Interactive Machine Learning and Explainability in Mobile Classification of Forest-Aesthetics

Simon Flutura, Andreas Seiderer, Tobias Huber, Katharina Weitz, Ilhan Aslan, Ruben Schlagowski, Elisabeth André
lastname@hcm-lab.de
Human-Centered Multimedia, University of Augsburg
Augsburg, Germany

Joachim Rathmann
joachim.rathmann@uni-wuerzburg.de
Chair of Geography and Regional Science, University of Würzburg
Würzburg, Germany

ABSTRACT

This paper presents an application that classifies forest's aesthetics using interactive machine learning on mobile devices. Transfer learning is used to be able to build upon deep ANNs (MobileNet) using the limited resources available on smart-phones. We trained and evaluated a model using our application based on a data-set that is plausible to be created by a single user. In order to increase the comprehensibility of our model we explore the potential of incorporating explainable Artificial Intelligence (XAI) into our mobile application. To this end we use deep Taylor decomposition to generate saliency maps that highlight areas of the input that were relevant for the decision of the ANN and conducted a user study to evaluate the usefulness of this approach for end-users.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*; Empirical studies in ubiquitous and mobile computing; • **Information systems** → Personalization.

KEYWORDS

interactive machine learning, artificial neural networks, scenic beauty classification, explainable AI

ACM Reference Format:

Simon Flutura, Andreas Seiderer, Tobias Huber, Katharina Weitz, Ilhan Aslan, Ruben Schlagowski, Elisabeth André and Joachim Rathmann. 2020. Interactive Machine Learning and Explainability in Mobile Classification of Forest-Aesthetics . In *6th EAI International Conference on Smart Objects and Technologies for Social Good (GoodTechs '20), September 14–16, 2020, Antwerp, Belgium.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3411170.3411225

1 INTRODUCTION

A dictum says, there's no accounting for taste. While there is a rough consensus on what is beautiful, it differs for every individual in the long run. This also holds true for the individual taste for landscapes, which can even be correlated to personality traits [1].

GoodTechs '20, September 14–16, 2020, Antwerp, Belgium
https://doi.org/10.1145/3411170.3411225

While traditionally technology and nature are seen as antipodes, mobile technology is pervading our lives, becoming invisible to our accustomed eye. This enables artificial intelligence to guide us in situations where traditional technology would be seen as being disruptive. With these recent trends, machine learning techniques can be integrated seamlessly into our daily routines and even benefit our health, for example by planning routes that contain landscapes which are the most recreational for us. In such contexts, personalized prediction models are desirable. Thus, individual taste and preferences can be taken into account to optimize the desired positive effects for a user. As forests have repeatedly proven their recreational effect in previous studies [17], this paper lays the conceptual and prototypical foundation for applications that use interactive machine learning to personalize the classification of a forest's aesthetic value. Forests vary in purpose and appearance, while there is work on route planning based on forests' walkability [9], a basis for scenic route planning in forests is lacking. To preserve the user's privacy we designed our application to train artificial neural networks (ANNs) locally without sending any private data such as images to remote servers. For model training and continuous user customization, we implemented and evaluated a transfer learning approach. Additionally, this allows using comparably few data amounts for training. We examine the hypothesis that (0) with just one user's photographs, we are able to train a model on a smartphone that is able to distinguish aesthetic from un-aesthetic forest-scenery. Since ANNs typically use millions of trained parameters, their decision process is incomprehensible for the user. To address this problem we explore the potential of augmenting the raw predictions of our classification network with saliency maps, a technique from the research field of explainable artificial intelligence (XAI) that highlights the regions of the input that were relevant for the prediction of the ANN. The interplay between user perception and visualization can vary from one field of application to another, thus evaluation of specific scenarios is advisable. This additional information might help the users to understand the classifier and enable them to choose additional training data to adapt the model to their personal needs, i.e. by including more pictures of objects they find particularly aesthetic. We conducted a user study to verify our hypothesis that this augmentation increases (1) the perceived transparency of the prediction model and (2) the competence of the user with regard to judging forest aesthetics. This study also contained an exploratory part that investigated whether our augmented predictions enabled the users to select additional training data to better customize the neural network to their personal aesthetic preferences.

