Journal of Applied Business and Technology

Increasing Trust in AI with Explainable Artificial Intelligence (XAI): A Literature Review

Dewi Nasien[a*], M. Hasmil Adiya[a], Devi Willeam Anggara[b], Zirawani Baharum[c], Azliza Yacob[d], Ummi Sri Rahmadhani[e]

[a]Department of Computer Science, Institut Bisnis dan Teknologi Pelita Indonesia, Indonesia
[b]School of Electrical and Informatics Engineering, Institut Teknologi Bandung, Indonesia
[c]Technical Foundation section, Universiti Kuala Lumpur, Malaysia
[d]Department of Computer Science, University College TATI, Malaysia
[e]Department of Electrical Engineering, Universitas Riau, Indonesia

Article History
Received
30 June 2024
Received in revised form
22 August 2024
Accepted
26 September 2024
Published Online
30 September 2024

*Corresponding author
dewinasien@lecturer.pelitaindonesia.ac.id

Abstract
Artificial Intelligence (AI) is one of the most versatile technologies ever to exist so far. Its application spans as wide as the mind can imagine: science, art, medicine, business, law, education, and more. Although very advanced, AI lacks one key aspect that makes its contribution to specific fields often limited, which is transparency. As it grows in complexity, the programming of AI is becoming too complex to comprehend, thus making its process a "black box" in which humans cannot trace how the result came about. This lack of transparency makes AI not auditable, unaccountable, and untrustworthy. With the development of XAI, AI can now play a more significant role in regulated and complex domains. For example, XAI improves risk assessment in finance by making credit evaluation transparent. An essential application of XAI is in medicine, where more clarity of decision-making increases reliability and accountability in diagnosis tools. Explainable Artificial Intelligence (XAI) bridges this gap. It is an approach that makes the process of AI algorithms comprehensible for people. Explainable Artificial Intelligence (XAI) is the bridge that closes this gap. It is a method that unveils the process behind AI algorithms comprehensibly to humans. This allows institutions to be more responsible in developing AI and for stakeholders to put more trust in AI. Owing to the development of XAI, the technology can now further its contributions in legally regulated and deeply profound fields.

Keywords: XAI, Artificial Intelligence, Machine Learning, XAI Model, XAI Implementation
DOI: https://doi.org/10.35145/jabt.v5i3.193
SDGs: Quality Education (4); Decent Work and Economic Growth (8); Peace, Justice and Strong Institutions (16)

1.0 INTRODUCTION

Artificial Intelligence (AI), as the name suggests, has a mind of its own. As it continues to advance and become more complex, it has become more and more of a challenge for humans to understand its inner workings. As of right now, AI models are considered "black boxes" due to their level of complexity, which makes them incomprehensible to humans. It is now difficult and sometimes impossible to trace how an AI's algorithm ended up with certain results or decisions. These models were created with the data directly, therefore not even the developers can tell exactly how the AI gave that conclusion. Explainable Artificial Intelligence (XAI) is a method of AI that deals with this problem (Javed et al., 2023). In the context of AI, explainability means "The ability to present in understandable terms to humans", which encompasses the ability to make humans understand the reasons behind a decision or result. The goal is to "unbox" the black-box model and help users understand the model's behaviour. If the user is an expert in AI, then the XAI must assist further to help that user comprehend the model's structure and behaviour at a suitable level. Either way, XAI is the technology that clarifies how AI works (Nagahisarchoghaei et al., 2023). A recent study measured explainability in XAI by similarity and stability. Similarity measures objectivity and trustworthiness in XAI by comparing the system's output to a manual result made by the expert; for example, a radiology image processing result is compared to annotations from an actual Radiologist. Stability measures the consistency of XAI's outputs by presenting the learned processes and comparing the results between tests in all the used methods (Siddiqui & Doyle, 2022). Since AI's contribution to society has been

