Its following special issue with journal Machine Learning is currently live for paper submissions (https://www.springer.com/journal/10994/updates/19536896).

Leiden University:

Leiden University hosted the annual ECOLE Summer School on the 8th and 9th of June 2021. Due to the Corona measures, it was decided that instead of a live event, the Summer School would be organized in an online environment. For the Summer School, international researchers that belong to the best in their field were invited to join and present their research. Speakers included Prof. Dr. Kate Smith Miles (University of Melbourne, Australia), Prof. Dr. Markus Zimmermann (Technische Universität München, Germany), Prof. Dr.-Ing. Yaochu Jin (University of Surrey, UK), Dr.-Ing. Carola Doerr (Sorbonne Université, Paris), Dr. Tobias Rodemann (Honda Research Institute Europe GmbH, Germany) and Dr. Carlos A. Coello Coello (CINVESTAV-IPN, Mexico). Their talks presented a mix of both theoretical and applied computer science, thereby contributing to the main goals of the ECOLE project.

Kate Smith Miles presented how a new methodology called Instance Space Analysis could be used to support the objective testing of algorithms. In her talk, she also showed how the methodology works using university timetabling. The talk of Markus Zimmermann demonstrated how the V-model of systems engineering can help in design issues, thereby drawing on examples from the automotive industry. Yaochu Jin showed how two recently developed algorithms can be used to solve high-dimensional and many-objective optimization problems. In the talk of Carola Doerr, the complementarity of theoretical and empirical research in optimization was stressed and different techniques to achieve this balance were demonstrated. Tobias Rodemann's talk demonstrated the importance of practice in optimization research. With examples from the energy and mobility sector he showed that the factor 'human' should not be neglected in the design of optimization methods. The day ended with the talk of Carlos A. Coello Coello, who summarized the history of multi-objective evolutionary algorithms and shared some provocative thoughts on how the field should develop in the future in his eyes.

The necessity to organize the Summer School online had two main advantages. Firstly, there was the opportunity to invite speakers that would otherwise, due to travel times and busy schedules, not have been able to join. Secondly, because of the worldwide experts speaking, it was decided to open up the Summer School to a broader audience. Apart from the ESRs and their supervisors, also people from the research groups of the ECOLE universities joined the lectures. On average, 35 to 40 people participated during the whole day. This also stimulated the discussions after each talk. Overall, the Summer School was positively received.

In the evening of the first day, a social event was organized for the ESRs and their supervisors. This provided the opportunity to create stronger social bonds and substitute part of the missing social interactions during the day.

The second day of the Summer School was dedicated to the ESRs presenting their research projects to the whole group.

HRI-EU:

Both summer schools and workshops (Fall 2019 and Fall 2020) organized by HRI-EU set a specific focus on the importance of research in optimization and machine learning methods applied to industrially relevant problems. The fact that real-world product optimization typically utilizes time-consuming digital simulations amplifies the importance of extracting and exploiting experience based on historical data. In the first event, HRI-EU staff gave insights into advanced evolutionary optimization and machine learning research as well as into project management techniques. Dr. Sebastian Schmitt and Dr. Mariusz Bujny (both HRI-EU) presented their most recent research results in structural design optimization and data analytics and their applications to industrial design problems. Dr. Krishna Rajan and Karin Haake (both HRI-EU) contributed talks on intellectual property including practical information on how to write a patent (compared to writing a scientific paper) and on advanced project management in an industrial research institute. Both talks were followed by a full day workshop dedicated to patenting praxis and project management for an intensive practical training given to all ESRs.
In the second event, which was organized as an online event due to the constraints imposed by the Covid-19 pandemic impact, internationally renowned scientists in optimization and learning were invited as speakers. This very special line-up of speakers was only possible in an online event. Speakers were: Prof. Yew-Soon Ong (Nanyang Technical University, Singapore), Prof. Yaochu Jin (University of Surrey, UK) and Prof. Kaisa Miettinen (University of Jyväskylä, Finland), who each gave a talk on their current cutting-edge research in the field of evolutionary, multi-objective optimization combined with machine learning. This summer school has been carried out in combination with the HRI-EU EGN symposium, a bi-yearly event where all members of the EGN are invited to attend talks by all current PhD-students, which typically leads to lively discussions on actual research. On this occasion, all ESRs presented their research to a broader audience and received valuable feedback for their progress.