This paper is structured into (1) related work (2) description of the data-set our work relies on, (3) the mobile prototype we developed, (4) the approach on explainable AI we chose as well as (5) the user study we conducted and (6) evaluation of the ANN and the user study.

2 RELATED WORK

People have used recreational green spaces for centuries, as can be seen in the cross-cultural practice of creating gardens and parks. Forests in particular can help to restore a lower blood-pressure after demanding cognitive work [17] and help us to relax through a multitude of fascinating features [18]. This is reflected in a recent trend called "Shinrin-yoku" or forest-bathing, where a person strives for relaxation by slowly wandering a forest, while perceiving the forest with all senses [24].

2.1 Scenic Beauty Assessment

Parsons et al. [25] argue, that decisions in scenic aesthetic environment classifications are holistic in an environmental psychological and cognitive scientific view and therefore sustainable. To support such decisions, assessment of aesthetic potential has been researched in scenic beauty estimation (SBE), e.g. by using validated scales to rate photographs [10]. One dilemma with aesthetic quality is according to [21], that it highly depends on the perspective of the observer. As a consequence, Bell et al. distinguish between 'expert-led' and 'perception-based' aesthetic criteria [7]. To better assess subjective criteria, visitor employed photography (VEP) [33] can be used. By analyzing images that were taken by the individual users, their personal preferences regarding features such as focus, viewing angle and situational taste can be assessed.

2.2 Aesthetic Landscapes in HCI

The perspective of using technology to support our goals of living a good life has led to a paradigm shift towards positive computing [8] where efficiency at work is no longer the sole purpose of computers. Consequentially, applications augmenting recreational forest walks are not a neglected topic in human computer interaction (HCI). The "Hobbit"-System by Posti et al. for instance, helps users to avoid other people in hiking tours by interpreting signals that are captured by mobile phones [27]. Other applications predict the aesthetics of scenery by using machine learning techniques such as artificial neuronal networks (ANNs). Such prediction models were trained on handcrafted features, for instance relating to a forest's structure [16] and directly on raw image data [26] to classify aesthetic value in scenery. Samsonov et al. for instance used ANN-based scenic beauty classifiers to select a passenger's seat on a bus-tour based on which side the view is more aesthetic [29]. Other projects explored tools that use these techniques to plan routes that are aesthetically more pleasing for travelers [22]. Most of the applications that consider scenic beauty for route planning rely on automatically generated images which are provided by the Google Streetview API and not by pictures that were taken by users. As a result, the user's individual aesthetic taste was not considered in these projects.

2.3 Interactive Machine Learning and Personalization

The term *Interactive Machine Learning* (IML) was introduced by Fails et al. who described a train-feedback-correct cycle which allows users to correct mistakes of a machine learning system [12]. Kelusza et al. took this idea a step further by focusing in depth on the transparency of machine learning classifiers to improve the user's mental model of a system and therefore the quality of his or her corrections. As part of their study, they generated detailed explanations for a naive bayes classifier within an e-mail categorization tool [19]. IML can also be used on mobile devices like a smartwatch [13]. In this case drinking activities were recognized.