exponentially increasing, the need for XAI has become urgent. This is especially true in certain fields that are regulated by law and fields that can directly and deeply impact human lives. An obvious example of this is medicine. Traditional AI is not very suitable for medicine due to its lack of understandability to not only the patients but also the doctors themselves. It is such a shame because AI can enable accurate and quick disease diagnosis, which can save lives. However, in certain nuanced cases, AI can also be prone to misdiagnosis. This can be fatal. The XAI solves that problem by being a "white box". The XAI will explain the reasoning behind the produced outputs and the contribution of each feature in the disease prediction model. This not only helps everyone understand, but more importantly, it helps the physicians keep the system accountable and auditable. This also helps AI experts and non-AI experts communicate better in collaborations, thus improving the system continuously (Gurmessa & Jimma, 2023). Another example of this is finance. In finance, according to the Bank of England, explainability in XAI is defined as a system that allows every interested stakeholder to comprehend the main processes behind the model's decisions. There is even a guideline created by the European Commission High-Level Expert Group on AI presented in the Ethics Guidelines for Trustworthy Artificial Intelligence in April 2019, in which there are key requirements that an AI system must meet for it to be deemed trustworthy to use in financial services. Among the points, three of them represent the need for XAI. The first one is human agency and oversight, which means that other than each decision must be understood by humans, there must also be human involvement within the loop to oversee the process. Number two is transparency, which encompasses AI's responsibility to provide clarity and understanding of every step of the process to every stakeholder involved upon request. Lastly, accountability establishes that AI systems should be developed to be accountable and auditable whenever necessary (Bussmann et al., 2020).

The purpose of XAI expands beyond just being a tool for understanding the non-AI experts involved in AI-based projects. As a system, AI is very advanced, yet it is not flawless. It can lead to incorrect decisions, inaccurate answers, and errors. As a society, we have kept increasing the use of AI in daily human activities to the point that we even have self-driving cars. If there were a crash, wouldn't we want to know where in the AI process did it go wrong? When classification findings of an AI have the potential to cause harm, the ability to know exactly how and why it can happen becomes an urgent need so that the system can be fixed and incidents can be avoided. The need for XAI is the need for fair and ethical decision-making (Javed et al., 2023). Another purpose of XAI is as simple and as complex as it sounds; curiosity. What XAI does is provide justifications for decisions made by the algorithm, and what curiosity can do is learn from it. This facilitates learning and further model development, thus producing new and improved models. As the XAI method keeps being implemented, the results and explanations should improve and become more consistent. It should increase in level of comprehensibility to humans, and it should increase human's trust and confidence in the decision-making system. As the system increases in quality, it is hoped that the system will later also be able to provide workable solutions to imperfections that will maintain transparency, trust, and fairness (Nagahisarchoghaei et al., 2023). Specifically in the medical field, XAI helps provide clarity by methods of visualizing, explaining, and analyzing deep learning models (Taşcı, 2023). In psychiatry, the need for comprehensible methods is heightened even more. When the machine is responsible for describing the syndromes, outcomes, disorders, signs/symptoms and eventually becomes determinant of the diagnosis, treatment, and medications, there's no room for imperceivable decisions (Joyce et al., 2023). Applications of AI in the medical field have been advancing over the years. At this moment, AI-assisted medical image diagnoses and prognoses are available. This has led to the development of XAI for facilitated magnetic resonance imaging (MRI). The technology is relatively new, yet very promising to the future medical field (Qian et al, 2023). Healthcare, criminal justice, and autonomous driving are all high-stakes situations that should never be belittled. The involvement of AI sure has advanced these fields. However, explainability is a major factor that can bridge AI capabilities to operational needs (Hu et al., 2023). There are three known challenges to the complexity of AI models: (1) Interpreting the results, (2) ensuring the system works properly and is understandable, and (3) justifying the results to supervisors. Based on these points, experts and policymakers in Europe (EBA, European Union, European Union, Parliament and Council, and the European Commission) are creating AI-specific regulations. It is becoming very important to know exactly how each feature influences the predictions made by AI, thus the need for XAI continues to increase (de Lange et al., 2022).

