User Feedback to Improve the Performance of a Cyberattack Detection Artificial Intelligence System in the e-Health Domain

Carmelo Ardito, Tommaso Di Di Noia, Eugenio Di Sciascio, Domenico Lofù, Andrea Pazienza, Felice Vitulano

User Feedback to Improve the Performance of a Cyberattack Detection Artificial Intelligence System in the e-Health Domain

Carmelo Ardito[1][0000−0002−6827−2585], Tommaso Di Noia[1][0000−0002−0939−5462], Eugenio Di Sciascio[1][0000−0002−5484−9945], Domenico Lofù[1,2][0000−0001−6413−9886], Andrea Pazienza[2][0000−0002−6827−2585], and Felice Vitulano[2][0000−0002−6059−8177]

[1] Politecnico di Bari – Via E. Orabona 4, Bari (I-70125), Italy
{carmelo.ardito, tommaso.dinoia, eugenio.disciascio, domenico.lofu}@poliba.it
[2] Innovation Lab, Exprivia S.p.A. – Via A. Olivetti 11, Molfetta (I-70056), Italy
{domenico.lofu, andrea.pazienza, felice.vitulano}@exprivia.com

Abstract. New and evolving threats emerge every day in the e-Health industry. The safety of e-Health's telemonitoring systems is becoming a prominent task. In this work, starting from a CADS (Cyberattack Detection System) model that uses artificial intelligence techniques to detect anomalies, we focus on the activity of interacting with data. Using a User Interaction Engine, a dashboard allows you to visually explore and view data from suspected attacks on healthcare professionals for a threat reaction. In particular, a User Feedback module is presented to interact with healthcare personnel and ask for a response on the anomaly detected.

Keywords: User Interaction · User Feedback · Artificial Intelligence for Security · Internet of Medical Things · e-Health.

1 Introduction

In the e-Health sector, the protection of patient telemonitoring systems is essential to ensure that they follow their clinical path without any kind of external intrusion. In particular, Artificial Intelligence (AI) and Machine Learning (ML) have become critical technologies in information security, as they are capable of rapidly analyzing millions of events and identifying many different types of threats.

Intrusion analysts infer the context of a security breach by using prior knowledge to discover events relevant to the incident and understand why it happened [1]. Although security tools have been developed that provide visualization techniques and minimize human interaction to simplify the analysis process, too little attention has been paid to humanizing the interpretation of security incidents. Simply reporting a cyberattack is not sufficient to allow the healthcare

professional to correct the patient's clinical path. These data must be represented graphically, which can be understood by the healthcare professional. The detection of the cyberattack must therefore be supported by systems that provide different forms of visualization and interaction, according to the different end-users, and that allows them to have the possibility to interactively manipulate graphical representations based on Visual Data Mining (VDM) techniques.

This paper is organized as follows. Section 2 provides an overview of related work and technologies which were investigated as background knowledge. Section 3 provides the main contribution of the paper regarding the user interaction with a Cyberattack Detection System (CADS). Finally, Section 4 concludes the paper, outlining future works.

2 Background and Related Work

Cyberattack Detection System (CADS) is software that automates the cyberattack detection process and detects possible cyberattacks. Anomaly detection typically works on monitored network traffic data. Indeed, the integration of healthcare-based devices and sensors within the Internet of Things (IoT) has led to the evolution of the Internet of Medical Things (IoMT) [12]. In particular, ML-based anomaly detection systems are critical for ensuring security and mitigating threats such as bogus data injection attacks [7]. Therefore, designing a distributed security framework for distributed IoMT applications is a challenging task due to the dynamic nature of IoMT networks such as IoT devices, edge devices, and the cloud. The line of research in this way is moving towards building robust anomaly-based intrusion detection systems that efficiently distinguish attack and normal observations in the IoMT environment, consisting of interconnected devices and sensors. For this reason, works in [3, 8, 9] has dealt with a clinical and operational context to develop integrated solutions for continuous care in which AI and IoMT are used at the Edge, with a people-centered approach that fit the needs of healthcare professionals and is incorporated into their workflows.

Figure 1 depicts the architecture of the CADS proposed in [2]. It focuses on the security of data transmitted from IoMT sensors to three different interconnected processing modules: a Clinical Pathway Anomaly Detection (CPAD), an Explainer module, and an User Interaction Engine. The latter is made up of three sub-modules, i.e.: a Visualization Framework, an User Interface, and an User Feedback system. The activity of interacting with the data is carried out by means of the User Interaction Engine, which provides a dashboard through which to visually explore the data to get a clearer view of what happened over a period of time. In particular, thanks to the *User Feedback* module, it is possible to implement a continuous improvement of the classification performance, and consequently of the anomaly detection, thus obtaining a more effective identification of threats. In this way, CADS is increasingly accurate in identifying threats and therefore more robust from a security point of view.

