e-ISSN: 3027-0650
Vol. 2, Issue 1, pp254-261, November 6 -9, 2024

An Adaptive and Scalable Ontology for Explainable Deep Classifier in Disease Surveillance

Kamal Bakari Jillahi[1], Aamo Iorliam[2], Gabriel Mshelia Mwajim[3], Shuaibu Anas[4]

{kamal.bakari@aun.edu.ng, aamo.iorliam@aun.edu.ng, gabriel.mshelia@aun.edu.ng}

Department of Computer Science, American University of Nigeria, Yola[1,3,4], Department of Data Science, American University of Nigeria, Yola[2]

Abstract. This research aims to improve explainability of predictions in disease surveillance by leveraging an ontology-based model. A Markov Decision Process (MDP) and a Q-Learning algorithms were proposed to update two public Ontologies making them both dynamic and Scalable in order to enhance the quality of explanations generated on the output of a deep learning classifier used for Morbidity/Mortality prediction of Malaria disease. The study uses Atlas Malaria dataset, OBO Malaria Ontology, SWEET Ontology and a Recurrent Neural Network thus, integrating domain-specific knowledge and data. The study compares the proposed model with a static model based on fidelity, interpretability, relevance, ROC and AUC metrics. The proposed model achieves a fidelity score of 0.92, compared to 0.75 for the static model, along with a higher interpretability score of 4.7/5 versus 3.9/5 for the static approach. Additionally, the relevance score for the dynamic ontology is 0.88, outperforming the static model's 0.72. The dynamic ontology also exhibits superior classification performance, with an AUC of 0.9532, significantly higher than the static model's AUC of 0.7968. These results demonstrate the dynamic ontology's effectiveness in improving both model performance and explanation quality in case studied.

Keywords: Evolving knowledge graphs, White-box AI, Semantic network, Ethical AI, Clarity, Public Health Surveillance.

1 Introduction

In the context of Artificial Intelligence (AI), explainability refers to the ability of a model to provide details or reasons to make clear how and why it made a specific decision or prediction. Explainability in AI systems boosts trust, transparency, and accountability by making them more understandable to users, decision-makers, and regulators [1]. It ensures fairness, detects biases, and improves model reliability. In fields like healthcare, security, finance, and law, explainability is crucial for validating AI's safety and ethical use [2]. It also fosters better human-AI collaboration, allowing users to refine, correct, and improve models, leading to more robust and trustworthy AI applications. Furthermore, it also allows for transparent, interpretable insight, ensuring evidence-based interventions [3]. For example, when an AI model predicts an imminent outbreak, explainability helps epidemiologists understand the factors driving the prediction, such as symptoms, environmental factors, or human mobility patterns. This transparency is essential for making informed decisions about public health measures. Furthermore, Explainable AI addresses ethical concerns in disease surveillance by identifying and mitigating biases in data. Whence, Black-box models can lead to unfair targeting of certain populations or regions. Explainable AI enhances accountability by allowing stakeholders to trace back AI decisions, justifying actions during health crises. This trust in AI-driven public health tools encourages their wider adoption and integration into real-time disease surveillance systems, thereby reducing the risk of biased decisions.

Ontology is a tool that is used in explainable AI to provide structured, domain-specific knowledge. This approach provides context-aware interpretations of AI decisions, especially in complex domains like healthcare, law, and scientific research. Ontologies also improve the consistency and transparency of AI models by aligning predictions with domain knowledge, reducing errors or biases [4]. Ontology-based explanations also facilitate better communication between AI systems and human experts, ensuring understandable and actionable reasoning behind AI decisions. Despite their benefits, ontology-based explainable classifiers face challenges related to scalability and dynamism [5]. Large, static ontologies can become outdated and cumbersome to manage, particularly when integrated with high-dimensional AI models in real-time surveillance systems [6][7]. This can lead to computational inefficiencies, slowing down the decision-making process, which is critical in the context of disease outbreaks. Additionally, the dynamic nature of disease surveillance, where new diseases or emerging variants continually shift the landscape, requires ontologies to be updated regularly [8][9]. Managing these updates in real time is a significant technical challenge, as outdated ontologies could lead to incorrect or irrelevant predictions [10]. For example [11], posited that future research in this area could focus on developing more adaptive ontology management systems that can evolve alongside AI models, ensuring that both remain accurate and relevant in rapidly changing environments.

