Trustworthy Explainability Acceptance: a New Metric to Measure the Trustworthiness of Interpretable AI Medical Diagnostic Systems

Davinder Kaur, Suleyman Uslu, Arjan Durresi, Sunil Badve, and Murat Dundar

Abstract We propose, *Trustworthy Explainability Acceptance* metric to evaluate explainable AI systems using expert-in-the-loop. Our metric calculates acceptance by quantifying the distance between the explanations generated by the AI system and the reasoning provided by the experts based on their expertise and experience. Our metric also evaluates the trust of the experts to include different groups of experts using our trust mechanism. Our metric can be easily adapted to any Interpretable AI system and be used in the standardization process of trustworthy AI systems. We illustrate the proposed metric using the high-stake medical AI application of Predicting Ductal Carcinoma in Situ (DCIS) Recurrence. Our metric successfully captures the explainability of AI systems in DCIS recurrence by experts.

Davinder Kaur
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: davikaur@iu.edu

Suleyman Uslu
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: suslu@iu.edu

Arjan Durresi
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: adurresi@iupui.edu

Sunil Badve
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: sbadve@iupui.edu

Murat Dundar
Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA e-mail: mdundar@iupui.edu

1 Introduction

The past decade has witnessed the massive deployment of algorithmic decision-making and artificial intelligence systems. These powerful, intelligent systems are used in many applications like business, healthcare, education, government, judicial, and many more. These systems have completely transformed our lives. With vast data and computation power availability, these systems have become very efficient but equally complex. Despite the many advantages, these systems have become opaque and can cause harm to the users and society. These systems have become black boxes, which are challenging to interpret, driving unfair and wrong decisions. Well-known examples of the harm caused by them are: self-driving car killed a pedestrian [37], recidivism algorithm used in our judicial system found to be biased against black people [3], recruitment algorithm used by a significant tech company found to be biased against women [9]. These examples show how important it is to make these systems safe, reliable, and trustworthy.

To prevent the harm caused by them, various government and scientific agencies have proposed guidelines and frameworks to make these systems trustworthy. European Union (EU) has presented ethical guidelines and frameworks to govern the development and working of AI systems [12] and also passed a law called GDPR(General Data Protection Regulation) [6] which gives users the "right of explainability" about the decisions made by AI systems. Standardization organization ISO also presented different approaches to establishing trust in AI systems [18]. Various solutions have been proposed to implement different requirements of trustworthy AI like fairness, accountability, explainability, privacy, and controllability[18]. All these requirements play an important role in making AI systems safe, reliable, and trustworthy. [20] [22] presented survey of all these requirements and their proposed methods. However, nowadays, great attention is given to make AI systems explainable and interpretable. Many methods have been proposed to make sure that people using the system and impacted by it should have a clear understanding of the system, its uses, and its limitations [16]. [1] [11] [17] presented an in-depth surveys reviewing different explainability methods. Despite the availability of many explainability approaches, there is a lack of evaluation metrics to quantitatively compare and judge them. Different researchers have presented different ways to evaluate them without any concrete metrics.

As the use of AI systems in high stake applications has increased, there is a growing need for standardization to govern the development and implementation of these systems [18]. The standardization process requires metrics [23]that will quantitatively compare and judge different explainability approaches and create a common language for the developers and the users of the system. Authors in [33] proposed a trustworthy AI metric for acceptance requirement. There is a need for such metrics for the explainability requirement of trustworthy AI. For this reason, in this paper, we have proposed a concrete, Trustworthy Explainability Acceptance metric and its measurement methodology. Our metric uses human-in-the-loop and is capable of quantifying the interpretable AI system's explanations by the experts. We have illustrated our metric using a high stake medical application involving predicting the recurrence of Ductal Carcinoma In Situ (DCIS). Our contributions can be summarized as follows:

- We propose in Section 3 our *Trustworthy Explainability Acceptance* metric for evaluating the Explainability of AI-based systems by field experts.
- The measurement procedure for the proposed metric is described in Section 3 and is based on the concept of a distance acceptance approach that is adaptable to a wide range of systems. In addition to the acceptance value, our metric provides the confidence of the acceptance.
- Our metric utilizes the trust of the experts in the given context, managed by our trust system, summarized in Section 2.
- Our metric can be measured in many points of the system to reach an assessment of the whole system, as discussed in Section 4.
- Finally, in Section 4, we illustrate the application of our trustworthy acceptance metric and its measurement methodology using an interpretable AI system for DCIS Recurrence Prediction.

