From Fitness Landscapes to Explainable AI and Back

Sarah L. Thomson
University of Stirling
Stirling, United Kingdom
s.l.thomson@stir.ac.uk

Jason Adair
University of Stirling
Stirling, United Kingdom
jason.adair@stir.ac.uk

Alexander E.I. Brownlee
University of Stirling
Stirling, United Kingdom
alexander.brownlee@stir.ac.uk

Daan van den Berg
VU University
Amsterdam, Netherlands
d.van.den.berg@vu.nl

ABSTRACT

We consider and discuss the ways in which search landscapes might contribute to the future of explainable artificial intelligence (XAI), and vice versa. Landscapes are typically used to gain insight into algorithm search dynamics on optimisation problems; as such, it could be said that they *explain* algorithms and that they are a natural bridge between XAI and evolutionary computation. Despite this, there is very little existing literature which utilises landscapes for XAI, or which applies XAI techniques to landscape analysis. This position paper reviews the existing works, discusses possible future avenues, and advocates for increased research effort in this area.

KEYWORDS

Fitness Landscapes, Search Landscapes, Neural Networks, Explainable AI, XAI

ACM Reference Format:
Sarah L. Thomson, Jason Adair, Alexander E.I. Brownlee, and Daan van den Berg. 2023. From Fitness Landscapes to Explainable AI and Back. In *Genetic and Evolutionary Computation Conference Companion (GECCO '23 Companion), July 15–19, 2023, Lisbon, Portugal.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3583133.3596395

1 INTRODUCTION

Size matters, and the maturation of artificial intelligence (AI) in recent years is arguably due to the rapid increase in processing and storage size, alongside increasing quantity and quality of available data. With the spectacular ascent of AI through the ranks of human labour came discussion surrounding *explainable* AI — pertaining to intuitive and accessible interpretations, mostly for machine learning (ML) models [3, 11, 12], but simultaneously triggering a paradigm shift within evolutionary computation (EC) circles [2, 25] and other AI communities. As an example: when a model analyses a patient's medical data and diagnoses a case of lung cancer, we want to know *how* — for both humanitarian and legal reasons — before accidentally administrating chemotherapy to treat a common cold.

GECCO '23 Companion, July 15–19, 2023, Lisbon, Portugal

ACM ISBN 979-8-4007-0120-7/23/07.
https://doi.org/10.1145/3583133.3596395

Fitness landscapes in EC are both a mathematical model and a visual metaphor used to analyze and indeed, *explain*, the interplay between an algorithm and an optimization problem instance [30, 41]. Providing insights into algorithmic decision-making, landscapes can serve as a 'bridge' between XAI and evolutionary computing [2, 45] but there have been very few studies which explicitly do so.

From here, we take the position that there are two directions in which landscapes and XAI can be combined: landscapes for improving or implementing XAI, or XAI for improving or understanding landscape analysis. We will elaborate in both directions, alongside some discussion of earlier work insofar as available.

2 LANDSCAPES & XAI

We begin with key definitions and context. A **fitness landscape** [40] is composed of three parts: (S, N, f) : S is the full set of possible solutions; N is the neighbourhood function, which assigns a set of adjacent solutions $N(s)$ to every $s \in S$ and $f : S \longrightarrow \mathbb{R}$ is a fitness function that provides a mapping from each solution to exactly one numerical fitness value (see Figure 1). There are many methods for XAI, and we do not discuss all of them here (and refer the interested reader to a recent survey [15]). Instead, we now focus on a subset: a few popular approaches which seem particularly relevant viewed through the lens of landscape analysis.

Shapley Additive Explanations (usually referred to as **SHAP**) [22] are a prevalent XAI method [1, 16, 48] which estimate the contribution of features to a prediction. SHAP works by training

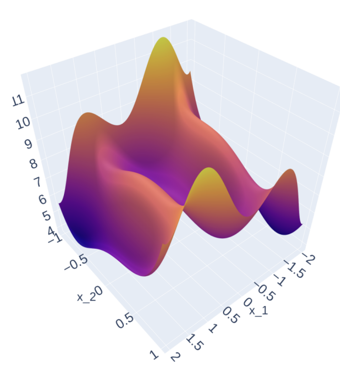

Figure 1: A typical fitness landscape; the two optimization variables are in the plane and the landscape's height denotes the fitness value. Adapted from [28]

models using different sets of features. The *marginal contribution* of a feature — for a particular observation — is obtained by subtracting the prediction of a model which *excludes* that feature from the prediction of the same model which *includes* the feature. Marginal contributions of the feature across all models which contain it are added together — resulting in a SHAP value for the feature-observation pair. The higher the absolute SHAP value, the more important the associated feature is taken to be in the model under study. While SHAP values constitute local explanations (that is, for a single prediction only), these can be aggregated for a set of observations to provide a global model explanation.

