1. Record Nr. UNINA9910872189603321

Autore	Longo Luca
Titolo	Explainable Artificial Intelligence : Second World Conference, XAI 2024, Valletta, Malta, July 17-19, 2024, Proceedings, Part IV
Pubbl/distr/stampa	Cham : , : Springer International Publishing AG, , 2024 ©2024
ISBN	3-031-63803-4
Edizione	[1st ed.]
Descrizione fisica	1 online resource (480 pages)
Collana	Communications in Computer and Information Science Series ; ; v.2156
Altri autori (Persone)	LapuschkinSebastian SeifertChristin
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- Contents - Part IV -- Explainable AI in Healthcare and Computational Neuroscience -- SRFAMap: A Method for Mapping Integrated Gradients of a CNN Trained with Statistical Radiomic Features to Medical Image Saliency Maps -- 1 Introduction -- 2 Related Work -- 3 Design and Methodology -- 3.1 The Approach -- 3.2 The Experiment -- 3.3 Evaluation of Saliency Maps -- 4 Results and Discussion -- 4.1 Discussion of Results -- 5 Conclusions and Future Work -- References -- Transparently Predicting Therapy Compliance of Young Adults Following Ischemic Stroke -- 1 Introduction -- 2 Related Studies -- 3 Materials and Methods -- 3.1 Participants and Clinical Settings -- 3.2 Cognitive Assessment -- 3.3 Computer-Based Rehabilitation Therapy -- 3.4 Modelling -- 3.5 Explanation Methods -- 4 Results -- 4.1 Experimental Data -- 4.2 Therapy Compliance Prediction -- 4.3 Explanation Reports -- 5 Discussion -- 6 Conclusions -- References -- Precision Medicine for Student Health: Insights from Tsetlin Machines into Chronic Pain and Psychological Distress -- 1 Introduction -- 2 Tsetlin Machines -- 3 Related Work -- 3.1 Pain and Psychological Distress -- 3.2 Explainable AI -- 4 Materials and Methods -- 4.1 The SHoT2018 Study -- 4.2 Models and Analyses -- 5 Results and Discussion -- 5.1 Performance -- 5.2 Interpretability Analysis -- 6 Conclusions and Future Work -- A Literal Frequency in the Tsetlin Machine -- References -- Evaluating Local Explainable AI

Techniques for the Classification of Chest X-Ray Images -- 1 Introduction -- 2 Previous Work -- 2.1 Explainable AI for X-Ray Imaging -- 2.2 Evaluation Metrics for XAI -- 3 Explainable AI Techniques -- 4 Analyzed Dataset -- 5 Proposed Metrics -- 6 Results and Evaluation -- 7 Conclusions -- References -- Feature Importance to Explain Multimodal Prediction Models. a Clinical Use Case. 1 Introduction -- 2 Related Work -- 2.1 Short-Term Complication Prediction -- 2.2 Multimodal Prediction Models -- 2.3 Explainability -- 3 Materials and Methods -- 3.1 Dataset -- 3.2 Machine Learning Models -- 3.3 Training Procedure and Evaluation -- 3.4 Explanation -- 4 Results -- 4.1 Model Performance -- 4.2 Explainability -- 5 Discussion -- 6 Conclusion -- A Hyperparameters -- B Detailed Feature Overview -- References -- Identifying EEG Biomarkers of Depression with Novel Explainable Deep Learning Architectures -- 1 Introduction -- 2 Methods -- 2.1 Description of Dataset -- 2.2 Description of Model Development -- 2.3 Description of Explainability Analyses Applied to All Models -- 2.4 Description of Approach for Characterization of M2 and M3 Filters -- 2.5 Description of Novel Activation Explainability Analyses for M2 and M3 -- 2.6 Key Aspects of Approach -- 3 Results and Discussion -- 3.1 M1-M3: Model Performance Analysis -- 3.2 M1-M3: Post Hoc Explainability Analysis -- 3.3 M2-M3: Characterization of Extracted Features -- 3.4 M2-M3: Post Hoc Spatial Activation Analysis -- 3.5 M2-M3: Post Hoc of Activation Correlation Analysis -- 3.6 Summary of MDD-Related Findings -- 3.7 Limitations and Future Work -- 4 Conclusion -- References -- Increasing Explainability in Time Series Classification by Functional Decomposition -- 1 Introduction -- 2 Background and Related Work -- 3 Method -- 4 Case Study -- 4.1 Sensor Model -- 4.2 Simulator -- 5 Application -- 5.1 Instantiation of the Generic Methodology -- 5.2 Influence of Data Representation and Decompositions -- 5.3 Influence of the Chunking -- 5.4 Datasets -- 6 Realization and Evaluation -- 6.1 Training and Testing of the Chunk Classifier -- 6.2 Training and Testing of the Velocity Estimator -- 6.3 Robustness Analysis of the Chunk Classifier -- 6.4 Testing of the Complete System -- 7 Explanations.
7.1 Dataset-Based Explanations -- 7.2 Visual Explanations -- 8 Conclusion and Future Work -- References -- Towards Evaluation of Explainable Artificial Intelligence in Streaming Data -- 1 Introduction -- 2 Related Work -- 2.1 Consistency, Fidelity and Stability of Explanations -- 3 Methodology -- 4 A Case Study with the Iris Dataset -- 5 Results Analysis -- 5.1 XAI Metric: Agreement (Consistency) Between Explainers -- 5.2 XAI Metric: Lipschitz and Average Stability -- 5.3 Comparison of Stability Metrics -- 5.4 Detailed Stability Comparison for Anomalies A1 and A2 -- 5.5 Quantification of Differences in Stability Between Ground Truth and Black-Box Explainers -- 6 Conclusion -- 7 Future Work -- References -- Quantitative Evaluation of xAI Methods for Multivariate Time Series - A Case Study for a CNN-Based MI Detection Model -- 1 Introduction -- 2 Background and State of the Art -- 2.1 Multivariate Time Series -- 2.2 MI Detection Use Case -- 2.3 Related Work -- 3 Methodology -- 3.1 Explanations for Time Series Data -- 3.2 Truthfulness Analysis -- 3.3 Stability Analysis -- 3.4 Consistency Analysis -- 4 Experimental Results -- 4.1 Results of the Truthfulness Analysis -- 4.2 Results of the Stability Analysis -- 4.3 Results of the Consistency Analysis -- 5 Discussion -- 6 Conclusion -- References -- Explainable AI for Improved Human-Computer Interaction and Software Engineering for Explainability -- Influenciæ: A Library for Tracing the Influence Back to the Data-Points -- 1 Introduction -- 2 Attributing Model Behavior