2 Background and Related Work

This section gives information about the need for the metrics to measure the trustworthy explainability acceptance and the trust mechanism on which the metric is based on.

2.1 Need for AI Explainability Metrics

Much research is done in designing the guidelines and the framework to make AI systems trustworthy. However, significantly less work is dedicated to creating measurement mechanisms to measure the trustworthiness of the AI system. This measurement mechanism is needed to quantify the system's trustworthiness, which will lead to more acceptance of the systems. Standardization organizations such as ISO [18] also raised the need for metrics' to measure the trustworthiness of the AI systems. Their standardization document has mentioned various challenges related to the implementation and use of AI systems. The primary concern they have raised is the over-reliance and under-reliance on the AI system. Over-reliance can happen if the user becomes too reliant on the automation without considering its limitations. It can lead to adverse outcomes. And under-reliance can occur if the user keeps on overriding/disagreeing with the correct AI system decisions. To avoid these issues, a quantitative measurement analysis is needed to effectively compute the trustworthiness of the system based on its past predictions and its explanation for that.

Different trustworthy requirements need different methods of evaluation. In this paper, we have focused on the explainability requirement of trustworthy AI systems. Over the past years, many explainability methods have been proposed to make AI systems transparent and understandable. However, there is still an implementation gap from research to practice. The main reason for such an implementation gap is the lack of methods to compare and evaluate these systems[4]. Very little research has been done in designing these evaluation methods. Different researchers have presented various measures, tools, and principles to develop these evaluation systems. Some researchers have presented measures that are important for explainability evaluation [18], some have presented fact sheet to evaluate explainability methods based on their functionality, usability, and safety [29], some have presented fidelity method of comparing the accuracy of an interpretable model with a black box for evaluation [14], some have proposed an evaluation approach based on comparing local explanations with ground truth [15]. Some researchers also suggested quantitative evaluation methods like faithfulness metric [2] and monotonicity [25] which evaluate the system by measuring how the change in feature importance weights affect the classifier probability. All these proposed methods do not capture the human aspect of explainability. It is essential to have human involvement to increase the confidence in AI systems [8]. There is a need for more quantitative evaluation metrics that can compare different explainability methods and quantify the human acceptance of those methods to increase the use and trust in them. Furthermore, such metrics can be used for the standardization of explainable AI solutions and later for their certification by the appropriate agencies.

2.2 Trust Mechanism

The proposed metric is based on our trust framework [28], which is composed of two parameters: impression and confidence. The impression is defined as the level of trust one entity has towards another entity. It is the comprehensive summary of all the measurements between two entities (P and Q) taken over time, as shown in Eq. 1. The more the impression, the more will be the trustworthiness of the system. $m^{P:Q}$ is the impression, and $r_i^{P:Q}$ is the i-th measurement from P to Q, where N is the total number of measurements.

$$m^{P:Q} = \frac{\sum_{i=1}^{N} r_i^{P:Q}}{N} \tag{1}$$

Confidence measures the certainty of the impression and is defined as how sure one entity is about its impression of another entity. It is inversely proportional to the standard error of the mean. It is calculated using the formula given in Eq. 2. $c^{P:Q}$ is the confidence that P has about his impression of Q.

$$c^{P:Q} = 1 - 2\sqrt{\frac{\sum_{i=1}^{N}(m^{P:Q} - r_i^{P:Q})^2}{N(N-1)}} \tag{2}$$