This study aims to create an adaptive and scalable Ontology for Mobidity/Mortality risk prediction for Malaria disease, ensuring adaptability and reliability in response to evolving disease dynamics. It aims to enhance transparency, trust, and decision-making by providing domain agreed, context-aware explanations, reducing bias, and promoting fairness in AI-driven decisions.

2 Review of Literature

In disease surveillance, artificial intelligence (AI) has become an indispensable tool for predicting outbreaks, identifying high-risk areas, and managing public health resources [10]. However, the increasing complexity of AI systems necessitates improved transparency and interpretability, since decisions based on these AI predictions can have lives-affecting consequences [12]. Explainable AI (XAI) aims to make the decision-making process of AI systems more understandable to humans, enhancing trust and accountability [13]. Ontologies as structured representations of domain-specific knowledge, are increasingly being explored as a means of improving the explainability of AI classifiers [14][15]. These knowledge-based systems can provide deeper, more contextually relevant explanations, which are especially useful in disease surveillance, where understanding the factors driving predictions is essential for timely and accurate public health interventions [16].

Other explainable AI methods used in disease surveillance, such as model-agnostic techniques like LIME, SHAP, or feature importance analysis like Heap-maps, have notable weaknesses compared to ontology-based approaches [17][18][19]. These methods often focus on explaining individual predictions based on correlations or approximations between input and output rather than leveraging structured domain knowledge, which can lead to explanations that lack context or domain relevance [22]. For example, in disease surveillance, these methods might indicate that certain symptoms or demographic features contribute to a prediction without providing deeper insights into how these factors interact within the disease transmission process. Additionally, model-agnostic methods may struggle with transparency in complex, high-dimensional models like deep neural networks, offering only superficial or generalized explanations that may not align with public health expertise [23] [24]. This can lead to a disconnect between the AI model's outputs and the actionable insights needed for effective disease control, making it harder for public health professionals to trust and act on the predictions.

Ontologies have long been used in healthcare for organizing and standardizing medical knowledge [25][26]. Resources such as OBO Foundry, SNOMED CT, ICD-10, and MeSH offer structured ways of representing symptoms, diagnoses, and treatments, enabling better data interoperability across healthcare systems [21]. In the context of AI, ontologies serve as a foundation for integrating domain knowledge directly into classifiers, enhancing the interpretability of their decisions [6]. In disease surveillance, ontologies can be used to represent relationships between symptoms, transmission vectors, environmental factors, and other epidemiological variables, providing AI models with a robust framework for classifying diseases and predicting outbreaks [27]. For instance, during the COVID-19 pandemic, ontology-driven models helped integrate real-time clinical data with historical outbreak patterns to better understand the spread of the virus and inform intervention strategies [10].

Traditional XAI methods like LIME and SHAP have been applied in disease surveillance to make AI predictions more transparent [13]. However, these methods often fall short when it comes to providing domain-specific explanations, relying heavily on statistical correlations rather than leveraging expert knowledge. Ontology-based classifiers offer a solution by embedding structured knowledge directly into AI models, allowing for explanations that are both context-aware and grounded in epidemiological expertise. For example, an ontology-driven AI system might classify a region as high-risk for a disease outbreak based on a combination of environmental factors, human mobility patterns, and local healthcare capacity. By explaining these relationships in terms that align with established disease transmission models, ontology-based systems provide public health officials with clearer insights into the factors driving AI predictions [16]. Research in this area demonstrates that such explanations can lead to more actionable and informed decisions, such as targeted vaccination campaigns or resource allocation [20].