NEC:
NEC led organization of the activities, including Winter School on Deep Learning (January 2020), Workshop on Mining Patterns from Complex Data (May 2020), and an international workshop on Learning and Mining with Industrial Data (November 2019). The winter school introduced diverse recent research advances in machine learning, including Reinforcement Learning and its Application to NLP Tasks, Graph Neural Networks, Deep Robot Vision, and Computer Vision with TensorFlow. The workshop also provided the ESRs with a training about deep learning platforms. Besides the technical part, the winter school organized the sessions about soft skills training for career development of the ESRs, such as speed reading, commercial potential of their research, and technology transfer. The workshop on Mining Patterns from Complex Data was held virtually due to the pandemic. The advances in the topics of Class Imbalance Learning, AI for the Biomedical Domain, and Complex Industrial Data Analysis, as well as other Machine Learning Applications were introduced to the ESRs. As regular, the ESRs also reported the new results and discussed the challenges with the supervisory board. Finally the consortium discussed the management matters, especially the upcoming deliverables, the potential impact of the pandemic on the project progress and the secondment of the ESRs. Last but not least, NEC led the organization of the international workshop on Learning and Mining with Industrial Data. The workshop is not only for the ESRs of the ECOLE project, but aimed to bring together researchers and practitioners from academia and industry to discuss challenges, emerging topics, and recent advances in Experienced-based Computation: Learning to Optimize. The workshop was selected to hold in conjunction with the top-tier data mining conference: IEEE International Conference on Data Mining (ICDM 2019). It effectively disseminated the results of the project to the community, and increased its impact in ML/DM/Optimization areas. There were >20 submissions, and >80 authors from more than 10 countries.

The ESRs participated in the organization work, and gained hands-on experience on organization, management and communication skills.

5. Professional Skill Development

Professional skill development for ESRs is covered from a whole range of activities, including (1) formal training course such as the Research Skill Training modules (RST) and the Personal and Career Skill Training modules (PCST). These are explained in Section 4 of this document and in Deliverable 4.1. (2) Skill development is embedded in structured training and monitoring of student (ESR) research progresses at the Universities of Birmingham and Leiden, e.g., through thesis groups and their regular meetings, where the students have to develop their presentation skills, communication skills, research skills, project management skills, time management skills, critical thinking, etc. (3) Professional skill development is also inherently embedded in frequent supervision meetings between supervisors and students, where students have to communicate with both academic and industrial supervisors and in return get different perspectives on his/her research work. Such three-legged approach has been shown to be very effective in supporting our ESRs in their professional skill development.

LS-Dyna Training

Continuous optimization problems in real-world application domains, e.g., mechanics, engineering, economics and finance, can encompass some of the most complicated optimization setups. Principal obstacles in solving the optimization tasks in these areas involve multimodality, high dimensionality, and unexpected drifts and changes in the optimization setup. Due to these obstacles and the black-box assumption on the optimization setup, traditional optimization schemes, e.g., gradient descent and Newton methods, are rendered inapplicable. The majority of the optimization schemes applied in these areas now focus on utilizing direct-search methods such as Evolutionary Algorithms (EAs) and Surrogate-Assisted Optimization (SAO). The purpose of this ECOLE training was to familiarize the ESRs with important issues faced in real-world scenarios, with a special emphasis on computational efficiency. The training included a comprehensive lecture and tutorial series lasting for five days in total. A particular emphasis was given to the software utilization for addressing issues in practical scenarios. Following topics were given special attention in the training:

1. Simulation Technology (Workflow and Fundamentals)
2. Direct-search Methods and Applications (with focus on EAs)
3. Robustness and Sensitivity Analysis

The training component devoted to “Simulation Technology” encompassed issues such as meshing, efficiency, and (numerical) modeling and approximation. The component related to “Direct-search methods and applications” provided a concise overview on different heuristics for tackling optimization problems in practice. Since unexpected drifts and changes in the optimization setup can affect the quality of the optimal solution, the training also included “Robustness and Sensitivity Analysis” component. The aim of this part was to provide a brief overview on techniques to handle uncertainty and noise in the optimization setup. Three practical examples, namely “Target-shape design Optimization”, “Topology Optimization”, and “Crashworthiness Optimization” were used to demonstrate the important concepts in depth. The skillset learned from lectures and tutorial will enable the ESRs to utilize “Ansys LS-Dyna” in their future research related to engineering and ICT sectors.