2.4 Explainable Artificial Intelligence

With AI and Big Data becoming ubiquitous buzzwords in media over the recent years, the research field of artificial intelligence gained popularity. While such systems like convolutional neural networks (CNNs) achieved remarkable prediction accuracies, their decision process remains opaque in part because of the huge amount of learned parameters that are used. Recent years therefore saw a resurgence of research on Explainable Artificial Intelligence, which aims to make opaque black-box models more interpretable for humans. One way of increasing the interpretability of such a model is the creation of saliency maps, which are heat-maps that highlight the parts of the input that were relevant for a certain prediction [2]. While some saliency map algorithms use gradients to identify inputs that would change the prediction the most [32], other algorithms try to approximate the classifier with a better understandable model [28]. In this paper we use Layer-wise Relevance Propagation (LRP) which directly uses the trained weights to calculate how much each input contributed to the prediction [6]. This method was for example used by [15] to create an explainable recommendation system that not only gives a recommendation but also identifies the features that mostly influenced this recommendation. Our study, in contrast, uses transfer learning instead of manual corrections to retrain an image classifier to adapt predictions to the user's individual aesthetic taste. The ANNs used for this application run on a mobile device previously bench-marked by Seiderer et al. [31]. As model transparency encourages users to engage in the IML cycle more frequently [20], but is non-trivial to achieve for complex classifier architectures, such as ANNs, we used specialized explaining techniques from the research field of explainable artificial intelligence.

3 DATA-COLLECTION

The data-set is based on pictures taken by geography students in an urban forest. Photographs were taken using digital compact cameras and had to be associated with either the label *aesthetic* or *unaesthetic*. To simulate the process of single user rating, the photographs were reviewed by a single student, discarding those not inline with his judgment. This resulted in a data-set of 202 pictures, balanced within the two classes, 101 each for aesthetic / unaesthetic forest (see Figure 1 for examples). A data-set of 200 pictures can be considered an effort practicable by a single layperson.

The non-aesthetic part consists of all pictures with obvious human impact ruining the beauty of the scene and all pictures showing

Figure 1: Images labeled *unaesthetic* (left), *aesthetic* (right).

natural destructive processes. These include tree stumps, bark carvings, piled up logs, markers and signs, fencing around the trees, wagons left on an old railroad track, small to medium-sized boulders and parts with either only sparse vegetation or branches lying around. Samples of other human constructions such as a Kneipp basin that not only had no water in it but was also in bad condition were labeled as non-aesthetic as well as images showing the results of erosive processes by forces of nature. This compares well to references in literature [33]. Lastly, dead parts of the forest were also captured by the students, thus finalizing our unaesthetic part. The aesthetic part on the other hand mainly contains pictures of well grown branches, single or multiple trees, in some cases separated by a small footpath. Water related pictures showing a clear creek, stream or lake with decent forest plant growth around it are also part of the aesthetic class. Some pictures were taken on the edge of the forest, thus showing the transition from grassland to forest with a shiny blue sky above. These, as well as fountains, sculptures and trees or tree stumps covered in moss or ivy are categorized as aesthetic, too.

4 PROTOTYPE

Based on the mobile machine learning scenario we present our prototype implementation consisting of a transfer learning setup and an explainable AI setup.

4.1 On-device Model Training

To implement our prototype we decided to use Python in combination with a Web-based GUI. This makes it possible to use a full version of TensorFlow with classification and training functionality on a smartphone while providing a GUI that is familiar to Android users. An overview of the communication between the components of the prototype can be seen in Figure 2. The prototype is implemented by using multiple components. The main component provides a minimal webserver with websocket capability. It is used to host an HTML 5 based GUI using Google's "Material Design"[1] which is also used by Android and thus looks familiar to Android users. The main component calls two separate components as processes for retraining and classification. The retraining is con-

Figure 2: Communication between prototype components.

ducted by an adapted version of the TensorFlow retrain example [11]. This adapted version provides more outputs to be able to show the progress of the operations in more detail. Image classification is done by another component which loads the specified model and

[1]https://material.io/design/

labels before classifying a picture file. The GUI consists of two tabs. The "home" tab (see Figure 3a) allows the selection of a collection, can show information about the classes of a collection, allows the user to start the model retraining as well as to revert to the previous model and allows the classification of a picture. The "settings" tab (see Figure 3b) can be used to add and remove collections and classes. Additionally, already existing pictures can be added to a selected class. Information about the classes of a collection is shown at the bottom. Before being able to start image classification the user has to create a collection. In the best case there are already some pictures of the desired classes stored on the device so that the user can just add them by using the button on the "settings" tab. There have to be at least two classes with at least five pictures in each class to be able to create a classification model. Such information is shown in a dialog if the user tries to use the image recognition without having created a model or wants to train a model without sufficient pictures and classes.