Local interpretable model-agnostic explanations (LIME) [33] is a common method for XAI with black-box models [9, 19, 23]. Given a single data observation, LIME samples slightly perturbed versions of that and uses the original model to predict the response for them; after that, LIME fits a separate linear model using this data. The associated coefficients are taken to be feature importances for the original model; these serve as explanations.

Counterfactual analysis [18] is a further XAI approach which could be used to better understand fitness landscapes and the problem-algorithm mechanics they capture. Counter-factual explanations take the form "if X had not occurred, then Y would not have occurred either." Returning to the lung cancer scenario: the system could indicate "You do not have lung cancer, because 19 patients with data very similar to yours that did have cancer all had platelet counts of over $456,784 \times 10^6/L$, while yours is only $276,004 \times 10^6/L$, meaning your probability of having lung cancer is 0.09%."[1]

3 LANDSCAPES FOR XAI

The notion of a **loss landscape** [20] is used to describe the topology of the parameter space which a neural network is learning to minimise the loss function. In this way, they can be used to understand the relationship between learned parameters and the quality of the model. Rather than gaining insight into an already-constructed single model, the landscape is in this case being used to explain the machine learning process *itself.* In this sense, loss landscapes are good at providing some explanation to practitioners of the learning process, and how to navigate it [6, 20]. In the literature, the notion of a loss landscape seems to have been applied exclusively in the context of neural networks. There is probably untapped potential in modeling other learning algorithms in this manner (constructing a landscape from mappings between parameters and loss). One example is logistic regression, where stochastic gradient descent can be used to minimise error. Although logistic regression has fewer parameters to learn than a neural network does, the fact that its parameters are weights (coefficients) for individual features may open the door for insight about the machine learning model and about the features in the dataset itself.

Fitness landscapes have been deployed previously to analyse **neural architecture search** spaces [29, 31]. The landscapes were found to be straightforward and well-suited to iterated local search. This research closely neighbours landscapes for XAI: it provides an understanding of the connection between a neural network architecture and the subsequent model quality for a given dataset. Instead of providing human-readable explanations for a particular model (as is the case in typical XAI), this type of analysis can provide numerical and visual explanations for what constitutes a good network topology for a dataset. One particularly poignant example is the work on weight-agnostic neural networks where evolutionary algorithms produce highly irregular networks of variable size [10]. The authors had intended to find as-small-as-possible and highly functional networks for simulated tasks such as controlling the movement of a bipedal walker (Figure 2). It is unclear how network size relates to the movement of the robot, and fitness landscape analysis could play a critical role in understanding this.

Another variation on the neural network theme was witnessed in the development of **neuroevolution trajectory networks** (NTNs) [37]. NTNs have recently been applied to uncover and explain the process of transfer learning in deep neural networks [36]. In this work, the authors used neuroevolution to train and refine a model on a series of MNIST datasets of increasing complexity; using NTNs, they were able to provide visual representations of which of the final model's hyper-parameters were derived from each dataset.

A **machine learning pipeline** is the sequence of processes which are carried out in order to perform modelling with a dataset. Machine learning pipelines have been formulated as optimization problems and subsequently been analysed with fitness landscape analysis in the literature [42, 43]. This facilitates the possibility of explanations for what the pipeline is doing. As one example: what if a the performance of a pipeline is *rugged* in its hyperparameters? It would explain why very similar pipelines have very different behaviour.