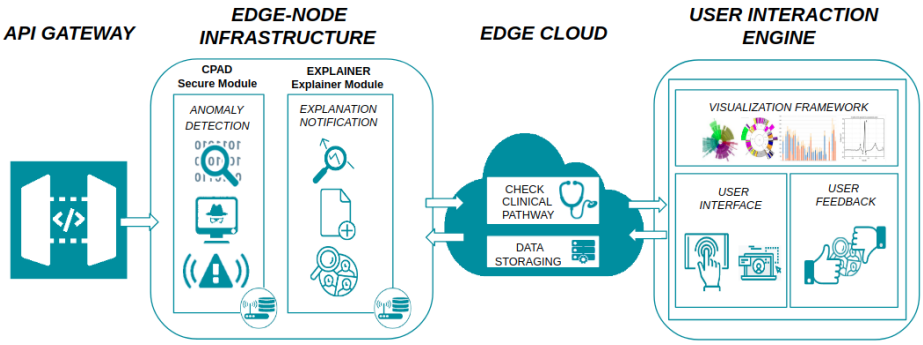

Fig. 1. Cyberattack Detection System Architecture

Below is a case study, where an e-Health telemonitoring system is, in turn, monitored by a CADS at the Edge. Let's consider a smart ECG device, which collects heartbeat information from a patient in remote assistance.

3 ECG User Interaction and User Feedback

The Interactive Machine Learning (IML) research domain incorporates human feedback into the model training process to develop high-performance ML models [4]. It is experience research again momentum [5].

Thanks to the use of examples provided by the system and human feedback [10, 11], it is possible to allow the algorithm to learn how best to behave when faced with a given ECG detection. In our case study, the contribution of the User Feedback to the system is the evaluation of the detected anomaly and the embedding of a doctor's feedback. The User Feedback module will generate for each detection ECG_i a feedback coefficient ϕ_i that represents the doctor's feedback on a given instance.

Definition 1 *Let ECG be the set of the heartbeat detections received from a smart ECG monitoring end-device, ECG_i be the i-th heartbeat detection. Then, the feedback coefficient is a function ϕ: $ECG \mapsto \{-1, 1\}$ such that any i-th user feedback related to the heartbeat detection ECG_i, is defined as follows:*

$$\phi_i = \begin{cases} +1 & \text{if } ECG_i \text{ is false positive or false negative} \\ -1 & \text{if } ECG_i \text{ is true positive or true negative} \end{cases} \quad (1)$$

Therefore, the User Feedback UF is a set of tuples such that, for any i-th pair of arguments (ECG_i, ϕ_i), a single element UF_i is defined as:

$$UF_i = (ECG_i, \phi_i) \quad (2)$$

In this way, the CADS will become more robust to external cyberattacks, since the User Feedback would report the opinion of the caregiver which will

confirm or not whether the *i*-the ECG detection is abnormal or not. In the User Interaction Engine, the *Visualisation Framework* represents the data orchestrator, handling and visualizing processed data coming from the various modules. It uses algorithms of VDM [6] that allow, through different visualization techniques, to interactively group data in a more efficient way, improving the data insight process. Afterwards, the *User Interface (UI)* included in the User Interaction Engine allows the user to interact with the data. In the case study, the UI allows the caregiver to interact with the ECG instances. In the CADS architecture, an Explanation module displays useful classification information, with which it is possible to visually manage each ECG detection. For instance, one would be able to no longer consider an ECG instance as an anomaly, or, more specifically, to improve the classifier performances by indicating the correct class of anomaly when a wrong one has been predicted. The interaction with the user, in this case a doctor, helps the system to be more and more reliable, as well as secure from cyberattacks.

Through the integration of CPAD, Explainer, and User Interaction Engine modules, the Visualization Framework will be able to manage anomalies detected as threat insights. These will be appropriately displayed on the UI which, in addition to allowing interaction with the anomalous data (in this case the ECG detection), will be able to display the threat representation through a dashboard. Thanks to the threats graphical representation in the dashboard, the user's reaction to the threat is improved. The visual process, whereby the healthcare professional is able to mark a detection as true or not, is a key scenario in which ML methods are combined with human feedback through interactive visualization. This process enables the fast prototyping of the ML model that can improve both the performance of the algorithm and human feedback. It will be also able to complete tasks where anomaly identification was not yet possible. The system then uses User Feedback to refine detection results and guide further analysis. Caregiver Feedback is therefore used as an essential source of ever-improving anomaly detection of ECG, which means that labeling the local environment will trigger global updates and thus guide further analysis.