BMW Training Event

The belated (due to COVID-19) ECOLE BMW training is scheduled to take place from 22nd to 26th November, 2021 in Munich, Germany. Apart from the scientific communications and research activities, all participating ESRs will join the following activities:

1. A guided factory tour focusing on:
    - BMW production, such as body, drivetrain and final assembly
    - Research and Development department
        - Simulation-based car development
        - Optimisation and data analytics in car development
2. Software Trainings
3. Networking and Career Events

The purpose of the guided tour is to enable the ESRs to have a deeper understanding of the industrial application, especially in the field of automotive industry, in order to help them realise the commercial potentials of their research in the near future. The ESRs will be equipped with aerodynamic software skills, whereas the networking and career events will provide the ESRs with the opportunities to expand their interpersonal connections with the industry and enhance their professional competitiveness.

6. ECOLE Management

COVID-19 has posed a big challenge to our project because ESR's mobility has been affected significantly due to travel restrictions. Some of the physical secondments could not be implemented as originally planned. Fortunately, our beneficiaries were able to be flexible and allowed for virtual online secondments. Our research has not been affected in a major way by COVID-19 because most of ESR's work could be done virtually online. This can be seen from the large number of publications generated by ESRs. As shown by the description of WP1-3 earlier in this document, we have achieved all our research goals of work packages.
In terms of project coordination and management, the four beneficiaries of the ECOLE project have worked closely as a team, meeting online every month since the outbreak of COVID-19. These are in addition to our six-monthly workshops and meetings for everyone in the project.

APPENDIX A – PhD Thesis Outlines

1. Thiago Rios (ESR1): Learning-based Representations of High-dimensional CAE Models for Automotive Design Optimization

In design optimization problems, engineers typically handcraft design representations based on personal expertise, which leaves a fingerprint of the user experience in the optimization data. Thus, learning this notion of experience as transferrable design features has potential to improve the performance of similar, yet more challenging, design optimization problems. However, engineering design data are unstructured, high-dimensional and often have no canonical representation, which poses several challenges for machine learning algorithms. Here, we utilize geometric deep learning techniques, in particular 3D point cloud autoencoders, to learn novel shape-generative models from engineering optimization data. We demonstrate that these autoencoders are scalable to high-dimensional engineering models and have comparable optimization performance to state-of-the-art representations. Furthermore, we propose a novel network feature visualization technique that provides a geometric interpretation of the knowledge stored in the network and allows one to select sub-sets of degrees of freedom to modify and optimize shapes. We exploit the agnosticism of the autoencoders' latent space to describe designs assigned to multiple optimization tasks and transfer knowledge between problems, which accelerated the optimizations and fostered commonality between the optimized shapes. Finally, based on a state-of-the-art point cloud autoencoder, we propose a novel deep-generative network, which exploits both data-driven and free-form deformation parameters handcrafted by human designers to generate simulation-ready representations, improving both, the computational efficiency and quality of the results obtained in a set of vehicle aerodynamic design optimization problems.

PhD-Thesis Outline:

1. Introduction
   - 1.1. Motivation
   - 1.2. Objectives and Research Questions
   - 1.3. Outline
2. Learning-based Automotive Design Optimization
   - 2.1. Design and Optimization in Automotive Engineering
   - 2.2. Learning Engineering Data
   - 2.3. Knowledge Transfer in Design Optimization
3. Geometric Representations for Shape Optimization Problems
   - 3.1. Geometric Data
   - 3.2. Design Representations
   - 3.3. Criteria for Selecting Design Representations
4. Learning on CAE 3D Point Cloud Data
   - 4.1. Related Work
   - 4.2. Point-based 3D Point Cloud Autoencoders
   - 4.3. Assessment of the Autoencoder Performance in Shape Optimization Problems
   - 4.4. Visualizing the Network Features
5. Fostering Knowledge Transfer with 3D Point Cloud Autoencoders
   - 5.1. Knowledge Transfer in Multi-task Evolutionary Optimization
   - 5.2. Information-theoretic Analysis of the Latent Representations
6. Bridging Morphing Techniques to 3D Point Cloud Autoencoder Representations
   - 6.1. Surface Reconstruction of Unorganized Point Clouds

6.2. Point2FFD: A Novel Deep-generative Model based on Engineering Experience
6.3. Benchmarking Point2FFD
6.4. Vehicle Aerodynamic Optimization using Point2FFD
7. Summary and Outlook