Trust is a tuple of impression and confidence. Trust can also be calculated if the two entities are not directly related to each other using trust propagation methods, namely transitivity and aggregation. Trust transitivity is needed when two entities are not communicating directly but through a third entity between them. Trust aggregation is used to calculate summarized trust when two or more different links are present between entities. Authors in [28] proposed various methods for calculating transitivity and aggregation. In this application, we have used averaging aggregation method which is presented in Eq. 3, and its error formula is shown in Eq. 4.

$$m_{T_1}^{PQ} \oplus m_{T_2}^{PQ} = \frac{m_{T_1}^{PQ} + m_{T_2}^{PQ}}{2} \tag{3}$$

$$e_{T_1}^{PQ} \oplus e_{T_2}^{PQ} = \sqrt{\frac{1}{2^2}((e_{T_1}^{PQ})^2 + (e_{T_2}^{PQ})^2)} \tag{4}$$

Our trust framework is validated and applied in various applications such as stock market prediction using twitter[27], fake users detection [19], crime prediction [21], and decision making systems in food-energy-water sectors[30, 31, 32, 34, 35, 36] .

3 AI Trustworthy Explainability Acceptance Metric

This section introduces our *Trustworthy Explainability Acceptance* metric and its measurement methodology. We assume an explainable AI system that provides reasoning for its decisions, and there is a group of experts that will evaluate the system based on their judgment. Each expert can agree or disagree with the explanation provided by the system.

The explainability acceptance is based on the closeness of the explanations. More the distance between the explanations, the lesser will be the acceptance. The explainability distance between the two explanations is calculated using the Euclidean distance formula. Each explanation is considered as a vector for distance calculation where different attributes of the explanation become a dimension. The explainability distance can be anything between 0 and 1, 1 being the maximum distance. For example, the explainability distance between two n-dimensional explanations X and Y represented as d_X^Y, is shown in Eq. 5 where X_i and Y_i are the values if i is the dimension for each explanation.

$$d_X^Y = \sqrt{\frac{\sum_{i=1}^{n} (X_i - Y_i)^2}{n}} \tag{5}$$

Explainability acceptance is calculated using the formula given in Eq.6. Explainability acceptance by the expert *e* for the system will be based on his/her explanation X and the explanation Y provided by the system. Explainability acceptance also lies between [0,1], 1 being the highest acceptance. The distance is bidirectional. That is, if the explanation provided by the system is not close enough to the explanation of the expert, the expert will have less explainability acceptance for the system.

$$A_e = 1 - d_X^Y \tag{6}$$

A certain number of experts will evaluate and rate the system with their explainability acceptances based on their reasoning to reduce the subjectivity. Each acceptance is considered a trust assessment, and we aggregate them using Eq. 3. We calculate the *Trustworthy Explainability Acceptance* metric by aggregating different experts' explainability acceptances weighted by their trust values, as shown in Eq. 7, where T_e is the trust of expert e and E is the total number of experts. The trust value is calculated using the impression and confidence described in Section 2.2.

$$Tw_A = \frac{\sum_e A_e T_e}{E} \tag{7}$$

The confidence of the measured *Trustworthy Explainability Acceptance* is calculated based on Eq. 2 and 4, as shown in Eq. 8 and Eq.9.

$$SE_{Tw_A} = \frac{\sqrt{\sum_e (Tw_A - A_e)^2}}{n} \tag{8}$$

$$c_{Tw_A} = 1 - 2(SE_{Tw_A}) \tag{9}$$

Therefore, our metric to measure the trustworthy explainability is the tuple (Tw_A, c_{Tw_A}). When the system needs to be evaluated based on different sample measurements we can use the aggregation method of our trust framework. The aggregated explainability acceptance means and the standard error for n samples are calculated based on the aggregation trust propagation method as shown in Eq. 10 and Eq. 11.

$$System_{Tw_A} = \frac{\sum_n Tw_A}{n} \tag{10}$$

$$SE_{System_{Tw_A}} = \sqrt{\frac{1}{n^2} \sum_n (SE_{Tw_A})^2} \tag{11}$$

4 Evaluating AI system for DCIS Recurrence Prediction

To illustrate the proposed *Trustworthy Explainability Acceptance* metric, we have evaluated an AI system for DCIS recurrence prediction. This section provides an overview of the DCIS Recurrence prediction problem, data, experimental setup, and results after assessing the prediction systems using the proposed metric.