4 Conclusion

An e-Health telemonitoring system cannot always have a security expert managing the security of the system. The proposed User Interaction and, in particular, User Feedback modules introduced in an AI-based CADS provide a visual representation of the results to the expert user engaged in a feedback response, determining whether the detected anomalous data are truly atypical by assessing the detection with positive feedback.Otherwise, the user provides negative feedback. Such interaction with a CADS has an impact on the processed health data, adjusting the visualization reports with the corrected measurements, shown in a useful dashboard. Future works will focus on the development of a further visual interface in which, thanks to the use of ML and VDM algorithms, it is possible to graphically represent both machine capability and human intelligence.

Acknowledgments

This work was partially funded by the European Union, Horizon 2020 research and innovation programme, through the ECHO project (grant agreement no 830943) and by the Italian P.O. Puglia FESR 2014 – 2020 (project code 6ESURE5) SECURE SAFE APULIA.

References

1. AfzaliSeresht, N., Liu, Q., Miao, Y.: An explainable intelligence model for security event analysis. In: Australasian Joint Conference on Artificial Intelligence. pp. 315–327. Springer (2019)
2. Ardito, C., Di Noia, T., Di Sciascio, E., Lofù, D., Pazienza, A., Vitulano, F.: An artificial intelligence cyberattack detection system to improve threat reaction in e-health. In: Proceedings of Italian Conference on Cybersecurity (ITASEC 2021) (2021), to be published.
3. Ardito, C., Noia, T.D., Fasciano, C., Lofú, D., Macchiarulo, N., Mallardi, G., Pazienza, A., Vitulano, F.: Towards a situation awareness for ehealth in ageing society. In: Proceedings of the Italian Workshop on Artificial Intelligence for an Ageing Society (AIxAS 2020). pp. 40–55 (2020)
4. Fails, J.A., Olsen Jr, D.R.: Interactive machine learning. In: Proceedings of the 8th international conference on Intelligent user interfaces. pp. 39–45 (2003)
5. Koch, J., Taffin, N., Beaudouin-Lafon, M., Laine, M., Lucero, A., Mackay, W.E.: Imagesense: An intelligent collaborative ideation tool to support diverse human-computer partnerships. Proceedings of the ACM on Human-Computer Interaction **4**(CSCW1), 1–27 (2020)
6. Kreuseler, M., Nocke, T., Schumann, H.: A history mechanism for visual data mining. In: IEEE Symposium on Information Visualization. pp. 49–56 (2004)
7. Luo, Y., Xiao, Y., Cheng, L., Peng, G., Yao, D.D.: Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities. arXiv preprint arXiv:2003.13213 (2020)
8. Pazienza, A., Anglani, R., Mallardi, G., Fasciano, C., Noviello, P., Tatulli, C., Vitulano, F.: Adaptive critical care intervention in the internet of medical things. In: 2020 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS). pp. 1–8. IEEE (2020)
9. Pazienza, A., Mallardi, G., Fasciano, C., Vitulano, F.: Artificial intelligence on edge computing: a healthcare scenario in ambient assisted living. In: Proceedings of the 5th Italian Workshop on Artificial Intelligence for Ambient Assisted Living 2019, AI*AAL@AI*IA 2019. pp. 22–37 (2019)
10. Shi, Y., Xu, M., Zhao, R., Fu, H., Wu, T., Cao, N.: Interactive context-aware anomaly detection guided by user feedback. IEEE Transactions on Human-Machine Systems **49**(6), 550–559 (2019)
11. Stumpf, S., Rajaram, V., Li, L., Burnett, M., Dietterich, T., Sullivan, E., Drummond, R., Herlocker, J.: Toward harnessing user feedback for machine learning. In: Proceedings of the 12th international conference on Intelligent user interfaces. pp. 82–91 (2007)
12. Yang, G., Jan, M.A., Menon, V.G., Shynu, P., Aimal, M.M., Alshehri, M.D.: A centralized cluster-based hierarchical approach for green communication in a smart healthcare system. IEEE Access **8**, 101464–101475 (2020)