A powerful combination of landscape analysis and XAI is using landscapes to understand **feature selection and feature set composition**; that is, modelling binary-encoded feature sets as solutions in an optimisation problem, with model quality as fitness. One study has already made steps in this direction, finding high amounts of neutrality in the landscape — indicating that many possible model configurations contain redundant features for the studied datasets [26]. Another direction along these lines is the estimation of distribution algorithms (EDAs), which maintain explicit probabilistic models of estimates for good components of solutions [13]. In one study, an EDA was used as a means for feature selection [27]; in another, fitness landscape analysis has been conducted for EDAs [24]. As far as we know, landscape analysis has not yet been conducted for EDAs with the feature selection problem. This approach could bring enhanced understanding of which (combinations of) features are associated with high machine learning model quality. For example, the landscape analysis might find that there are mostly gentle gradients on the evolutionary trajectory. From an XAI perspective, this could imply that similar good-quality feature sets are associated with similar model quality. The joint distribution over the features represented by the probabilistic model which the EDA keeps could also provide valuable data about which features interact well together. As an example, in the feature selection problem for a model predicting someone's lifetime breast cancer risk, an EDA might capture the fact that the interaction between "feature var_1 = *smoker*" and "feature var_{23} = *family history*" is important for model accuracy.

One study on **evolving rule sets** for predictive models [38] opens up the path towards modelling the rule set landscape. This

[1]Numbers are fictional

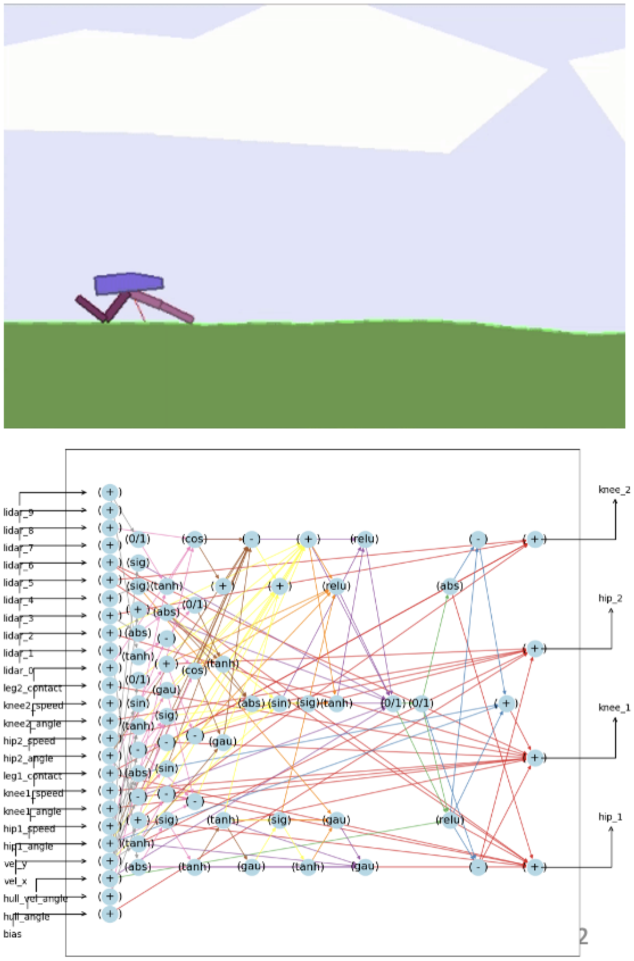

Figure 2: Neural architecture search for the problem of 'teaching' a bipedal entity to walk can be done with an evolutionary algorithm. Image by Mostafa Doroodian, adapted from work by Gaier & Ha [10]

could bring insight into important components, hierarchical inferencing, or domination in good rule sets. The results could be considered XAI, although the explanatory analysis pertains to the model construction process instead of the final model. Furthermore, pitfalls such as overfitting are often present in these systems, which could leave their mark on the associated landscapes as well. Another possible way forward in using landscape analysis for XAI is visualizing multi-objective modelling processes as fitness landscapes. For example, a practitioner might want to balance the possibly **conflicting objectives** of having a model with maximum accuracy but with an architecture requiring low computational cost. This could happen in a healthcare setting, where hardware capacity may be limited but the accuracy of the model is extremely important. This balance of objectives could be modelled using multi-objective optimization and the pareto front of solutions (models) consequently visualised. This could help both practitioners and stakeholders understand the modelling process.

During the execution of SHAP, many different model configurations (perturbed versions of the original model) are tried out in order to establish feature importances. Model configurations (i.e., the feature sets) could be taken as candidate solutions and their quality as fitness. The landscape of model configurations could then be analysed and visualised in order to understand what SHAP is uncovering about features more comprehensively.