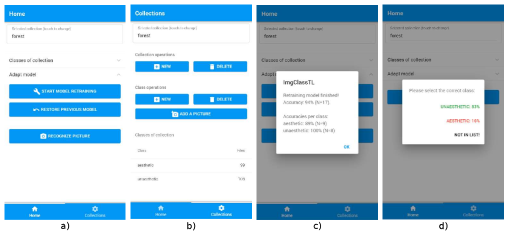

Figure 3: a) Main screen for model retraining, restoration and image recognition. The prototype can handle multiple collections (here: forest) with multiple classes (here: aesthetic / unaesthetic). b) Overview over collections and classes with possibilities to adapt them. c) Evaluation result after training a model. In this case 17 pictures were used and all 8 unseen unaesthetic pictures were classified correctly. d) If the recognition result was incorrect (user selected "No") this dialog allows to select the right class or add the image to a completely new class. The recognition probabilities are shown to make the process more transparent.

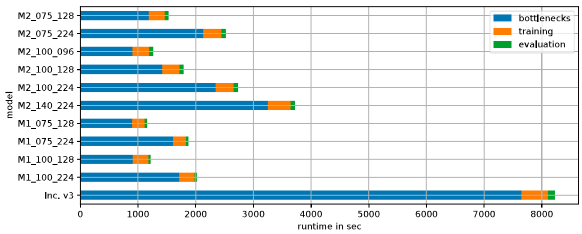

Figure 4: Runtimes of bottleneck (feature vector) creation, training and evaluation on a Nexus 6P smartphone. ANNs are MobileNet 1 (M1) and 2 (M2) at input from 75 by 128 pixels (M2_075_128) to to 100 by 224 pixels and Inception V3

If sufficient data is available, the model can be retrained. Information about the retraining process is shown in a dialog. In the end the performance of the model is shown to the user per class (see Figure 3c). This can give the user first hints about the practical performance of the model. After this process the recognition can be used by pressing on the "recognize picture" button. This opens first the default camera app of the system and after taking a picture the file is classified with the current model. The program will show as text how certain it is with its result based on the output of the classifier. It can select from "I'm sure the picture contains a ...", "I'm quite sure ..." and "I'm not sure but the picture seems to contain ...". Additionally, it asks whether this is correct. If the user selects "Yes" the new picture is added as a new sample to the recognized class, if "No" a list is opened where the user can select from all classes which shows the probabilities the classifier returned for each class and enables the user to create a new class (see Figure 3d). If the user restarts the training process the new samples are included in the data set. In case the user notices that a model is broken but the one before worked it is possible to revert to the last trained model. If the user added more data it is very likely that the problem gets resolved and the user can use a newly trained model on the data set for image recognition. We tested the transfer-learning process on a Nexus 6P smart-phone, see Figure 4. This shows that MobileNet with 128 pixel input has a good ratio when it comes to on-device training. Because of this and the fact that it achieves good results, it is the foundation of our evaluation regarding ANN-model in Section 6.1 as well as user-study in Section 6.2.

4.2 Explainable AI Approach

To identify the regions of an input picture that were especially relevant for our classification models we use an XAI method called deep Taylor decomposition which was initially introduced by Montavon et al. [23]. This method increases the transparency of a classifier by creating saliency maps that visualize the relevance of each pixel of the input for the decision of the network (see Figure 5). Deep

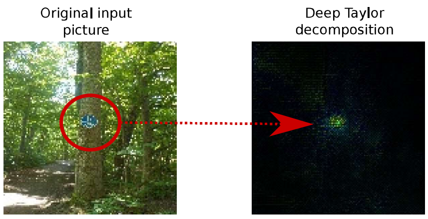

Figure 5: An example for deep Taylor decomposition. Highlighted on the left, the relevant parts of the image for the prediction. In this example, the decomposition clarifies that the network's prediction is mainly based on the street sign.