2. Sneha Saha (ESR2): Multi-Criteria Preference-Aware Design Optimization of 3D Designs
In the early design phase of automotive digital development, engineers exploit their experience to address the challenges in design ideations and optimization problems to develop a new prototype. However, engineers may profit from synergies hidden in history design data from past optimizations to consider multi-criteria for designing a prototype and simulating accurate performance estimations of this prototype. Further, engineers also need to use expertise to manually select design features and domains for engineering optimization. However, since most of the design processes and computer simulations currently rely on virtual models developed with CAE tools, the design of multiple products generates sufficient digital data. In our research, we imagine a cooperative design system in the automotive domain that can learn from these digital data and engineer's expertise to provide assistance to the engineers for several tasks in the design process, e.g., design ideations considering multiple criteria and performance estimations of design prototypes. We address three main aspects for the realization of such a system through our research. First, we propose methods to capture the engineer's expertise approach and provide guidance during the design process. Second, recent advancements in deep learning methods show promising results for learning and compressing 3D data allowing engineers to generate a low-dimensional representation of the 3D design. We propose our deep-generative variational autoencoder model for vehicle data that learns low-dimensional representations of the design data in an unsupervised fashion. Finding representations in a data-driven fashion results in representations that are agnostic to downstream tasks performed on these representations and are believed to capture relevant design features. Additionally, we evaluate this data-driven representation for generating realistic novel shapes, surrogate modeling for performance predictions, and multi-criteria-based optimization tasks. We explored the feasibility of this learned representation for multi-objective optimization and exploited the learned knowledge to improve convergence faster to optimal solutions, to address better the engineer's preference. Thus, in sum, we address three important aspects of building a design assistance system in the automotive domain through our research.

PhD-Thesis Outline:
1. Introduction: This chapter explains the motivation of our research, research questions with an overview of each of the chapters of the thesis.
2. Related Literature: This chapter includes an overview of prior research of 3D design data, deep learning models for learning 3D data and preference-aware design optimization.
3. Data generation from a 3D design task: This chapter addresses methods for understanding the level of experience of the designers and different types of data generated by modifying 3D designs.
4. Learning 3D design data using predictive models: This chapter introduces the two deep learning models – first, we introduce a recurrent neural network to learn from 3D sequence data and, second, a deep generative variational autoencoder to learn from 3D representations.
5. Preference-aware design proposal generation: This chapter focus on considering multi-criteria and engineer's preference for a real-world inspired design optimization scenario. We analyzed the feasibility of the low-dimensional representation of the autoencoder model for multi-objective optimization tasks and propose a method to improve convergence of optimization algorithms.

6. Multi-criteria performance analysis of the generated shapes: In this chapter, we address methods for improving the information content in latent space of the autoencoder models to build surrogate model and propose data augmentation method for improving the performance of surrogate models.
7. Conclusion and outlook

3. Sibghat Ullah (ESR3): Model-Assisted Robust Optimisation for Continuous Black-Box Problems
Continuous optimisation problems in real-world application domains e.g., mechanics, engineering, economics and finance, can encompass some of the most complicated optimisation setup. Principal obstacles in solving these problems involve multimodality, high dimensionality, and unexpected drifts and changes in the optimisation setup. This thesis is dedicated to efficiently solve expensive-to-evaluate black-box problems in the face of uncertainty and noise. To gain efficiency, the so-called "Surrogate Modeling" approach is employed in this thesis. Fundamental research questions regarding the suitability of surrogate modeling are answered with the help of empirical studies in the first part of the thesis. The empirical studies in the first part take into account the impact of several external factors such as the problem landscape, dimensionality, noise level and modeling technique. The second part of the thesis emphasizes on the so-called "Bayesian Optimisation" algorithm, computational cost of robustness and practical applications. In particular, the so-called "Moment-Generating Function of the Improvement" is extended to the robust scenario for Bayesian Optimisation algorithm. The computational cost of robustness analyses the algorithmic time complexity for achieving robustness in Bayesian Optimisation. Consequently, efficient robustness criteria are recommended to practitioners for practical scenarios. All in all, the thesis emphasizes on analyzing, developing and comparing techniques for addressing noisy black-box problems with a special regard to computational efficiency.
PhD-Thesis Outline:
Part I – Fundamentals