4 XAI FOR LANDSCAPES

Machine learning models have been employed several times for exploring relationships between topological landscape features and optimization performance [7, 21, 46]. Recently, authors have proposed using XAI to understand this type of model by using SHAP [44], but LIME and counterfactual explanations could similarly be applied to this kind of model. Along the same vein, there has also been work towards deliberately minimizing the number of features involved in **performance prediction** with the explicit aim of building somewhat-interpretable models [32].

An optimization problem search space could be modelled using a supervised machine learning approach. The input features would be formed by chromosomes and fitness would serve as response variable. This process of **machine learning the search space** is typically the approach taken to construct surrogate fitness models [17] and in Bayesian optimization [14]. The probabilistic mechanics of EDAs might also be said to be such a model — being biased towards generation of high-fitness solutions. Indeed, the general idea of analysing models to better understand optimization problems was first suggested in the form of mining the probabilistic models of EDAs [34, 35]. Such an analysis was then applied in practice via analysis of coefficients in an interpretable linear probabilistic model of fitness [5]. More recently, the probing of machine learning surrogate models fitted to population data from an EA similar to LIME has been used to quantify the sensitivity of the fitness function to features (variables) [39, 47].

We observe that, in the future, **decision trees** could be used to better understand fitness landscapes. Decision trees are widely agreed to be inherently interpretable because their decision-making process is simply a set of human-readable rules [4, 8]. A decision tree could be built using a sample of candidate solutions to an optimisation problem as training data, with fitness values serving as the labels. After that, the rules learned from the data could be used to gain insight into the landscape; e.g., for the feature selection problem on a cancer dataset: "feature var_4 = *tumour diameter*" and "feature var_{31} = *tumour fractal dimension*" must be present for a feature set to be a local optimum. There is no requirement that the optimisation problem is feature selection, however; it could also be a Travelling Salesperson or Quadratic Assignment instance (or, indeed, any arbitrary optimisation problem). There could potentially be a feature which often presents as the root node in decision trees of the search space, thus emphasising its salience as a component. More broadly, feature importances could be extracted from the decision tree to estimate the effect of different solution components. Counterfactual explanations would also be possible using the "solutions as features, fitness as response variable" model. If the planning of a chemotherapy drug regimen was modelled as an optimisation problem, then an example might be: "if $var_{11} \geq$

= 5.1413", the fitness would not have reached the level of global optima."

5 POSITION

While fitness landscape analysis is becoming common practice in areas such as combinatorial optimisation and evolutionary computing, there is little existing work which applies its powers to XAI. One can assist the other here: landscape analysis could help with XAI, and XAI methods could be deployed to better understand landscapes. Although not all discussed avenues are XAI in the 'traditional' sense of the word, we call upon the community to question, discuss, challenge or oppose this position to further the exploration of fitness landscapes in XAI, and vice versa. We advocate for increased research effort in this area and look forward to your initiatives.

REFERENCES

[1] Miller Janny Ariza-Garzón, Javier Arroyo, Antonio Caparrini, and Maria-Jesus Segovia-Vargas. 2020. Explainability of a machine learning granting scoring model in peer-to-peer lending. *Ieee Access* 8 (2020), 64873–64890.

[2] Jaume Bacardit, Alexander EI Brownlee, Stefano Cagnoni, Giovanni Iacca, John McCall, and David Walker. 2022. The intersection of evolutionary computation and explainable AI. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 1757–1762.

[3] Vaishak Belle and Ioannis Papantonis. 2021. Principles and practice of explainable machine learning. *Frontiers in big Data* (2021), 39.

[4] Dimitris Bertsimas and Georgios Antonios Margonis. 2023. Explainable vs. interpretable artificial intelligence frameworks in oncology. *Translational Cancer Research* 12, 2 (2023), 217.

[5] A.E.I. Brownlee, J.A.W. McCall, and Q. Zhang. 2013. Fitness Modeling With Markov Networks. *IEEE T. Evolut. Comput.* 17, 6 (2013), 862–879. https://doi.org/10.1109/TEVC.2013.2281538

[6] Huanran Chen, Shitong Shao, Ziyi Wang, Zirui Shang, Jin Chen, Xiaofeng Ji, and Xinxiao Wu. 2023. Bootstrap Generalization Ability from Loss Landscape Perspective. In *European Conference on Computer Vision*. Springer, 500–517.

[7] Fabio Daolio, Arnaud Liefooghe, Sébastien Verel, Hernán Aguirre, and Kiyoshi Tanaka. 2015. Global vs local search on multi-objective NK-landscapes: contrasting the impact of problem features. In *proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*. 369–376.