Taylor decomposition calculates the relevance of the input pixels by performing a layer-wise relevance propagation similar to the one introduced by Bach et al. [5] (in fact the latter approach can be embedded into the theoretical framework of deep Taylor decomposition). During this propagation the algorithm assigns a relevance value R_i to each neuron of the neural network. This process starts at the output layer, where the relevance of the prediction we want to analyze is defined as the activation of the respective neuron. The output neuron's relevance is then successively propagated backwards to each previous layer until it reaches the input layer where the neurons correspond to the input pixels. At each step of this propagation a Taylor approximation is used to approximate how relevant a neuron x_i^l in layer l was for a neuron x_j^{l+1} of the subsequent layer $l+1$. For this approximation we need to describe the relevance of x_j^{l+1} as a function $R_j^{l+1}(x^l)$ which depends on the neurons x_i^l of the previous layer. This can for example be done through previous relevance propagation steps. After such a function is found, one can decompose it using the Taylor series

$$R_j^{l+1}(x^l) = \sum_i \frac{\partial R_j^{l+1}}{\partial x_i^l}\Big|_{\widetilde{x}^l}(x_i^l - \widetilde{x}_i^l) + \varepsilon, \tag{1}$$

with Taylor residual ε and base point $\widetilde{x}^l$ which is chosen depending on x_j^{l+1}. The propagated relevance from x_j^{l+1} to x_i^l is given by $\frac{\partial R_j^{l+1}}{\partial x_i^l}\Big|_{\widetilde{x}^l}(x_i^l - \widetilde{x}_i^l)$ if one assumes that ε is small enough. Depending on the choice of base point $\widetilde{x}^l$ we obtain different deep Taylor methods. For our study we use the deep Taylor implementation of the INNvestigate framework [3].

5 USER STUDY

Figure 6: Examples of each block in the questionnaire.

To gather information regarding the reception of our approach, we conducted an online survey. Participants were presented six images in two modes (see Figure 6), one with deep Taylor decomposition saliency maps and one without, thus resulting in an overall iteration count of 12 images per participant. To prevent a bias related to learning effects, the two modes were permuted randomly and balanced. Both groups' questionnaires started out asking for personal demographic and tech-affinity information and were concluded by asking if each block's explanations were perceived as sufficient, how false-classifications would be handled in an interactive system as well as final feedback. Each of the 12 iterations contained questions regarding label agreement comprehension, regarding the system's classification, and improvement regarding sharpening visual sensing. Additionally, image regions perceived as important for classification were inquired. To explore the potential of our system to assist the user to customize the model to their personal aesthetic preferences we also included an open question. Here the participants had to describe which additional training data they would use to change the hypothetical prediction of a model for the input picture in Figure 5 from 'unaesthetic' to 'aesthetic'.

6 EVALUATION

Our prototype's evaluation is split in evaluation of the model (Section 6.1) that was trained using our data-set (Section 3), as well as evaluating deep Taylor decomposition as UI-feature (Section 6.2).

6.1 ANN-Model

The retrained model we use is based upon MobileNet V2 [30] with an input of 128 by 128 pixel. To gain a representative evaluation k-folds cross-validation with 5 splits is employed on overall 202 images distributed evenly onto two classes. Training took place in eight epochs and batches of 20 while having the first 55 layers frozen. The number of epochs was determined using a grid search. This results in an accuracy of 79.7% (confusion matrix in Table 1).

Table 1: Results of 5-Fold cross-validation (202 images).

	Confusion Matrix	
Predicted:	**Aesthetic**	**Unaesthetic**
Aesthetic	76	23
Unaesthetic	18	85
Accuracy		**79.7%**

6.2 User Study

We conducted a study with 31 participants of a mean age of 31 (SD = 10.54) with 18 males and 13 females. On a Likert scale from 1 to 5 when considering their machine-learning knowledge our participants scored an average of 2.48 (SD = 0.99). Participants were divided into two groups, one of them seeing the XAI images first, the other group seeing them secondly. The participants were assigned randomly. Since we employed a total of six t-tests, we used the Holm correction to adjust our p-values that are presented in the following. Independent t-tests were employed in the next three reported results, when only the answers raised on the participants' first encounter with the images were used, be it with (group A) or without (group B) deep Taylor visualizations.