Through Data Influence -- 2.1 Notation -- 2.2 Influence Functions -- 2.3 Kernel-Based Influence -- 2.4 Tracing Influence Throughout the Training Process -- 3 API -- 4 Conclusion -- References -- Explainability Engineering Challenges: Connecting Explainability Levels to Run-Time Explainability -- 1 Introduction -- 2 Explainability Terminology.
3 MAB-EX Framework for Self-Explainable Systems -- 4 Explainability Requirements in Software-Intensive Systems -- 5 Integration of Explainability Levels into the MAB-EX Framework -- 6 The Role of eXplainable DRL in Explainability Engineering -- 7 Conclusion -- References -- On the Explainability of Financial Robo-Advice Systems -- 1 Introduction -- 2 Related Work -- 3 Background -- 3.1 Financial Robo-Advice Systems -- 3.2 XAI and the Law -- 3.3 EU Regulations Relevant to Financial Robo-Advice Systems: Scopes and Notions -- 4 Proposed Methodology -- 5 Robo-Advice Systems -- 6 Legal Compliance Questions for Robo-Advice Systems -- 7 Case Studies -- 7.1 Requested Financial Information -- 7.2 Personas -- 7.3 Results: Robo-Generated Financial Advice -- 8 Threats to Validity -- 9 Discussion -- 10 Conclusion and Future Work -- References -- Can I Trust My Anomaly Detection System? A Case Study Based on Explainable AI -- 1 Introduction -- 2 Literature Review -- 3 Preliminaries -- 3.1 VAE-GAN Models -- 3.2 Semi-supervised Anomaly Detection Using Variational Models -- 3.3 Explaining Anomaly Maps Using Model-Agnostic XAI Methods -- 3.4 Comparing Explained Anomalies Against a Ground Truth -- 4 Experimental Evaluation -- 5 Conclusions -- References -- Explanations Considered Harmful: The Impact of Misleading Explanations on Accuracy in Hybrid Human-AI Decision Making -- 1 Introduction -- 2 Background and Related Work -- 3 How Explanations Can Be Misleading -- 4 Methods -- 5 Results -- 5.1 Impact on Accuracy -- 5.2 Impact on Confidence -- 6 Discussion -- 7 Conclusion -- References -- Human Emotions in AI Explanations -- 1 Introduction -- 2 Related Literature -- 3 Method -- 4 Results -- 5 Robustness Check -- 6 Discussion -- 7 Conclusion -- References -- Study on the Helpfulness of Explainable Artificial Intelligence -- 1 Introduction -- 2 Measuring Explainability.
2.1 Approaches for Measuring Explainability -- 2.2 User Studies on the Performance of XAI -- 3 An Objective Methodology for Evaluating XAI -- 3.1 Objective Human-Centered XAI Evaluation -- 3.2 Image Classification and XAI Methods -- 3.3 Survey Design -- 4 Survey Results -- 4.1 Questionnaire Responses -- 4.2 Qualitative Feedback -- 5 Discussion -- 6 Conclusion -- Appendix A Additional Visualizations -- Appendix B Demographic Overview of Participants -- References -- Applications of Explainable Artificial Intelligence -- Pricing Risk: An XAI Analysis of Irish Car Insurance Premiums -- 1 Introduction -- 2 Background and Related Work -- 3 Data and Methods -- 4 Results -- 5 Discussion and Conclusion -- References -- Exploring the Role of Explainable AI in the Development and Qualification of Aircraft Quality Assurance Processes: A Case Study -- 1 Introduction -- 1.1 Background -- 1.2 Related Work -- 2 Description of Use Case -- 3 XAI Methods Applied to Use Case -- 4 Insights in the xAI Results -- 4.1 Experiment Results -- 4.2 xAI on New Model -- 5 Discussing xAI w.r.t. Development and Qualifiability -- 6 Conclusion -- References -- Explainable Artificial Intelligence Applied to Predictive Maintenance: Comparison of Post-Hoc Explainability Techniques -- 1 Introduction -- 2 Proposed Methodology -- 3 Post-Hoc Explainability Techniques -- 3.1 Impurity-Based Feature Importance -- 3.2 Permutation Feature Importance -- 3.3 Partial Dependence Plot (PDP) -- 3.4 Accumulated Local Effects (ALE) -- 3.5 Shapley Additive Explanations (SHAP) -- 3.6

Local Interpretable Model-Agnostic Explanations (LIME) -- 3.7 Anchor -- 3.8 Individual Conditional Expectation (ICE) -- 3.9 Discussion on Implementation and Usability of the Techniques -- 4 Conclusions -- References.

A Comparative Analysis of SHAP, LIME, ANCHORS, and DICE for Interpreting a Dense Neural Network in Credit Card Fraud Detection.