4.1 Background

Ductal carcinoma in situ (DCIS) is a non-obligate precursor lesion that is managed aggressively. Different DCIS trials have documented that the addition of radiation [7][10] and endocrine therapy [5] will result in reductions of recurrence rates. DCIS if left untreated, has a chance to progress to invasive carcinoma in only 20-40 % cases [13][24][26]. This has led to significant concerns regarding the over-treatment of patients. There is a need for an objective tool that helps identify women who are unlikely to recur and perform evidence-based de-escalation of additional therapies to avoid aggressive treatments. The development of such a tool requires a deep interpretable machine learning system for computer-aided recurrence prediction for DCIS, and the implementation and use of such a system require doctors' acceptance and trust.

4.2 Data

Original diagnostic slides from the patients diagnosed with DCIS in 2009 - 2012 within Indiana University Hospital System and at the Eskenazi City hospital were reviewed, and clinical data were obtained. Any case that was upgraded to invasive carcinoma was excluded. This review identified around 20 cases each of recurrent and non-recurrent DCIS with at least 8-year follow-up data. After excluding the missing cases (including referral/ second opinion cases) and the cases with a scant amount of DCIS, 30 cases (15 recurring and 15 non-recurring) were available for our studies.

The machine learning system uses these patient cases with eight years of follow-up data to determine DCIS recurrence. For simplicity, the image analysis part was performed manually by an internationally recognized breast pathologist. A thorough review of the histological slides was carried out, and the areas of DCIS were identified. The pathologist has categorized each slide based on 25 attributes. The values of the 25 attributes for all 30 cases served as the input for the machine learning model. A supervised machine learning model support vector machine (SVM) is used for classification. Using this model, we were able to get 83% accuracy on a leave-one-out cross-validation basis. The model has found 13 features useful for predicting DCIS recurrences and their corresponding optimized weights. Table 1. provides the list of these morphological features and their corresponding descriptors. As our approach is not to quantify a given system but to develop a proof of the concept of the metric, we have simulated another algorithm profile and four pathologist profiles by changing the weight of one of the morphological feature descriptors to keep it simple. This is done to simulate how different pathologists can have different opinions based on their experience level related to DCIS prediction. One AI system may find some features more important than others for prediction.

Table 1 Relevant Histo-morphological Criteria for Predicting DCIS Recurrence

Morphological Features	Descriptors
Architecture Solid	yes, no
Architecture Other	cribriform, micropaa, papillary, other
Lumina	regular, irregular
Nuclear Shape	round, oval, irregular
Nuclear Size	small, intermediate, large
Nuclear Pleomorphism	mild, moderate, prominent
Nuclear Membrane	smooth, irregular
Nuclear Spacing	even, uneven
Nucleoli	present, absent
Nucleolar Shape	round, oval, pleom, n/a
Mitosis	abnormal, normal
Necrosis	absent, focal, comedo
Immune Cells with Circumferential Distribution	yes,no

4.3 Experiments and Results

In our study, we assumed that the AI system is better in predicting the recurrence of DCIS than pathologists, and there is a need to evaluate the system to create trust and acceptance of the system among pathologists before deploying it in the hospital setting. An appropriate organization responsible for testing and certifying AI systems has employed high trust expert pathologists to evaluate the system using their expertise. We used our explainability acceptance metric to measure the system's acceptance by comparing the explanations provided by the algorithm to the cognitive reasoning of expert pathologists.

To demonstrate our proposed metric, we assumed two interpretable AI systems (System 1 and System 2) that need to be evaluated. Four simulated expert pathologists with high trust are deployed to do the evaluation. The interpretable AI systems generate morphological feature descriptors and corresponding weights along with their predictions. The pathologists evaluate the output provided by the system based on their expertise and provide their weights for the morphological feature descriptors.