First, we tested if the participant's agreement with the classification of the network differs between group A and group B, where group A has no higher agreement with the classification of the network compared to group B. There was no statistical difference ($t(166) = -0.06$, $p = 1.0$) between group A ($M = 3.42$, $SD = 1.37$) and group B ($M = 3.43$, $SD = 1.25$). **Second**, we evaluated the question if XAI compared to the UI without Deep-Taylor explanations helps to sharpen the participant's view. There was no statistically significant difference ($t(166) = 1.86$, $p = .09$) between group A ($M = 2.92$, $SD = 1.17$) and group B ($M = 2.58$, $SD = 1.14$). **Third**, we evaluated if XAI (A) helps users to comprehend the system's decision when compared to the UI without XAI (B). There was a statistically significant difference ($t(166) = 2.63$, $p = .02$) between group A ($M = 3.57$, $SD = 1.22$) and group B ($M = 3.08$, $SD = 1.22$), where group A found the system's decisions more comprehensive than group B.

Thereafter, paired t-tests were used to evaluate the difference between the UI with XAI and the UI without XAI for each participant also including the images rated in the participant's second block. Here the agreement with the system's classification can not, again as expected, be significantly discriminated ($t(185) = 2.60$, $p = 1.0$) between XAI (A) and non-XAI (B) with group A ($M = 3.44$, $SD = 1.36$) and group B ($M = 3.40$, $SD = 1.30$).

XAI does help to sharpen the user's view significantly ($t(185) = 2.60$, $p = .025$) when considering the UI-implementation with XAI (A) compared to the UI without XAI (B) with group A ($M = 2.70$, $SD = 1.17$) and group B ($M = 2.38$, $SD = 1.10$).

Compared to the independent t-test the effect regarding the helpfulness of XAI (A) to comprehend the systems decision in relation to the UI without XAI (B) increased and is significant ($t(185) = 6.0$, $p = .0003$) with group A ($M = 3.26$, $SD = 1.18$) and group B ($M = 2.47$, $SD = 1.23$).

To evaluate the open question about what kind of additional pictures the participants would choose to change the prediction of the model for the picture in Figure 5 from "unaesthetic" to "aesthetic", we took inspiration from Anderson et al. [4]. We asked three experts, that were involved in the training process of the model, how they would answer the question. All three of them answered with variations of "*I would include additional pictures that contain street signs and label them with unaesthetic*". Based on this answer we identified two concepts that should be part of a correct answer: (1) referring to the street sign and (2) choosing an appropriate labeling. Afterwards, we checked which participants included those priorly defined concepts in their answer.

Table 2: Results of whether the saliency maps enabled the participants to choose training data to adapt the network to their personal aesthetic preferences.

concepts included in answer	2	1	0
# participants	7	10	6
average ML experience	2.9	2.2	1.8

Out of our 31 participants 8 answered, they did not understand the question. The results of the remaining 23 are shown in Table 2.