1. Introduction
   - 1.1. Robust Optimisation – A Short Guide
   - 1.2. Aims and Objectives
   - 1.3. Organization and Contributions
2. Background
   - 2.1 Robust Optimisation
     - 2.1.1. Uncertainties and Noise in Optimisation Problems
     - 2.1.2. Cases of Uncertainty and Noise
     - 2.1.3. Robustness in Continuous Black-Box Optimisation
   - 2.2. Surrogate Modeling
     - 2.2.1. Introduction
     - 2.2.2. Response Surface Models
     - 2.2.3. Kriging Models
   - 2.3. Summary and Discussion
3. Model-Assisted Robust Optimisation
   - 3.1. Robust Optimisation via Surrogate Modeling
     - 3.1.1. Sampling Plans
     - 3.1.2. Preparing Data and Choosing a Modeling Approach
     - 3.1.3. Parameter Estimation and Training
     - 3.1.4. Model Testing
     - 3.1.5. Empirical Comparison of Modeling Approaches for Robust Optimisation
     - 3.1.6. The "Curse of Dimensionality" and How to Avoid It

3.2. Summary and Discussion

Part II – Advanced Concepts

4. Robust Bayesian Optimisation
   - 4.1. Bayesian Optimisation
     - 4.1.1. Infill Criteria for Bayesian Optimisation
     - 4.1.2. Extending Bayesian Optimisation for Problems with Uncertainty
     - 4.1.3. Empirical Comparison of Infill Criteria for Problems with Uncertainty
   - 4.2. Summary and Discussion
5. Cost of Robustness
   - 5.1. Computational Cost of Robustness
     - 5.1.1. Empirical Investigation on Computational Cost of Robustness
   - 5.2. Optimality vs Robustness
     - 5.2.1. Empirical Investigation on Optimality as Cost of Robustness
   - 5.3. Summary and Discussion
6. Engineering Applications
   - 6.1. Real-World Engineering Case-Studies
     - 6.1.1. Aero-foil Design
     - 6.1.2. The Nowacki Beam
     - 6.1.3. Target-Shape Design Optimisation
   - 6.2. Summary and Discussion
7. Conclusions and Outlook

4. Duc Anh Nguyen (ESR4): Automatic algorithm configuration for parameter tuning of modelling and optimization algorithms

A plethora of algorithmic strands exists for applying machine learning to real-world problems. However, for efficient implementation of these algorithms, the practitioners have to make several high-level decisions such as choosing which approach to use on a given data set, whether and how to preprocess data and tuning hyperparameters. Therefore, in order to efficiently apply the machine learning techniques in solving real-world problems, we focus on the research topic of how to automatically select the optimally suited algorithm and its optimized hyperparameter settings for an arbitrary real-world problem from a given portfolio of the existing works and achievements with less human effort. In other words, we take care of the final step of every research line to make the research achievements accessible to the non-expert. We exploit the recent advancements in Automated machine learning (AutoML) for solving these problems. First, we propose a method to automatically address the class imbalance problems. The problem is tackled with the help of Tree-structured Parzen Estimators (TPE), a special class of Bayesian Optimization (BO) algorithms. Furthermore, we proposed a novel sampling approach to maximize the coverage of the AutoML search space already at the stage of the initial sampling of BO to improve the performance of BO used for AutoML optimization. Finally, based on the proposed sampling approach, we proposed a new robust BO variant for the AutoML optimization problem and a new robust AutoML framework where the proposed approach is used as the underlying optimization approach.

PhD-Thesis Outline:

1. Introduction

1.1. Motivations
1.2. Research questions
1.3. Summary of Contributions
1.4. Outline of the thesis, summarizing each section
2. Background and Related works
2.1 Automated Machine learning optimization problem
2.1.1. Hyperparameter optimization (HPO) problem
2.1.2. Algorithm selection (AS) problem
2.1.3. Automated machine learning (AutoML) optimization problem
2.2 Bayesian optimization
2.2.1. Initial sampling approaches
2.2.2. Probabilistic models
2.2.3. Infill criteria
2.2.4. The state-of-the-art algorithms
2.3 Chapter conclusion.
3. Efficient AutoML via combinational sampling
3.1 Introduce the combination sampling approach
3.2 BO4AutoML: A Robust Bayesian optimization approach for AutoML
3.3 RobustAutoML: A Robust AutoML framework
3.4 Experimental Evaluation.
3.5 Chapter conclusion.
4. Divide and conquer strategy for AutoML optimization:
4.1 Contest procedure approach: Hybridizing bayesian optimization and hyperband.
4.2. Racing procedure approach for controlling optimization processes.
4.3 Experimental Evaluation
4.4 Chapter conclusion.
5. On the use of AutoML optimization in real-world applications:
5.1 AutoML for classification problems
5.2 AutoML for Imbalanced classification problems
5.3. Industrial applications
5.4 Chapter conclusion
6. Conclusions.
6.1 Summary and discussion.
6.2 Future work.