An evaluation of a system starts by comparing the system explanation/weights of the feature descriptors with the weights of the pathologists, which is given based on their expertise and experience. For each pathologist, a non-zero distance is measured between the pathologist's explanation and the explanation provided by the system using Eq. 5. After having the individual distances, the acceptance rate for each pathologist is calculated by Eq. 6. Then we averaged all the individual acceptances weighted by their trust and calculated the confidence of the acceptance using Eqs. 7, 8, and 9. The trust-weighted average acceptance and its confidence constitutes our Trustworthy Explainability Acceptance metric. We have performed the same tasks for System 2.

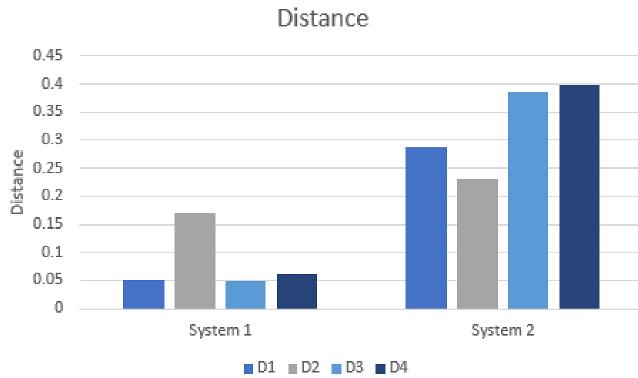

Fig. 1 Explainability distance between explanations provided by the system and the reasoning provided by pathologists based on their expertise.

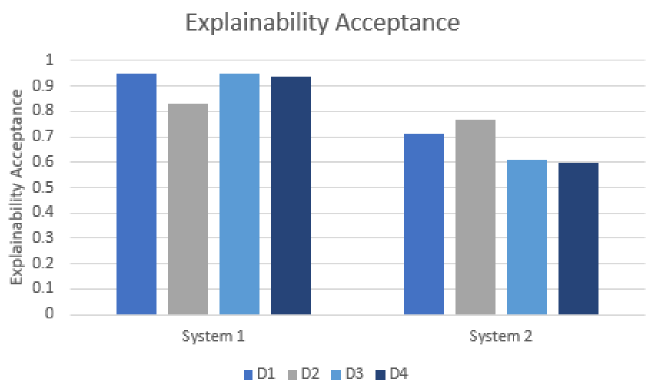

Fig. 2 Measured explainability acceptances of each pathologist D for System 1 and System 2.

Fig. 1 shows the explainability distance of all the pathologists for System 1 and System 2., calculated using Eq. 5. Fig. 2 shows the explainability acceptance of the systems calculated based on the difference in the opinions, calculated using Eq. 6. Fig.3 shows *Trustworthy Explainability Acceptance* metric, Tw_A, calculated using Eq.7, and the corresponding confidence values are calculated using Eq. 9. The variation in the acceptance shows how one system with similar accuracy as the other one can be more accepted based on the feature importance and explanation provided by it. The certifying agencies could use this type of evaluation metrics to standardize and certify AI systems based on the evaluation provided by top experts regarding the AI systems' explainability. For example, in our experiment, System 1 is more acceptable than System 2 regarding explainability. Therefore, our metric provides a framework to measure the acceptance of AI systems by experts based on system explainability.

Fig. 3 *Trustworthy Explainability Acceptance* of System 1 (confidence: 0.99) and System 2 (confidence: 0.98)

5 Conclusions

We presented *Trustworthy Explainability Acceptance* metric for evaluating explainable AI systems using expert-in-the-loop. This evaluation method provides a quantitative way to compare and judge different interpretable AI systems and provides a common language to all the various stakeholders involved in the designing, development, testing, standardization, and implementation phase of AI systems. Our metric measures the distances between the explanation provided by the system and the reasoning provided by the experts. Based on these distances and the trust of the experts, we calculated the *Trustworthy Explainability Acceptance* and its confidence. Our metric trust mechanism will help differentiate between different experts based on their expertise and reputation. Our *Trustworthy Explainability Acceptance* metric can be applied to any interpretable AI system that can use the concept of distance measurements.