7 DISCUSSION

We trained a ANN based on MobileNet v2 with a limited data-set containing 202 images (transfer learning) that could be created by a single user without effort. Our model was able to discriminate aesthetic from unaesthetic forest-scenes with an accuracy of 79.7% and therefore is a indicator, that we are able to create sufficient models with limited resources (hypothesis 0). The visualizations helped users to sharpen their view significantly, that is improving their competence when it comes to judging the aesthetics of forest views. The biggest effect could be found in regard of Deep-Taylor visualizations helping the user to comprehend the system's decision. To explore the benefit of additional saliency maps for the interactive machine learning process we asked the participants which additional training data they would use to change a certain prediction of the system. The results of this question showed that 74% of the participants, that understood the question, were able to identify at least one of the two concepts we identified as relevant for choosing training data to adapt the network to fit a specific aesthetic preference. After investigating, we saw a connection between the participant's understanding of machine learning algorithms and the amount of concepts they were able to recognize. The participants that identified both relevant concepts had the highest average machine learning experience while the group of participants that found none of the concepts had the lowest average machine learning experience. On the one hand we think these results indicate that augmenting the raw prediction with saliency maps improves the users' ability to personalize the system. On the other hand the connection to machine learning expertise shows that additional work is needed to increase the user's understanding of the training process of the network. This might be achieved by giving the users additional instructions on how to use the system but it also highlights the need for XAI approaches to convey knowledge about the underling machine learning system. Since visualizations highlighting regions of interest for the classifier within images were significantly rated to help the users to sharpen their view, the visualization goes beyond being a benefit just in the context of the interactive machine learning process. Deep Taylor visualizations hold the potential to be beneficial in the users' ability to judge the aesthetic quality of forest scenery.

8 CONCLUSION

We presented a prototype that enables smart-phone users to use transfer learning on their devices interactively. An accuracy of 79.7% was achieved by the prototype within the scenario of forest-aesthetic classification, based on a limited data-set consisting of images gained by Visitor Employed Photography. Finally, we presented insights from a user-study examining the use of deep-Taylor visualizations that improve the comprehensiveness of the machines decision process but also fosters the user's capabilities in judging a forest scene's aesthetics. Since forests aesthetics is relevant for users' wellbeing it could be further researched by including mobile physiological and environmental sensors like used in the study conducted in [14].

REFERENCES

[1] R.P. Abello and F.G. Bernaldez. 1986. Landscape preference and personality. *Landscape and Urban Planning* 13 (1986), 19 – 28.

[2] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[3] Maximilian Alber, Sebastian Lapuschkin, Philipp Seegerer, Miriam Hägele, Kristof T Schütt, Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller, Sven Dähne, and Pieter-Jan Kindermans. 2018. iNNvestigate neural networks! *arXiv preprint arXiv:1808.04260* (2018).

[4] Andrew Anderson, Jonathan Dodge, Amrita Sadarangani, Zoe Juozapaitis, Evan Newman, Jed Irvine, Souti Chattopadhyay, Alan Fern, and Margaret Burnett. 2019. Explaining Reinforcement Learning to Mere Mortals: An Empirical Study. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 1328–1334. https://doi.org/10.24963/ijcai.2019/184

[5] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. 2015. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE* 10, 7 (July 2015). https://doi.org/10.1371/journal.pone.0130140

[6] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one* 10, 7 (2015), e0130140.

[7] Simon Bell. 2001. Landscape pattern, perception and visualisation in the visual management of forests. *Landscape and Urban planning* 54, 1-4 (2001), 201–211.

[8] Rafael A Calvo and Dorian Peters. 2014. *Positive computing: technology for wellbeing and human potential.* MIT Press.

[9] Yuelu CHEN, Tianzhong ZHAO, Gang WU, and Feixiang CHEN. 2018. Optimal Walking Path Analysis Method for Forest Region. *Transactions of the Chinese Society for Agricultural Machinery* 6 (2018), 23.

[10] Terry C Daniel and Ron S Boster. 1976. Measuring landscape esthetics: the scenic beauty estimation method. *Res. Pap. RM-RP-167. US Department of Agriculture, Forest Service, Rocky Mountain Range and Experiment Station. 66 p.* 167 (1976).

[11] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2013. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. *CoRR* abs/1310.1531 (2013). arXiv:1310.1531

[12] Jerry Alan Fails and Dan R Olsen Jr. 2003. Interactive machine learning. In *Proceedings of the 8th international conference on Intelligent user interfaces*. ACM, 39–45.