5. Jiawen Kong (ESR5): Novel approaches for class imbalance classification

The class-imbalance problem commonly refers to a type of classification problem where the samples in different classes have a gap in terms of the number of samples. This gap will make the classifiers biased towards the majority class(es) and make the minority class(es) to be overlooked, which hinders the performance of the classifiers. However, the underrepresented (minority) class is usually the class of interest in the problem from the application point of view. In real-world classification problems, there are many applications that suffer from the class-imbalance problem, for instance, fault diagnosis, anomaly detection, medical diagnosis and etc. In these problems, it is much more important to correctly identify the minority samples. The price of misclassifying the minority samples would be a huge loss of money in fault diagnosis, an unqualified product in anomaly detection and a person's life in medical diagnosis. Hence, it is significant to improve the class-imbalance classification. In this thesis, we first empirically investigate the efficiency of state-of-the-art resampling techniques. We also study the hyperparameter optimisation on class-imbalance classification. Furthermore, we propose to improve class-imbalance classification by anomaly detection and using the proposed ideas on engineering applications. Finally, we extend our research direction to multi-class imbalanced classification in both benchmarks and applications.

PhD-Thesis Outline:

1. Introduction
    1.1. General Description of the Class-Imbalance Problem
    1.2. Significance of Studying Class-Imbalance Problem
    1.3. Research Questions and Objectives of the Thesis
    1.4. Contributions of the Thesis
    1.5. Overview of the Structure of the Thesis
2. Class-Imbalance Classification
    2.1. Problem Statement (Theoretically)
        2.1.1. Formal Description
        2.1.2. Performance Measures
        2.1.3. Real-World Applications
    2.2. State of the Art
        2.2.1. Cost-Sensitive Learning
        2.2.2. Data-Level Approaches
        2.2.3. Algorithm-Level Approaches
        2.2.4. Ensemble Learning
    2.3. Research Topics in the Literature (Apart from the four main ones in Section 2.b)
        2.3.1. Multi-Class Class-Imbalance Learning
        2.3.2. Data Complexity in the Class-Imbalanced Domain
        2.3.3. Online Class-Imbalance Learning
        2.3.4. Other Topics
3. An Empirical Investigation Comparing Several Oversampling Techniques (PAKDD paper)
    3.1. Introduction
    3.2. Background
        3.2.1. Oversampling Techniques
        3.2.2. Data Complexity
    3.3. Experimental Setup
    3.4. Simulation Analysis and Discussion
        3.4.1. Experimental results for different classifiers
        3.4.2. Correlation between data complexity and classification performance
    3.5. Efficient Oversampling Strategies for Engineering Vehicle Mesh Dataset
        3.5.1. Generation of a Synthetic Data Set
        3.5.2. Results and Discussion
    3.6. Conclusions
4. Hyperparameter Optimisation on Class-Imbalance Problems (SSCI paper)
    4.1. Introduction
    4.2. Background

4.2.1. Resampling Techniques
4.2.2. Hyperparameter Optimisation
4.3. Related Works
4.4. Experimental Setup
4.4.1. Dataset description
4.4.2. Experimental Results and Discussion
4.5. Conclusions
5. Improving Imbalanced Classification by Anomaly Detection (PPSN paper)
5.1. Introduction
5.2. Background
5.2.1. Resampling Techniques
5.2.2. Outlier Factor
5.2.3. Four Types of Samples in the Imbalanced Learning Domain
5.3. Experimental Setup
5.4. Experimental Results and Discussion
5.4.1. Results
5.4.2. Feature Importance Analysis
5.5. Conclusions
6. Paper on multi-class imbalanced classification
7. Conclusions.
7.1 Summary of achievements
7.2 Summary of future work