Acknowledgements This work was partially supported by the National Science Foundation under Grant No. 1547411, by the U.S. Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) (Award Number 2017-67003-26057) via an interagency partnership between USDA-NIFA and the National Science Foundation (NSF) on the research program Innovations at the Nexus of Food, Energy and Water Systems, and by an IUPUI iM2CS seed grant.

References

1. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (xai). IEEE access **6**, 52,138–52,160 (2018)
2. Alvarez-Melis, D., Jaakkola, T.S.: Towards robust interpretability with self-explaining neural networks. arXiv preprint arXiv:1806.07538 (2018)

3. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: Machine bias. propublica. See https://www.propublica. org/article/machine-bias-risk-assessments-in-criminal-sentencing (2016)
4. Arya, V., Bellamy, R.K., Chen, P.Y., Dhurandhar, A., Hind, M., Hoffman, S.C., Houde, S., Liao, Q.V., Luss, R., Mojsilović, A., et al.: One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. arXiv preprint arXiv:1909.03012 (2019)
5. Bartlett, J.M., Thomas, J., Ross, D.T., Seitz, R.S., Ring, B.Z., Beck, R.A., Pedersen, H.C., Munro, A., Kunkler, I.H., Campbell, F.M., et al.: Mammostrat® as a tool to stratify breast cancer patients at risk of recurrence during endocrine therapy. Breast cancer research **12**(4), 1–11 (2010)
6. Calder, A.: EU GDPR: a pocket guide. IT Governance Ltd (2018)
7. Correa, C., McGale, P., Taylor, C., Wang, Y., Clarke, M., Davies, C., Peto, R., Bijker, N., Solin, L., Darby, S., et al.: Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. Journal of the National Cancer Institute. Monographs **2010**(41), 162–177 (2010)
8. Danny Tobey, M.: Explainability: where ai and liability meet: Actualités: Dla piper global law firm (2019). URL https://www.dlapiper.com/fr/france/insights/publications/2019/02/explainability-where-ai-and-liability-meet/
9. Dastin, J.: Amazon scraps secret ai recruiting tool that showed bias against women. San Fransico, CA: Reuters. Retrieved on October **9**, 2018 (2018)
10. DeSantis, C.E., Fedewa, S.A., Goding Sauer, A., Kramer, J.L., Smith, R.A., Jemal, A.: Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. CA: a cancer journal for clinicians **66**(1), 31–42 (2016)
11. Došilović, F.K., Brčić, M., Hlupić, N.: Explainable artificial intelligence: A survey. In: 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 0210–0215. IEEE (2018)
12. EC: Ethics guidelines for trustworthy ai (2018). URL https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
13. Esserman, L.J., Thompson, I.M., Reid, B., Nelson, P., Ransohoff, D.F., Welch, H.G., Hwang, S., Berry, D.A., Kinzler, K.W., Black, W.C., et al.: Addressing overdiagnosis and overtreatment in cancer: a prescription for change. The lancet oncology **15**(6), e234–e242 (2014)
14. Freitas, A.A.: Comprehensible classification models: a position paper. ACM SIGKDD explorations newsletter **15**(1), 1–10 (2014)
15. Guidotti, R.: Evaluating local explanation methods on ground truth. Artificial Intelligence **291**, 103,428 (2021)
16. Gunning, D.: Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web **2**(2) (2017)
17. Gunning, D., Aha, D.: Darpa's explainable artificial intelligence (xai) program. AI Magazine **40**(2), 44–58 (2019)
18. Information Technology– Artificial Intelligence – Overview of trustworthiness in artificial intelligence . Standard, International Organization for Standardization (2020)
19. Kaur, D., Uslu, S., Durresi, A.: Trust-based security mechanism for detecting clusters of fake users in social networks. In: Workshops of the International Conference on Advanced Information Networking and Applications, pp. 641–650. Springer (2019)
20. Kaur, D., Uslu, S., Durresi, A.: Requirements for trustworthy artificial intelligence–a review. In: International Conference on Network-Based Information Systems, pp. 105–115. Springer (2020)
21. Kaur, D., Uslu, S., Durresi, A., Mohler, G., Carter, J.G.: Trust-based human-machine collaboration mechanism for predicting crimes. In: International Conference on Advanced Information Networking and Applications, pp. 603–616. Springer (2020)
22. Kumar, A., Braud, T., Tarkoma, S., Hui, P.: Trustworthy ai in the age of pervasive computing and big data. In: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 1–6. IEEE (2020)