Scalability is a significant challenge when using ontologies for explainable AI, particularly in large, dynamic domains like disease surveillance. Ontologies, which represent domain knowledge through structured concepts and relationships, can become exceedingly complex as the scope of the domain expands. Managing and querying large ontologies in real-time can lead to performance bottlenecks, especially when integrated with deep learning models that process vast amounts of data. As ontologies grow, the computational resources required to maintain and utilize them also increase, posing difficulties for systems that need to provide quick, on-demand explanations. Additionally, integrating large ontologies with high-dimensional machine learning models requires sophisticated optimization techniques to ensure that the AI system remains efficient and scalable.

Dynamism is another issue, as ontologies must evolve with changing knowledge and real-time data. In fast-moving fields like disease surveillance, where new diseases, symptoms, or environmental factors constantly emerge, static ontologies quickly become outdated. AI models relying on such ontologies may produce explanations that no longer align with the latest scientific understanding or surveillance data. Maintaining up-to-date and relevant ontologies requires systems that can dynamically adjust and incorporate new information without manual intervention. This introduces further complexity in the form of real-time ontology updating, knowledge reconciliation, and conflict resolution, which are necessary to ensure that AI systems continue to provide accurate and reliable explanations as the underlying knowledge base evolves.

Ontology-based explainable classifiers represent a promising approach for enhancing transparency and trust in AI systems used for disease surveillance. By grounding AI models in structured, domain-specific knowledge, these classifiers provide explanations that are both interpretable and contextually relevant to public health professionals. This enables more accurate, trustworthy, and actionable decision-making, particularly in high-stakes situations like epidemic outbreaks [16][28]. However, challenges related to scalability and the dynamic updating of ontologies remain significant, highlighting the need for further research in creating adaptable, scalable systems that can evolve with emerging health threats [6][29].

3 Data and Methods

This work aims to develop a dynamic ontology to provide explanations for deep learning classifier models, such as Recurrent Neural Networks (RNNs), specifically for morbidity/mortality risk prediction tasks for Malaria disease. The model will uses Malaria Ontology and Semantic Web for Earth and Environment Technology (SWEET) Ontology to build an ontology-based explanations model which provide explanations on features warranting for the classicification, focusing on why certain individuals are at higher risk of mobidity or mortality.

Data gathered from reliable sources, such as news reports on disease burden, prevalence, and mortality rate across different geographical locations, are used to keep the Malaria Ontology and Semantic Web for Earth and Environment Technology (SWEET) Ontology updated. These ontologies serve as the knowledge base, enabling the system to interpret and explain the factors influencing the prediction of the classifier.

To ensure the ontologies remain adaptable and scalable, a Q-learning algorithm is developed. This algorithm decides when and how to update the ontologies based on a utility function it learns over time. By weighing the benefits of updating the ontology against potential system overload or irrelevant changes, that will help the system to maintain efficiency and accuracy in prediction. This ensures a more adaptable and explainable deep learning system that can dynamically adjust to new data and evolving disease patterns. Hence, this is presented in figure 1.