2.0 METHODOLOGY

Methods of XAI

The main difference between AI and XAI is its presentability (Joyce et al., 2023), which stems from the main drivers of the technology itself; explainability and transparency (Fritz-Morgenthal et al., 2022), which supports the system to be responsible, comprehensible, auditable, traceable, and trustworthy. An XAI model should be able to explain

its capabilities and features to improve its usability, explain what it has done, what it is doing now, and what it will do next, and identify the critical information it relies on. Noting this concept in mind, here are some XAI methods:

1. Self-Explainable Model
   This model of XAI provides explainability from the beginning, which is at the training stage. It does so by building a decision tree or applying interpretability directly to the structure of the model, adding an attention layer to deep learning models. Intrinsic explainability can be achieved by Linear, and Logistic Regression, Decision Trees, K-Nearest Neighbors, Rule-Based Learners, General Additive Models, and Bayesian Models (Gurmessa & Jimma, 2023).
2. Post-Hoc Explainable Model
   This model requires creating a secondary model to explain an existing model after training the black box model. The main difference between self-explainable and post-hoc explainable models comes from the trade-off between fidelity and accuracy. An intrinsically explainable model could provide an accurate and undistorted explanation of the cost of sacrificing prediction performance. This model, in particular, can be divided into three distinct categories based on the explanation it provides, which are model explanation, outcome explanation, and model inspection (Gurmessa & Jimma, 2023).
3. Global Explainability Model
   This is a collaborative effort to increase transparency and trust in machine learning models. Users and developers can now examine worldwide data instances and understand how a model works. The two ways to construct globally interpretable models are using all data to fit the model or using all instances of a few features (Gurmessa & Jimma, 2023; Brusa et al., 2023). Global methods were selected: Kernel SHAP, Tree SHAP, Accumulated Local Effects (ALE) (Brusa et al., 2023).
4. Local Explainability Model
   The local explainability model, on the other hand, focuses on local interpretability, examines a model's individual prediction, figures out why the model makes its decision locally and helps uncover the causal relations between a specific input and its corresponding model prediction (Nagahisarchoghaei et al., 2023). A well-known local explanation method is called LIME. Lime can explain any black-box classifier with two or more classes. All it takes is for the classifier to implement a function that takes raw text or a numpy array and outputs the probability for each class. Other than LIME, there are also Anchor Explainer and Integrate Gradient Explainer (Brusa et al., 2023).
5. Model-Specific Explanation Method
   These are XAI models that are specifically built and can only be used with certain machine learning models. Model-specific techniques make a fast single pass back through the neural network (Qian et al., 2023).
6. Model-Agnostic Explanation Method
   Model-agnostic methods can be implemented across various machine-learning methods and are not specifically attached to a certain model. Model-agnostic explanation methods require an extensive perturbation of the input images to measure the change in the output caused by the perturbations. Model-agnostic techniques overwhelm model-specific techniques in terms of the potential of XAI techniques to be "plug-and-play". Consisting of perturbation-based visual explanation, model-agnostic techniques have the highest ease of use, enabling them to be applied to any trained neural network to provide a visual explanation (Qian et al., 2023).

XAI in Healthcare

In healthcare, there are three types of treatments given to patients: curative, symptomatic, and preventive. Curative treatment is given to cure the patient of the illness; this is very common, and results vary greatly. In certain illnesses, curative treatment is simple and effective. For some others, it is unfortunately difficult and can be ineffective at times (Tian et al., 2024). Symptomatic treatment relieves symptoms of the illness, making the patient more comfortable during recovery, but it does not cure the disease or eliminate the cause. In some cases, symptomatic treatment is the only treatment given because no medication is available to rid the cause of the illness, leaving the job to the patient's immune system. This is often the case for viral infections. Preventive treatment is given to prevent the illness before it happens. This can be done before the illness manifests itself, with or without the risk factor being present, but especially when it does. Preventive treatment can consist of screening, vaccination/immunization, supplementation, and medication only if necessary, should anything come up during the screening process (Pillay et al., 2024). The involvement of AI in healthcare is hoped to smooth out these processes, making it easier on the healthcare providers, the patients, and the families of the patients. The most common use of AI in healthcare is for screening purposes. For many conditions, medical imaging, such as MRI, is an effective detection method. In the data analysis of MRI results, AI has played an important part in performing classification and segmentation tasks. In this case, XAI has been developed to provide insights into the

process, which is incredibly valuable to transparency and understanding where the result came from and gives a chance to audit and enhance. The MRI is used for many parts of the human body, such as the brain, breast, liver, musculoskeletal, gastrointestinal, prostate, and whole body. Knowing how the system came up with the result is crucial to the trustworthiness of this technology (Qian et al., 2023).