23. Lakkaraju, H., Adebayo, J., Singh, S.: Tutorial: (track2) explaining machine learning predictions: State-of-the-art, challenges, and opportunities. In: NeurIPS 2020. NeurIPS Foundation (2020)
24. Lester, S.C., Connolly, J.L., Amin, M.B.: College of american pathologists protocol for the reporting of ductal carcinoma in situ. Archives of pathology & laboratory medicine **133**(1), 13–14 (2009)
25. Luss, R., Chen, P.Y., Dhurandhar, A., Sattigeri, P., Zhang, Y., Shanmugam, K., Tu, C.C.: Generating contrastive explanations with monotonic attribute functions. arXiv preprint arXiv:1905.12698 (2019)
26. Moran, M.S., Schnitt, S.J., Giuliano, A.E., Harris, J.R., Khan, S.A., Horton, J., Klimberg, S., Chavez-MacGregor, M., Freedman, G., Houssami, N., et al.: Society of surgical oncology–american society for radiation oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages i and ii invasive breast cancer. International Journal of Radiation Oncology* Biology* Physics **88**(3), 553–564 (2014)
27. Ruan, Y., Durresi, A., Alfantoukh, L.: Using twitter trust network for stock market analysis. Knowledge-Based Systems **145**, 207–218 (2018)
28. Ruan, Y., Zhang, P., Alfantoukh, L., Durresi, A.: Measurement theory-based trust management framework for online social communities. ACM Transactions on Internet Technology (TOIT) **17**(2), 1–24 (2017)
29. Sojda, R.S.: Empirical evaluation of decision support systems: Needs, definitions, potential methods, and an example pertaining to waterfowl management. Environmental Modelling & Software **22**(2), 269–277 (2007)
30. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M.: Decision support system using trust planning among food-energy-water actors. In: International Conference on Advanced Information Networking and Applications, pp. 1169–1180. Springer (2019)
31. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M.: Trust-based game-theoretical decision making for food-energy-water management. In: International Conference on Broadband and Wireless Computing, Communication and Applications, pp. 125–136. Springer (2019)
32. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M.: Trust-based decision making for food-energy-water actors. In: International Conference on Advanced Information Networking and Applications, pp. 591–602. Springer (2020)
33. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M.: Trustworthy acceptance: A new metric for trustworthy artificial intelligence used in decision making in food-water-energy sectors. In: International Conference on Advanced Information Networking and Applications. Springer (2021)
34. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M., Tilt, J.H.: Control theoretical modeling of trust-based decision making in food-energy-water management. In: Conference on Complex, Intelligent, and Software Intensive Systems, pp. 97–107. Springer (2020)
35. Uslu, S., Kaur, D., Rivera, S.J., Durresi, A., Babbar-Sebens, M., Tilt, J.H.: A trustworthy human–machine framework for collective decision making in food–energy–water management: The role of trust sensitivity. Knowledge-Based Systems **213**, 106,683 (2021)
36. Uslu, S., Ruan, Y., Durresi, A.: Trust-based decision support system for planning among food-energy-water actors. In: Conference on Complex, Intelligent, and Software Intensive Systems, pp. 440–451. Springer (2018)
37. Wakabayashi, D.: Self-driving uber car kills pedestrian in arizona, where robots roam. The New York Times **19** (2018)