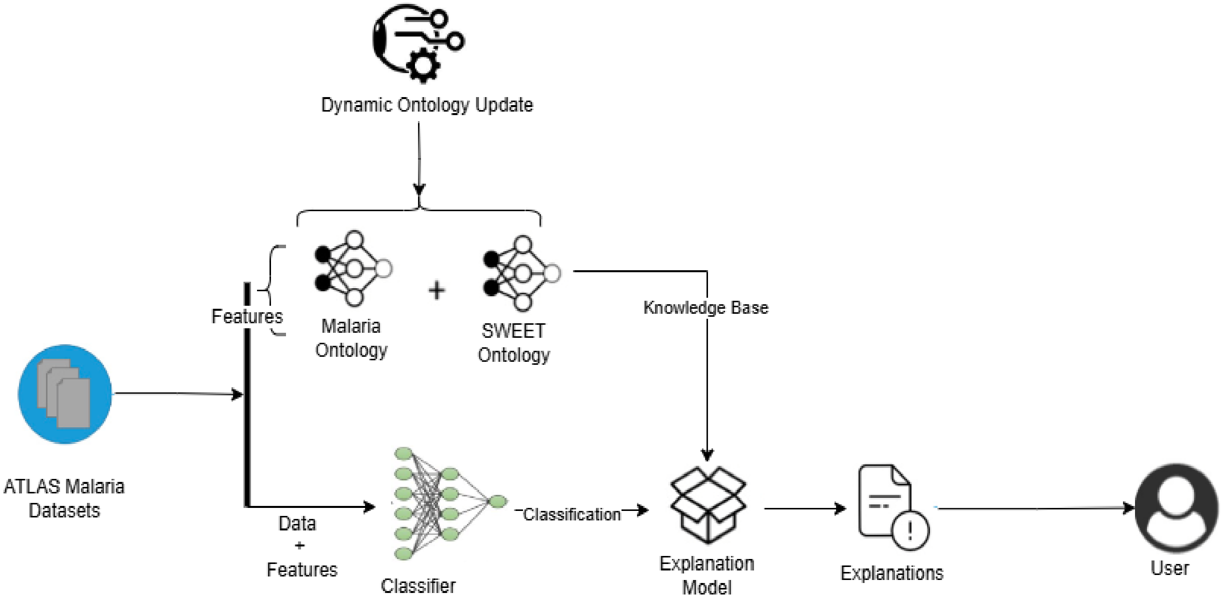

Figure 1: Dynamic Ontology Update Mechanism

Dataset/Ontologies

1. **Atlas Project Malaria Data**
   This dataset provides comprehensive global maps and data related to malaria transmission, prevalence, incidence and environmental factors that affect mosquito populations, with a strong focus on sub-Saharan Africa. The dataset is widely used in predictive models for malaria transmission, hotspot identification, and mortality/morbidity risk classification.
   URL: https://data.malariaatlas.org/

2. **Malaria Ontology**:
   The **Malaria Ontology** focuses on malaria-related knowledge, including the biology of the *Plasmodium* parasite, mosquito vectors, clinical manifestations, treatment, and prevention strategies. It organizes and standardizes data for malaria research and interventions, making it easier to analyze and integrate data from different sources.
   URL: http://purl.obolibrary.org/obo/IDOMAL_0002350

3. **Semantic Web of Earth and Environment Terms (SWEET) Ontology**:
   The **SWEET Ontology** is a comprehensive framework for representing Earth and environmental sciences, including geophysical phenomena, climate, and ecosystems. It supports integration and sharing of environmental data across domains by providing a common vocabulary. SWEET is essential for projects related to climate change, environmental monitoring, and sustainability, ensuring interdisciplinary data can be effectively linked and understood.
   URL: https://www.earthdata.nasa.gov/community/sweet

4 Methods

Problem Definition:

- Let S be the state space, where each state $s \in S$ represents a specific state of the ontology (e.g., concepts, relationships, structures).
- Let A be the action space, where each action $a \in A$ corresponds to a potential modification of the ontology (e.g., adding, modifying, or deleting concepts/relations).
- The agent's goal is to find a policy π:S→A, which maximizes the expected cumulative reward R, by updating the ontology dynamically as new data becomes available.

Markov Decision Process (MDP):

- **State**: $s_t \in S$ at time t, representing the current state of the ontology.
- **Action**: $a_t \in A$, an action which modifies the ontology's structure at time t.
- **Transition Function**: $T(s_{t+1} | s_t, a_t)$ represents the probability of moving from state s_t to state s_{t+1} after taking action a_t.
- **Reward**: $r_t = R(s_t, a_t)$, a scalar reward obtained after taking action a_t in state s_t.