[8] Arun Das and Paul Rad. 2020. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371* (2020).

[9] Freddy Gabbay, Shirly Bar-Lev, Ofer Montano, and Noam Hadad. 2021. A lime-based explainable machine learning model for predicting the severity level of covid-19 diagnosed patients. *Applied Sciences* 11, 21 (2021), 10417.

[10] Adam Gaier and David Ha. 2019. Weight agnostic neural networks. *Advances in neural information processing systems* 32 (2019).

[11] Leonida Gianfagna and Antonio Di Cecco. 2021. Explainable AI: Needs, Opportunities, and Challenges. *Explainable AI with Python* (2021), 27–46.

[12] Hani Hagras. 2018. Toward human-understandable, explainable AI. *Computer* 51, 9 (2018), 28–36.

[13] Mark Hauschild and Martin Pelikan. 2011. An introduction and survey of estimation of distribution algorithms. *Swarm Evol. Comput.* 1, 3 (2011), 111 – 128. https://doi.org/10.1016/j.swevo.2011.08.003

[14] Philipp Hennig, Michael A Osborne, and Hans P Kersting. 2022. *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press.

[15] Andreas Holzinger, Anna Saranti, Christoph Molnar, Przemyslaw Biecek, and Wojciech Samek. 2022. Explainable AI methods-a brief overview. In *xxAI-Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*. Springer, 13–38.

[16] Tom Jansen, Gijs Geleijnse, Marissa Van Maaren, Mathijs P Hendriks, Annette Ten Teije, and Arturo Moncada-Torres. 2020. Machine learning explainability in breast cancer survival. In *Digital Personalized Health and Medicine*. IOS Press, 307–311.

[17] Yaochu Jin. 2011. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm Evol. Comput.* 1, 2 (2011), 61–70. https://doi.org/10.1016/j.swevo.2011.05.001

[18] Mark T Keane and Barry Smyth. 2020. Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (XAI). In *Case-Based Reasoning Research and Development: 28th International Conference, ICCBR 2020, Salamanca, Spain, June 8–12, 2020, Proceedings 28*. Springer, 163–178.

[19] Murat Kuzlu, Umit Cali, Vinayak Sharma, and Özgür Güler. 2020. Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools. *IEEE Access* 8 (2020), 187814–187823.

[20] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. 2018. Visualizing the loss landscape of neural nets. *Advances in neural information processing systems* 31 (2018).

[21] Arnaud Liefooghe, Fabio Daolio, Sébastien Verel, Bilel Derbel, Hernan Aguirre, and Kiyoshi Tanaka. 2019. Landscape-aware performance prediction for evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation* 24, 6 (2019), 1063–1077.

[22] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[23] Pavan Rajkumar Magesh, Richard Delwin Myloth, and Rijo Jackson Tom. 2020. An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. *Computers in Biology and Medicine* 126 (2020), 104041.

[24] Marcella SR Martins, Mohamed El Yafrani, Roberto Santana, Myriam Delgado, Ricardo Lüders, and Belaïd Ahiod. 2018. On the performance of multi-objective estimation of distribution algorithms for combinatorial problems. In *2018 IEEE congress on evolutionary computation (CEC)*. IEEE, 1–8.

[25] Yi Mei, Qi Chen, Andrew Lensen, Bing Xue, and Mengjie Zhang. 2022. Explainable Artificial Intelligence by Genetic Programming: A Survey. *IEEE Transactions on Evolutionary Computation* (2022).

[26] Werner Mostert, Katherine M Malan, Gabriela Ochoa, and Andries P Engelbrecht. 2019. Insights into the feature selection problem using local optima networks. In *European Conference on Evolutionary Computation in Combinatorial Optimization (Part of EvoStar)*. Springer, 147–162.

[27] Geoffrey Neumann and David Cairns. 2016. A targeted estimation of distribution algorithm compared to traditional methods in feature selection. In *Computational Intelligence: Revised and Selected Papers of the International Joint Conference, IJCCI 2013, Vilamoura, Portugal, September 20-22, 2013*. Springer, 83–103.

[28] Dante Niewenhuis and Daan van den Berg. 2020. Making Hard(Er) Bechmark Test Functions. In *IJCCI*. 40–48. https://doi.org/10.5220/0011405300003332

[29] Gabriela Ochoa and Nadarajen Veerapen. 2022. Neural Architecture Search: A Visual Analysis. In *International Conference on Parallel Problem Solving from Nature*. Springer, 603–615.