6. Stephen Friess (ESR6): Predicting Inductive Biases for Search-based Optimization
Within recent years, knowledge transfer research in population-based optimization has gained significant traction. Notable work takes inspiration from research on transfer learning as a way to exploit complementarities between problems. However, these approaches suffer from a lack of understanding on what constitutes problem similarity in the first place. Thus, it is elusive in regard to what contributes to performance improvements in a knowledge transfer scenario. Uncontrolled, it has the capability to jeopardize algorithm performance in what is dubbed as 'negative transfer'. The objectives of our research are therefore the following: First of all, we want to clarify on what can be considered to be knowledge within population-based algorithms in the first place. And second, we want to understand how it can be consolidated and harnessed in a domain-dependent manner. We find within our research, that essentially optimization algorithms need to form inductive biases. These can be for instance represented through search operators which can be problem-tailored by incorporating information about the problem-structure. Secondly, we need a way of arbitration between operators. Such can be done through methodology from algorithm behavior studies. We investigated such an approach and showed that graph structures can particularly realize performance gains in a paper accepted at IEEE IJCNN 2021. In principle, for application scenarios constructing both, operators and a predictor is problematic for application scenarios. However, one can reframe it as a problem of predicting operator configurations. We demonstrate the efficacy so this approach in a scenario with an application to shape optimization.

PhD-Thesis Outline:
A. Abstract

B. Acknowledgments
1. Introduction
1.1 The Contemporary State of Evolutionary Computation
1.2 The Value of Knowledge in Real-World Optimization Problems
1.3 Thesis Aim, Structure & Contributions
2. Literature Review
2.1 Foundations from Adaptive Systems Research
2.2 Computational Models of Problem-Solvers
2.3 Open Questions
3. Inductive Biases in Search-based Optimization
3.1 Introduction / Structure
3.2 Generating Statistics from Algorithm Runs
3.3 Discussion of Methods for Operator Design
3.4 Model Selection and Hyperparametertuning
3.5 Properties of the Learned Search Operators
3.6 On the Efficacy on Benchmark Problems
3.7 Investigation of Problem and Statistical Similarity with Performance
3.8 Incorporating Mechanisms for Self Adaption
3.9 Upscaling to High Dimensional Problems
4. Feature Extraction from Optimization Problems
4.1 Introduction / Structure
4.2 Construction of Search Space Partitions
4.3 Post-Processing of Optimization Data
4.4 Generation of Optimization Data from Algorithms
4.5 Comparison and Trade-Offs of the Different Approaches
4.6 Extended studies
5. Predicting Operators for CMA-ES Algorithms
5.1 Introduction
5.2 The CMA-ES Algorithm
5.3 The FFD Technique
5.4 Design Optimization Problems
5.5 Operators in Design Optimization Problems
6. Conclusions and Outlook
6.1 On the Feasibility of End-to-End Learning for Search-based Optimization
6.2 Knowledge Transfer: Knowing When and How to Transfer Knowledge
6.3 Outlook for Future Research

7. Gan Ruan (ESR7): Knowledge transfer in evolutionary dynamic multi-objective optimization
There are many real-world dynamic multi-objective optimization problems (DMOPs) where multiple conflicting objectives may change over time. It is therefore significant to study how to solve them. However, most existing evolutionary algorithms for solving DMOPs neglects the useful experience or knowledge obtained when solving past problems. Recently, people firstly apply transfer learning for solving DMOPs, named as transfer learning-based dynamic multi-objective evolutionary algorithm (Tr-DMOEA). However, it has been found that transfer learning does not always help in Tr-DMOEA and the used Gaussian kernel function is not ideal. Therefore, we try to answer the research question of when and how to transfer knowledge in Tr-DMOEA and proposed an improved version of Tr-DMOEA in a

paper accepted by IEEE SSCI 2019. In addition, it has been also found that Tr-DMOEA is extremely computational cost. We did computational study on how to improve the efficiency and effectiveness of Tr-DMOEA in a paper accepted by IEEE CEC 2020. In DMOPs, the number of objectives may also vary over time. They are called DMOPs with a changing number of objectives. The existing algorithm for solving DMOPs with a changing number of objectives is unable to provide enough diversity for DMOPs with a changing number of objective where the problem property is complex. Therefore, a knowledge transfer-based technique was proposed to improve the diversity while maintain the convergence of a population. This work is under preparation with experimental studies. Through literature review, it has also been found that most existing DMOEAs using transfer learning algorithms always regard solutions found in the past problem as the only knowledge to be transferred. However, the experience of search for optimal solutions for the problem has not been explored and taken into consideration. Therefore, in the rest time of PhD period, we intend to study on what experience to be extracted from the search process, when and how to transfer the search experience for solving DMOPs.