Q-Learning Algorithm

The Q-learning algorithm will be used to learn an action-value function Q(s, a), which represents the expected cumulative reward of taking action *a* in state *s*, and following the optimal policy thereafter. The algorithm is presented below:

1. **Initialize the Q-table**:

$$Q(s,a) \longleftarrow 0 \quad \forall s \in S, \forall a \in A$$

2. **Policy**: An ϵ-greedy policy is used to balance exploration and exploitation:

$$a_t = \begin{cases} \text{random action with probability } \epsilon, \\ \max_a Q(s_t, a) \text{with probability } 1 - \epsilon \end{cases}$$

3. **Q-Value Update**: After taking action a_t in state s_t, receiving reward r_t, and observing the next state s_{t+1}, the Q-values are updated as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$$

where:

- $\alpha \in [0,1]$ is the learning rate,
- $\gamma \in [0,1]$ is the discount factor,
- $\max_a Q(s_{t+1}, a)$ is the maximum Q-value for the next state s_{t+1}, which corresponds to the best possible future action.

4. **State Transition**: The next state s_{t+1} is determined by the transition dynamics of the environment:

$$s_{t+1} \sim T(s_{t+1} | s_t, a_t)$$

In this case, this refers to the ontology being updated based on the action a_t, and how the new data or structural modification impacts the ontology.

5. **Termination**: The process repeats until a terminal state or the end of a predefined number of episodes, after which the Q-values reflect the optimal policy.

5 Results

The performance of the system was assessed using algorithmic and human centered metrics. For the algorithmic metrics, **Receiver Operating Curve (ROC) and Area Under Curve (AUC)** were used while for the human centered metrics: **Fidelity, Interpretability**, and **Relevance**. The results are presented below:

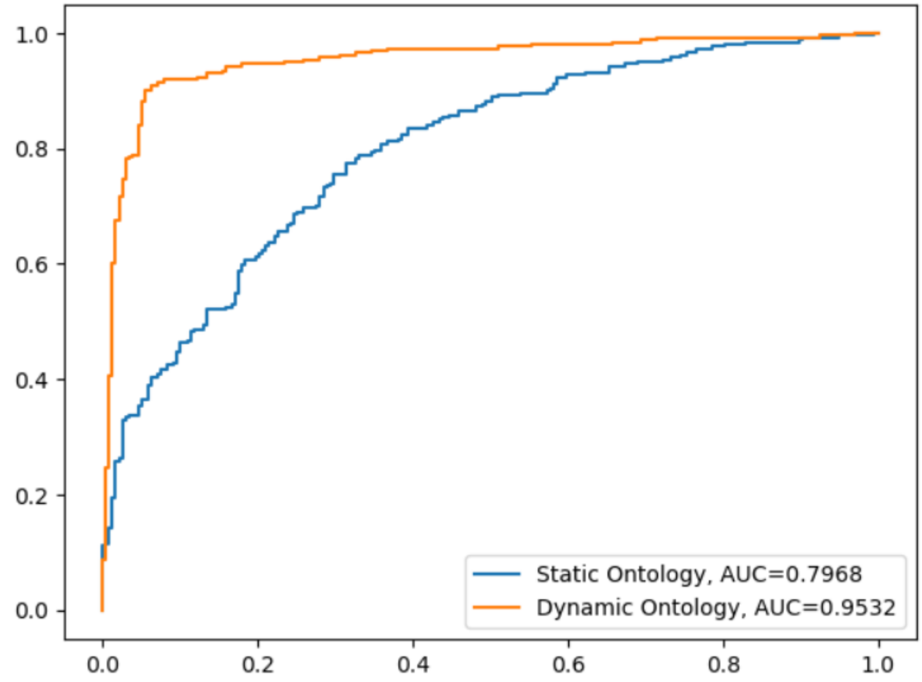

Figure 2: Comparision of the proposed dynamic ontology and static ontology

From Figure 2 above, the ROC curve compares the proposed model with static model: the **Static Ontology** (blue curve) and the **Dynamic Ontology** (orange curve), with the **AUC (Area Under the Curve)** as the primary performance metric. The **Static Ontology** classifier yields an **AUC of 0.7968**, indicating moderate classification capability, but it struggles to minimize false positives while maintaining high true positive rates, as reflected by the curvature away from the top-left. The **Dynamic Ontology** classifier significantly outperforms the static version, with an **AUC of 0.9532**. This higher AUC demonstrates the model's ability to more accurately balance between detecting true positives and minimizing false positives across thresholds. The curve for the dynamic model approaches the ideal performance boundary, suggesting better discriminative power and reduced error rates. Thus, the dynamic ontology-based approach shows superior classification performance, offering more reliable predictions.