Stroke is the second leading cause of death and a significant contributor to disability worldwide. There are two types of strokes, ischemic and hemorrhagic. Blood-clot blockages cause ischemic stroke, while hemorrhagic stroke is caused by rupture of weak blood vessels. In either case, the blood supply to the brain is disturbed, resulting in severe neurological impairment. Stroke is diagnosed through a thorough examination involving assessment of heart rate and blood pressure, blood tests for cholesterol and diabetes, cardiac evaluations, and brain assessments with computed tomography (CT) scans and magnetic resonance imaging (MRI) scans. These tests give important data to identify the cause of stroke, thus determining the next steps in the treatment plan. Treating stroke is a race against time; early detection and immediate care are crucial. As more and more people are at risk, the development of precise and effective prediction systems for early stroke detection becomes urgent. In this case, XAI can be a brilliant collaborator to assist doctors and radiologists in detecting stroke. The XAI can detect subtle differences in the imaging results; thus, early signs can be found, and early interventions can be done. With XAI, doctors would understand which feature contributes to the outcome, thus keeping the system accountable and auditable. This finding can improve approaches to stroke diagnosis and treatment at large (Gurmessa & Jimma, 2023).

Cancer diagnosis remains a puzzle in the medical field, yet its urgency is unquestionable. Getting the diagnosis as early as possible can make a huge difference in the treatment plan and chance of success. This makes cancer diagnosis ongoing research carried out by numerous institutions and authors, each striving to achieve a more efficient and accurate diagnostic method, prevention, and treatment option to increase the patient's chance of survival. Cancer diagnosis is typically achieved using medical imaging tests, such as MRI, mammograms, microscopic images, ultrasound, etc. Recently, computer-aided diagnosis (CAD) started to be used to enhance the cancer diagnosis process by assisting doctors in spotting cancer lesions, leading to higher accuracy and fewer human mistakes due to medical practitioners' fatigue. This also lessens the workload for doctors, allowing them to focus more on the care and treatment of the patient. To prevent misdiagnosis, which can happen in the cancer detection process, some CADs are built with the ability to differentiate normal and abnormal tissues, tumours and cancer, as well as malignant and benign lesions. This system is developed using AI, and has lately been transitioning into being developed using XAI techniques instead. Since cancer diagnosis is very detrimental to the survival of the patient, this system needs to be trustworthy in its diagnosing capabilities. Due to this, transparency, traceability, and auditability are important. That is why XAI is perfect for this function, as it helps doctors understand the system more deeply. In cancer detection, the XAI will utilize algorithms like deep learning to analyze medical images, and the results will be its diagnosis and an explanation of how the system concludes it that way. This can be done using the XAI highlighting specific areas of the image that make the system identify it as cancer lesions. Moreover, the XAI can also provide transparency for the patient and the patient's family (Alkhalaf et al., 2023).

A brain tumour can be detected faster and more accurately using XAI methods, and given their acute and fatal nature, any little difference in detection time can make a huge difference in the chance of survival. This system aids radiologists and eliminates the possibility of human error caused by healthcare worker's fatigue [6].

Alzheimer's Disease is a type of Dementia in which the patient's brain experiences damage and decrease in functionality, causing the patient to be forgetful, less willing to try new things, experience difficulty in decision-making, and can eventually lead to death. Alzheimer's Disease happens because of the death of neuron cells, causing a part of the brain to shrink in size (otherwise known as brain atrophy). Alzheimer's Disease is classified based on its severity; preclinical AD, mild cognitive impairment (MCI), and clinically diagnosed AD. An important note here is that a patient in the middle stage (MCI) has the potential to either recover or get worse. Patients in this stage have shown progress to get better and fully recover, but some have also gotten worse. Researchers are focusing on learning how to stop MCI from getting worse. Alzheimer's Disease is diagnosed through brain imaging. Lately, methods of machine learning (ML) have been employed. It's time XAI is used for this purpose so that not only the accuracy can be improved, but also audibility and trust in the system (Amoroso et al., 2023).