[13] Simon Flutura, Andreas Seiderer, Ilhan Aslan, Chi-Tai Dang, Raphael Schwarz, Dominik Schiller, and Elisabeth André. 2018. DrinkWatch: A Mobile Wellbeing Application Based on Interactive and Cooperative Machine Learning. In *Proceedings of the 2018 International Conference on Digital Health (DH '18)*. Association for Computing Machinery, New York, NY, USA, 65–74.

[14] Simon Flutura, Andreas Seiderer, Ilhan Aslan, Michael Dietz, Dominik Schiller, Christoph Beck, Joachim Rathmann, and Elisabeth André. 2019. Mobile Sensing for Wellbeing Estimation of Urban Green Using Physiological Signals. In *Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good (GoodTechs '19)*. Association for Computing Machinery, New York, NY, USA, 249–254.

[15] Francesco Fusco, Michalis Vlachos, Vasileios Vasileiadis, Kathrin Wardatzky, and Johannes Schneider. 2019. RecoNet: An Interpretable Neural Architecture for Recommender Systems. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 2343–2349.

[16] Ali Jahani. 2019. Forest landscape aesthetic quality model (FLAQM): A comparative study on landscape modelling using regression analysis and artificial neural networks. *Journal of Forest Science* 65 (03 2019), 61–69.

[17] Rachel Kaplan. 2001. The nature of the view from home: Psychological benefits. *Environment and behavior* 33, 4 (2001), 507–542.

[18] Stephen Kaplan. 1995. The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology* 15, 3 (1995), 169 – 182. Green Psychology.

[19] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th international conference on intelligent user interfaces*. ACM, 126–137.

[20] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell me more?: the effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1–10.

[21] Andrew Lothian. 1999. Landscape and the philosophy of aesthetics: is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape and urban planning* 44, 4 (1999), 177–198.

[22] Satomi Manzaki, Ayame Kano, Narihiro Haneda, Chihiro Sato, and Naohito Okude. 2018. Country Road Finder: Exploring Beauty when Driving Around. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*. ACM, New York, NY, USA, 21–25.

[23] Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, and Klaus-Robert Müller. 2017. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition* 65 (2017), 211–222.

[24] Bum Jin Park, Yuko Tsunetsugu, Tamami Kasetani, Takahide Kagawa, and Yoshifumi Miyazaki. 2010. The physiological effects of Shinrin-yoku (taking in the forest atmosphere or forest bathing): evidence from field experiments in 24 forests across Japan. *Environmental health and preventive medicine* 15, 1 (2010), 18.

[25] Russ Parsons and Terry C Daniel. 2002. Good looking: in defense of scenic landscape aesthetics. *Landscape and Urban Planning* 60, 1 (2002), 43 – 56.

[26] Yuen Peng Loh, Song Tong, Xuefeng Liang, Takatsune Kumada, and Chee Seng Chan. 2017. Understanding scenery quality: A visual attention measure and its computational model. In *Proceedings of the IEEE International Conference on Computer Vision*. 289–297.

[27] Maaret Posti, Johannes Schöning, and Jonna Häkkilä. 2014. Unexpected Journeys with the HOBBIT - The Design and Evaluation of an Asocial Hiking App. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS*.

[28] 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*. ACM, 1135–1144.

[29] Pavel Samsonov, Florian Heller, and Johannes Schöning. 2017. Autobus: Selection of Passenger Seats Based on Viewing Experience for Touristic Tours. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 321–326.

[30] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[31] Andreas Seiderer, Michael Dietz, Ilhan Aslan, and Elisabeth André. 2018. Enabling Privacy with Transfer Learning for Image Classification DNNs on Mobile Devices. In *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good (Goodtechs '18)*. ACM, New York, NY, USA, 25–30.

[32] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2013. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *CoRR* abs/1312.6034 (2013). arXiv:1312.6034

[33] J. G. Taylor, K. J. Czarnowski, N. R. Sexton, and S. Flick. 1995. The importance of water to Rocky Mountain National Park visitors: an adaptation of visitor-employed photography to natural resources management. *Journal of Applied Recreation Research* 20, 1 (1995), 61–85.