[30] Erik Pitzer, Andreas Beham, and Michael Affenzeller. 2013. Automatic algorithm selection for the quadratic assignment problem using fitness landscape analysis. In *European conference on evolutionary computation in combinatorial optimization*. Springer, 109–120.

[31] Isak Potgieter, Christopher W Cleghorn, and Anna S Bosman. 2022. A Local Optima Network Analysis of the Feedforward Neural Architecture Space. In *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[32] Quentin Renau, Johann Dréo, Carola Doerr, and Benjamin Doerr. 2021. Towards explainable exploratory landscape analysis: Extreme feature selection for classifying BBOB functions. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*. Springer, 17–33.

[33] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[34] Roberto Santana, Concha Bielza, Jose A Lozano, and Pedro Larranaga. 2009. Mining probabilistic models learned by EDAs in the optimization of multi-objective problems. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*. 445–452.

[35] R. Santana, P. Larrañaga, and J. A. Lozano. 2007. *Challenges and open problems in discrete EDAs*. Technical Report EHU-KZAA-IK-1/07. Department of Computer Science and Artificial Intelligence, University of the Basque Country. http://www.sc.ehu.es/ccwbayes/technical.htm

[36] Stefano Sarti, Nuno Laurenço, Jason Adair, Penousal Machado, and Gabriela Ochoa. 2023. Under the Hood of Transfer Learning for Deep Neuroevolution. *Lecture Notes in Computer Science* (2023), 640–655. https://doi.org/10.1007/978-3-031-30229-9_41

[37] Stefano Sarti and Gabriela Ochoa. 2021. *A NEAT Visualisation of Neuroevolution Trajectories*. 714–728. https://doi.org/10.1007/978-3-030-72699-7_45

[38] Hormoz Shahrzad, Babak Hodjat, and Risto Miikkulainen. 2022. Evolving explainable rule sets. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 1779–1784.

[39] Manjinder Singh, Alexander EI Brownlee, and David Cairns. 2022. Towards explainable metaheuristic: mining surrogate fitness models for importance of variables. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 1785–1793.

[40] Peter F. Stadler. 2002. Fitness landscapes. *Biological Evolution and Statistical Physics. Lecture Notes in Physics* 585 (2002), 183–204.

[41] Mohammad-H Tayarani-N and Adam Prügel-Bennett. 2016. An analysis of the fitness landscape of travelling salesman problem. *Evolutionary computation* 24, 2 (2016), 347–384.

[42] Matheus Cândido Teixeira and Gisele Lobo Pappa. 2022. Analysis of Neutrality of AutoML Search Spaces with Local Optima Networks. In *Intelligent Systems*, João Carlos Xavier-Junior and Ricardo Araújo Rios (Eds.). Springer International Publishing, Cham, 473–487.

[43] Matheus C Teixeira and Gisele L Pappa. 2022. Understanding AutoML search spaces with local optima networks. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 449–457.

[44] Risto Trajanov, Stefan Dimeski, Martin Popovski, Peter Korošec, and Tome Eftimov. 2021. Explainable landscape-aware optimization performance prediction. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 01–08.

[45] Risto Trajanov, Stefan Dimeski, Martin Popovski, Peter Korošec, and Tome Eftimov. 2022. Explainable Landscape Analysis in Automated Algorithm Performance Prediction. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*. Springer, 207–222.

[46] Marco Veluscek, Tatiana Kalganova, and Peter Broomhead. 2015. Improving ant colony optimization performance through prediction of best termination condition. In *2015 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 2394–2402.

[47] Aidan Wallace, Alexander EI Brownlee, and David Cairns. 2021. Towards explaining metaheuristic solution quality by data mining surrogate fitness models for importance of variables. In *Artificial Intelligence XXXVIII: 41st SGAI International Conference on Artificial Intelligence, AI 2021, Cambridge, UK, December 14–16, 2021, Proceedings 41*. Springer, 58–72.

[48] Christopher Yeung, Ju-Ming Tsai, Brian King, Yusaku Kawagoe, David Ho, Mark W Knight, and Aaswath P Raman. 2020. Elucidating the behavior of nanophotonic structures through explainable machine learning algorithms. *ACS Photonics* 7, 8 (2020), 2309–2318.