Brain-computer interface (BCI) is a system that converts neurological signals from the brain into commands that are understandable and actionable by a computer device. This system helps patients with neurological conditions by facilitating motor tasks according to the commands. There are various BCI devices currently in use, such as wheelchairs, prosthetics, robotic arms, and word processors. Many patients place their hopes in BCI. It helps people with terminal neurological conditions do daily activities independently. Conditions that require such help are experienced by millions of people in the world. Currently, people put hope in machine learning-based BCI as a non-invasive and better-quality performer of this task. It is hoped that this system can differentiate more kinds of motor tasks and perform each effectively just based on the command. A study in 2023

suggested a machine learning-based BCI capable of analyzing EEG signals directly from motor imagery to differentiate various motor tasks based on BCI competition III dataset IVa and optimize the features to distinguish the neural activity pattern better using the whale optimization algorithm (WOA), and then use machine learning to enhance precision of the EEG signal analysis by improving upon the chosen features. The system later used the LIME XAI technique to explain the individual contributions of the features in the predictions made by the model. This resulted in a huge success, with overall accuracy of 98.6% and with the XAI increasing trust and giving insights into the system's decisions (Hashem et al., 2023).

Functional brain development analysis is a process that has been done within the last decades using non-invasive and portable neuroimaging techniques, for example, functional Near- Infrared Spectroscopy (fNIRS), which has allowed researchers to study the functional development of the human brain, which has been contributing to the study of Developmental Cognitive Neuroscience (DCN). The challenge is that the fNIRS data are still quite limited. A new method developed in 2023 has emerged as a possible solution to this challenge. The innovation is called Fuzzy Cognitive Maps (EFCMs) for Effective Connectivity (EC) analysis of infants' fNIRS data. This technology can display interconnections between cortical areas and specify the direction of the EC. To complete this innovation, another method is proposed to give insights into the activation level of the cortical regions of the brain. This method is called Multivariate Pattern Analysis (MVPA), and XAI powers it. It can investigate visual and auditory processing in six-month-old infants with 67.69% accuracy. The system identifies patterns of cortical interactions that happen as a response to stimuli. This helps DCN studies tremendously. However, its abilities are limited to analysing cross-sectional DCN studies. To collaborate and complete these two existing innovations, a third system is developed, which is a novel time-dependent XAI (TXAI) system based on Temporal Type-2 Fuzzy Sets (TT2FS). This system is great for empirical studies and for real-life intelligent environment datasets to solve time-dependent classification problems. This system has a classification accuracy of 94.08%. The TXAI has the potential to report the evolution of the functional brain development process, as it uses temporal trajectories, which can assist in the delineation of brain developmental trajectories (Kiani, 2022).

Clinical Decision Support System (CDSS) is a type of computerized system that provides information based on patient data to be used as consideration by healthcare providers to make decisions on diagnosis and treatment. This is helpful as healthcare providers can use the information intelligently to improve the accuracy and effectiveness of the treatment. To achieve this, CDSSs are developed with the capabilities to diagnose, predict treatment response, give a personalized treatment plan, give a prognosis, and prioritize patient care according to the risks involved. This system is also helpful in care facilities with limited resources, in which the CDSS can give suggestions on the priorities objectively and accurately to accommodate everybody. This system can be made AI-based or knowledge-based. In the AI-based version, the system observes clinical data from the past to produce prediction models to examine the new input variables. The results given by CDSS are not to be taken directly as a treatment plan but are meant to be used as assistance or extra consideration by doctors. These recommendations can aid doctors in their practices, increase clinical choices and reduce human mistakes in medical diagnosis and treatments. It is also a fair system as it is objective and purely logic-based. However, any bias in training data can result in inaccurate prediction, which is why XAI is important for this kind of system so that the process can be monitored and audited as needed (Javed et al., 2023).

The potential of XAI in healthcare is endless. Other than being an invaluable diagnostic tool, XAI can be developed into a personalized connected healthcare system (Chandra et al., 2023) or mental health companion (Joyce et al, 2023). On a larger scale, with enough data, XAI can be developed into a medical image analysis and personalized medicine system based on medical records on a population level, with capabilities to analyze combined data from hospitals, and analyzed thoroughly by academic institutions (Galić et al., 2023). This is something impossible for AI without XAI due to transparency and non-audibility issues.

XAI in Business and Finance

Finance is a heavily regulated and guarded field due to its nature of significance and direct impact on people on a personal and population level. Today's economy is complicated, to say the least. It is a vast and continuously evolving field of study and professional work, much like technology. For a while, finance and tech go hand in hand in supporting each other's growth. However, with AI, it is a bit tough to put the two and two together since the algorithm is considered a 'black-box' model. In finance, transparency and integrity are essential. When the algorithm is too complex to be understood by even the developers and cannot be explained to the stakeholders, the system cannot be utilized. That is where XAI come in, with its explainability being the solution that bridges this gap in trust.