PhD-Thesis Outline:

1.Introduction
- 1.1. Motivations
- 1.2. Research Questions
- 1.3. Contributions
- 1.4 Gaps in the Current Literature
- 1.5 List of Publications Generated from the Thesis

2. Background and Literature Review
- 2.1. Dynamic Multi-objective Optimization (DMO)
- 2.2. Knowledge Transfer in Evolutionary Computation

3. Knowledge Transfer for Dynamic Multi-objective Optimization with Changing Pareto Set (PS) Position and/or Pareto Fronts (PFs)
- 3.1. When and How to Transfer Knowledge in Transfer Learning-based Dynamic Multi-objective Evolutionary Algorithms (Tr-DMOEAs)
- 3.2. Improve the Efficiency of Knowledge Transfer in Tr-DMOEA

4. Knowledge Transfer for Dynamic Multi-objective Optimization with a Changing Number of Objectives
- 4.1. Expand/Contract PS for DMO with a Changing Number of Objectives and Fixed PS Position
- 4.2. Transfer Knowledge to Tackle DMOPs with a Changing Number of Objectives and Changing PS Position and/or PFs

5. Conclusion and Future Work

8. Giuseppe Serra (ESR8): Explainable models for preference learning with user-generated data

The focus of this work is on developing statistical machine learning and text mining methods for automatically detecting customers' opinions on aspects and features of products, and for analyzing users' behaviors starting from large collections of unstructured data. Since user-generated data can be of different forms (e.g. texts, time series), we aim at developing different ideas considering the different nature of the data. With success of ML/AI in a variety of applications, they have been pervasive in diverse vertical sectors. Trustworthiness of ML/AI is thus an emerging concern, and the corresponding techniques are being investigated to give users the confidence in embracing AI-based solutions. To engender trust in AI systems, they should firstly be explainable. We aim at studying explainable AI (XAI) approaches to solve the problem in the context of personalized optimization for product design. In the context of the ECOLE project, the final target is to incorporate the valuable information contained in the user-generated into industrial applications. Ideally, we could use this information to provide useful insights for product feature optimization. For example, in a product design problem, we could utilize user design preferences extracted from the textual data for adding some aesthetic constraint to the multi-criteria optimization framework, to create a new design that could potentially satisfy users more than before.

1 Introduction
- 1.1 Project Aim
- 1.2 Motivations

2 Background
- 2.1 Introduction
- 2.2 Interpretable AI
- 2.3 Machine learning for user-generated data

3 Research questions
- 3.1 Can we improve the interpretability of common representation learning techniques (e.g. embedding) by exploiting human-understandable information, such as textual data?
- 3.2 Can we introduce some metrics to help assessing the quality of the results in terms of textual explainability?
- 3.3 In the context of recommendation tasks, given the nature of collaborative filtering (CF) data, do we really need to employ 'deep' architecture and complicated representations?
- 3.4 Does the interpretability constraint affect the learning task performances?

4 Interpreting Node Embedding with Text-labeled Graphs
- 4.1 Introduction
- 4.2 Problem statement
- 4.3 Objectives
- 4.4 Related works
- 4.5 Proposed solution
- 4.6 Experiments
- 4.7 Conclusions and future works

5 Product Rating Prediction through Interpretable Latent Class Modeling of User Reviews.
- 5.1 Introduction
- 5.2 Problem statement
- 5.3 Objectives
- 5.4 Related works
- 5.5 Proposed solution
- 5.6 Experiments
- 5.7 Conclusions and future works

6 Explainers for Graph Neural Networks
- 6.1 Introduction
- 6.2 Problem statement
- 6.3 Objectives
- 6.4 Related works
- 6.5 Proposed solution
- 6.6 Experiments
- 6.7 Conclusions and future works

7 Learning Sparsity of Representations with Discrete Latent Variables
7.1 Introduction
7.2 Problem statement
7.3 Objectives
7.4 Related works
7.5 Proposed solution
7.6 Experiments
7.7 Conclusions and future works
8 Summary and conclusions
8.1 Future works