Human Centered Evaluation

Fidelity

Fidelity refers to how accurately the explanations generated by the ontology reflect the decision-making process of the deep classifier. Thus, high fidelity indicates that the ontology explanations closely match the classifier’s internal logic, making the model’s decision-making transparent and understandable. Here, fidelity was quantified using a fidelity score that represents the proportion of cases where the explanation matched the classifier’s output. The Proposed system achieved an average fidelity score of **0.92** compared to 0.75 achieved by the static model. This means that the dynamic ontology-based explanations accurately represented the decision-making process in 92% of the cases, closely aligning with the classifier’s outputs. This high fidelity score suggests that the system reliably mirrors the model’s reasoning, allowing users to trust the explanations as faithful representations of the classifier's predictions.

Interpretability

Interpretability measures how understandable the explanations are to human. The goal was to ensure that the explanations could be easily interpreted and applied in real-world disease control settings. A group of 20 individuals were asked to evaluate the interpretability of explanations generated by the ontology system. Each individual reviewed 10 disease detection cases and rated the interpretability on a Likert scale from 1 (very difficult to understand) to 5 (very easy to understand). The average interpretability rating was **4.7/5** compared to 3.9/5 for the static based ontology system, indicating that the majority of individuals found the explanations to be highly understandable and actionable. Participants noted that although both ontology's explanations provided clear terms, the dynamic system provided concise rationales behind the deep classifier's decisions, particularly in identifying up to date correlations between symptoms, demographics, and disease risk factor of individuals.

Relevance

An explanation is considered relevant if it provides actionable insights that enhance the user's understanding of the underlying data and model predictions. The relevance of explanations was assessed by analyzing feedback from users. A relevance score was computed based on the proportion of cases where a user deemed the explanations directly applicable to model decision process. An average relevance score of **0.88** outperforming the static model which achieved a relevance score of 0.72, indicating that 88% of the explanations provided meaningful and applicable insights for disease surveillance.

The Performance of the two systems in all the three human centered metrics, are summarized in Table 1 below.

Metric	**Dynamic Ontology Explanations**	**Static Ontology Explanations**
Area Under Curve	0.9532	0.7968
Fidelity	0.92	0.75
Interpretability	4.7/5	3.9/5
Relevance	0.88	0.72

Table 1: Comparative performance of Dynamic ontology-based explanations versus traditional Static methods.

6 Discussion

The results demonstrate the efficacy of the proposed Dynamic and Scalable ontology-based explanation framework in improving the transparency and usability of deep classifiers in disease surveillance. The model's accuracy and high fidelity score ensures that the explanations accurately reflect the classifier's internal logic, while the interpretability and relevance metrics confirm that the system provides understandable and useful insights for users. Furthermore, the system guarantees alignment with domain knowledge since the ontologies are domain agreements.

References

[1] Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(suppl_1), D267-D270.

[2] Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule-based expert systems: The MYCIN experiments of the Stanford heuristic programming project*. Addison-Wesley.

[3] Shaban-Nejad, A., Michalowski, M., & Buckeridge, D. L. (2017). Health intelligence: how artificial intelligence transforms population and personalized health. *NPJ Digital Medicine*, 1(1), 1-5.

[4] Chi, Y. L., & Goldstein, B. A. (2021). Explainable artificial intelligence in public health: A scoping review. *Journal of Public Health Management and Practice*, 27(2), E83-E92.