In banks, XAI is being developed for credit assessment. As we know, credit scoring models have to be explainable, as required by financial authorities. The XAI model aims to predict credit default on a bank dataset. The interpretation is made of explanatory variables affecting the prediction. The most important variables are the

volatility of the utilized credit balance, remaining credit in the percentage of total credit and the duration of the customer relationship. The main contributions of XAI in this regard are to further its implementation in banking, improve the interpretability and reliability of the models, and analyze the potential economic value of improved credit scoring models (de Lange et al., 2022). The XAI is also good for risk management, in Fintech for example (Bussmann et al., 2020) or as a financial risk management calculator in general (Fritz-Morgenthal et al., 2022).

Leaning into the business sector, in Small-and-Medium-sized Enterprises (SMEs), key application areas of XAI include data analysis and insights, inventory management, financial analysis, personalized recommendations, fraud detection, process automation, predictive maintenance, customer support, and recruitment selection process (Han et al., 2023). For example, the XAI can be used for customer segmentation in the product development stage. Both feature-based and data-based explanations are valuable to the customer segmentation process in product development. It is a very iterative process, so the XAI involvement process would start from the model development selection, hyperparameter tuning, and model evaluation to troubleshooting. Feature-based explanations can facilitate these. Since this is a data-driven design, the customer's heterogeneous data must be homogenized. The system would go through the training stage and testing stage as a normal AI would, but this time, every part of the process would be explainable. The AI prediction on customer segmentation is essentially concerning humans. In light of human-centred design and fairness, it is important to understand explicitly why a prediction is made on a particular customer instance through local feature-based explanations (Hu et al., 2023).

XAI in Cyber Security

Hoax is fake information often found on the internet. With the current speed of the internet, hoaxes can spread fast like wildfire. Hoax news is often deceitful and seems true at first glance, leading readers to believe it as it is. A hoax can be weaponized to shape public perception of certain topics by tricking them into believing what is unreal. Research has proven that people often have trouble distinguishing genuine information from hoaxes; most people won't go the extra length to verify the legitimacy of information or news and its source. To counter this growing problem, an XAI system was developed using the Decision Tree algorithm. The system detects hoaxes and notifies readers on a web browser. The XAI model allows experts and non-experts to know why the system concludes something is a hoax or not and get a good grasp on the flow, allowing users to trust both the news and the system that verifies it (Imanuel et al., 2022).

Cyber risk management is defined as "any risk emerging from intentional attacks on information and communication technology (ICT) systems that compromises the confidentiality, availability, or the integrity of data or services". Cyber security is urgently important now more than ever, given what's at stake for most people online today. Cyber risk management has previously been done using only AI, which is effective, however the system doesn't allow just anyone to understand. When something is as profound as this, understandability is key, which is why XAI methods are better suited and has lately been developed for it (Giudici & Raffinetti, 2021).

XAI in Human Life Advancement

Smart city has been a dream of mankind for a while now. It is defined as technology-enabled, socially intensive, and environmentally friendly urban areas. Making the dream of a smart city into a reality would require many kinds of technology, such as blockchain, IoT, big data, 5G and beyond technologies, digital twins, AR/VR, and computer vision. In many aspects of life in a smart city, XAI can be the solution for efficiency and effortlessness in day-to-day life, such as for energy management, temperature management, education, health and human services, water management, air quality management, traffic management, payments and finance, smart parking, and trash management. What really sets apart an XAI-powered smart city to regular cities we're used to would be its ability to use energy and resources more efficiently, protect the environment, and improve its citizens' lives. When looked in depth, this can really be one of the key technologies in leading the development of up-and-coming smart cities (Javed et al., 2023).

A big part of the appeal of AI is its decision-making ability that's deemed objective and accurate. In various fields, XAI has been employed to aid decision-making, such as in healthcare, manufacturing, banking, education, insurance, autonomous driving, etc. Human understanding is usually achieved through graphical or textual explanations, which is why any system created to assist decision-making should be able to provide such explanations. Even though AI-assisted decision making is convenient, it must also be transparent and responsible, especially when task with something high risk such as automated driving and healthcare (Nagahisarchoghaei et al., 2023). In bioinformatics for example, since the model is data-driven it is crucial for it to be interpretable on every possible level by every party involved in the process to ensure safety and trustworthiness. A lack of explanation can stop collaboration and further development due to comprehensibility issues, which can be solved by XAI (Karim et al., 2023).

Soil fertility refers to the ability of soil to be planted, which means the soil has characteristics that support plants' growth, such as certain sets of nutrients and minerals as well as a certain level of moisture and density. Multiple physical, chemical, and biological parameters play a part in determining whether a soil is fertile or not. A type of XAI model developed for this purpose has successfully used measurement data of each parameter to determine the soil fertility with the accuracy of 97.02%, allowing the system to efficiently and effectively help people decide how to make the best use of land (Giudici & Raffinetti, 2021).

3.0 CONCLUSION

Given its many potentials and capabilities, XAI can change the world for the better in many ways. Thus far, XAI has deepened our knowledge and contributed to the fields of healthcare, finance, and cyber security. Its reach has surpassed many expectations of the technology, and this is only just the beginning (Nagahisarchoghaei et al., 2023). As technology evolves, surely our need for AI will only grow. That being said, without XAI, the contribution of AI to society would not have gone any further. The need for responsible AI has risen and will only keep rising at this rate. Transparency and trustworthiness become crucial, and XAI is just the solution to that (Fritz-Morgenthal et al., 2022). With the help of XAI, it is hoped that AI can increase its contribution to society, especially in profound fields such as healthcare and heavily regulated fields such as finance (Galić et al., 2023). Given the rising demand, XAI will continue to be developed and improved so that the trade-off can be minimized and its implementation in various fields can be increased (Bussmann et al., 2020).

Acknowledgement

The authors are grateful to the Research Institute and Community Service Institut Bisnis dan Teknologi Pelita Indonesia, which has facilitated the success of this research.

References

Alkhalaf, S., Alturise, F., Bahaddad, A. A., Elnaim, B. M. E., Shabana, S., Abdel-Khalek, S., & Mansour, R. F. (2023). Adaptive Aquila Optimizer with Explainable Artificial Intelligence-Enabled Cancer Diagnosis on Medical Imaging. *Cancers*, 15(5), 1492. https://doi.org/10.3390/cancers15051492

Amoroso, N., Quarto, S., Rocca, M. L., Tangaro, S., Monaco, A., & Bellotti, R. (2023). An eXplainability Artificial Intelligence approach to brain connectivity in Alzheimer's disease. *Frontiers in Aging Neuroscience*, 15. https://doi.org/10.3389/fnagi.2023.1238065

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in Fintech Risk Management. *Frontiers in Artificial Intelligence*, 3(26). https://doi.org/10.3389/frai.2020.00026

Brusa, E., Cibrario, L., Delprete, C., & Di Maggio, L. G. (2023). Explainable AI for Machine Fault Diagnosis: Understanding Features' Contribution in Machine Learning Models for Industrial Condition Monitoring. *Applied Sciences*, 13(4), 2038. https://doi.org/10.3390/app13042038

Chandra, H., Pawar, P. M., R. Elakkiya, Tamizharasan, P. S., Raja Muthalagu, & Alavikunhu Panthakkan. (2023). Explainable AI for Soil Fertility Prediction. *IEEE Access*, 11, 97866–97878. https://doi.org/10.1109/access.2023.3311827

de Lange, P. E., Melsom, B., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for Credit Assessment in Banks. *Journal of Risk and Financial Management*, 15(12), 556. https://doi.org/10.3390/jrfm15120556

Fritz-Morgenthal, S., Hein, B., & Papenbrock, J. (2022). Financial Risk Management and Explainable, Trustworthy, Responsible AI. *Frontiers in Artificial Intelligence*, 5(1), 1-14. https://doi.org/10.3389/frai.2022.779799

Galić, I., Marija Habijan, Hrvoje Leventić, & Krešimir Romić. (2023). Machine Learning Empowering Personalized Medicine: A Comprehensive Review of Medical Image Analysis Methods. *Electronics*, 12(21), 4411. https://doi.org/10.3390/electronics12214411