[5] Jiménez-Ruiz, E., & Grau, B. C. (2011). LogMap: Logic-based and scalable ontology matching. *Proceedings of the 10th International Semantic Web Conference (ISWC)*, 273-288.

[6] Shaban-Nejad, A., & Lavigne, M. (2018). Ontology-based real-time surveillance system for preventing disease outbreaks. *Journal of Biomedical Semantics*, 9(1), 1-12.

[7] Rosse, C., & Mejino Jr, J. L. (2003). A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of Biomedical Informatics*, 36(6), 478-500.

[8] Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics*, 77(2), 81-97.

[9] Grimm, S. (2009). Knowledge representation and ontologies. In Scientific data mining and knowledge discovery: Principles and foundations (pp. 111-137). Berlin, Heidelberg: Springer Berlin Heidelberg.

[10] Walter, T., Parreiras, F. S., & Staab, S. (2014). An ontology-based framework for domain-specific modeling. Software & Systems Modeling, 13(1), 83-108.

[11] Ayranci, P., Lai, P., Phan, N., Hu, H., Kolinowski, A., Newman, D., & Dou, D. (2022). OnML: an ontology-based approach for interpretable machine learning. Journal of Combinatorial Optimization, 44(1), 770-793.

[12] Sharma, S., & Jain, S. (2024). OntoXAI: a semantic web rule language approach for explainable artificial intelligence. Cluster Computing, 1-25.

[13] Yildirim, M., Okay, F. Y., & Özdemir, S. (2024). A comparative analysis on the reliability of interpretable machine learning. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 30(4), 494-508.

[14] Tasioulis, T., & Karatzas, K. (2023). Reviewing Explainable Artificial Intelligence Towards Better Air Quality Modelling. In Environmental Informatics (pp. 3-19). Cham: Springer Nature Switzerland.

[15] Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., ... & Bischl, B. (2020, July). General pitfalls of model-agnostic interpretation methods for machine learning models. In International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers (pp. 39-68). Cham: Springer International Publishing.

[16] Kaushik, R. Explainability in Machine Learning: Bridging the Gap Between Model Complexity and Interpretability. Edu Journal of International Affairs and Research, ISSN, 2583-9993.

[17] Liyanage, H., Krause, P., & De Lusignan, S. (2015). Using ontologies to improve semantic interoperability in health data. BMJ Health & Care Informatics, 22(2).

[18] Naqvi, M. R., Elmhadhbi, L., Sarkar, A., Archimede, B., & Karray, M. H. (2024). Survey on ontology-based explainable AI in manufacturing. Journal of Intelligent Manufacturing, 1-23.

[19] Tiddi, I., & Schlobach, S. (2022). Knowledge graphs as tools for explainable machine learning: A survey. Artificial Intelligence, 302, 103627.

[20] Gardy, J. L., & Loman, N. J. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. Nature Reviews Genetics, 19(1), 9-20.

[21] Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R., & Krogan, N. J. (2020). A systems approach to infectious disease. Nature Reviews Genetics, 21(6), 339-354.

[22] Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., & Gurram, P. (2017, August). Interpretability of deep learning models: A survey of results. In 2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI) (pp. 1-6). IEEE.

[23] Ivanović, M., & Budimac, Z. (2014). An overview of ontologies and data resources in medical domains. Expert Systems with Applications, 41(11), 5158-5166.

[24] Jurisica, I., Mylopoulos, J., & Yu, E. (2004). Ontologies for knowledge management: an information systems perspective. Knowledge and Information systems, 6, 380-401.

[25] Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021, May). Expanding explainability: Towards social transparency in ai systems. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1-19).

[26] Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., ... & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Information Fusion, 96, 156-191.

[27]Rasheed, K., Qayyum, A., Ghaly, M., Al- Fuqaha, A., Razi, A., & Qadir, J. (2022). Explainable, trustworthy, and ethical machine learning for healthcare: A survey. Computers in Biology and Medicine, 149, 106043.

[28] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.