Giudici, P., & Raffinetti, E. (2021). Explainable AI methods in cyber risk management. *Quality and Reliability Engineering International*, 38(3), 1318-1326. https://doi.org/10.1002/qre.2939

Gurmessa, D. K., & Jimma, W. (2023). A comprehensive evaluation of explainable Artificial Intelligence techniques in stroke diagnosis: A systematic review. Cogent Engineering, 10(2), 1-20. https://doi.org/10.1080/23311916.2023.2273088

Han, T. A., Pandit, D., S Joneidy, Hasan, M. M., Hossain, J., M Hoque Tania, Hossain, M. A., & N Nourmohammadi. (2023). An Explainable AI Tool for Operational Risks Evaluation of AI Systems for SMEs. *In 2023 15th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 69-74. https://doi.org/10.1109/skima59232.2023.10387301

Hashem, H. A., Abdulazeem, Y., Labib, L. M., Elhosseini, M. A., & Shehata, M. (2023). An Integrated Machine Learning-Based Brain Computer Interface to Classify Diverse Limb Motor Tasks: Explainable Model. *Sensors*, 23(6), 3171. https://doi.org/10.3390/s23063171

Hu, X., Liu, A., Li, X., Dai, Y., & Nakao, M. (2023). Explainable AI for customer segmentation in product development. *CIRP Annals*, 72(1), 89-92. https://doi.org/10.1016/j.cirp.2023.03.004

Imanuel, J., Kintanswari, L., Vincent, Lucky, H., & Chowanda, A. (2022). Explainable Artificial Intelligence (XAI) on Hoax Detection Using Decision Tree C4.5 Method for Indonesian News Platform. *2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, 63–68. https://doi.org/10.1109/icsintesa56431.2022.10041567

Javed, A. R., Ahmed, W., Pandya, S., Maddikunta, P. K. R., Alazab, M., & Gadekallu, T. R. (2023). A Survey of Explainable Artificial Intelligence for Smart Cities. *Electronics*, 12(4), 1020. https://doi.org/10.3390/electronics12041020

Joyce, D. W., Kormilitzin, A., Smith, K. A., & Cipriani, A. (2023). Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *Npj Digital Medicine*, *6*(1), 1-8. https://doi.org/10.1038/s41746-023-00751-9

Karim, R., Islam, T., Shajalal, Oya Beyan, Lange, C., Cochez, M., Dietrich Rebholz-Schuhmann, & Decker, S. (2023). Explainable AI for Bioinformatics: Methods, Tools and Applications. *Briefings in Bioinformatics*, 24(5), 1-22. https://doi.org/10.1093/bib/bbad236

Kiani, M. (2022). Explainable artificial intelligence for functional brain development analysis: methods and applications. *University of Essex (United Kingdom)*. http://repository.essex.ac.uk/33114/%0AExplainable

Nagahisarchoghaei, M., Nur, N., Cummins, L., Nur, N., Karimi, M. M., Nandanwar, S., Bhattacharyya, S., & Rahimi, S. (2023). An Empirical Survey on Explainable AI Technologies: Recent Trends, Use-Cases, and Categories from Technical and Application Perspectives. *Electronics (Basel)*, 12(5), 1-41. https://doi.org/10.3390/electronics12051092.

Pillay, J., Rahman, S., Klarenbach, S., Reynolds, D. L., Tessier, L. A., Guylène Thériault, Persaud, N., Finley, C., Leighl, N., Matthew, Garritty, C., Traversy, G., Tan, M., & Hartling, L. (2024). Screening for lung cancer with computed tomography: protocol for systematic reviews for the Canadian Task Force on Preventive Health Care. *Systematic Reviews*, 13(1), 1-18. https://doi.org/10.1186/s13643-024-02506-3

Qian, J., Li, H., Wang, J., & He, L. (2023). Recent Advances in Explainable Artificial Intelligence for Magnetic Resonance Imaging. *Diagnostics,* 13(9), 1571. https://doi.org/10.3390/diagnostics13091571

Siddiqui, K., & Doyle, T. E. (2022). Trust Metrics for Medical Deep Learning Using Explainable-AI Ensemble for Time Series Classification. In 2022 *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 370-377. https://doi.org/10.1109/ccece49351.2022.9918458

Taşcı, B. (2023). Attention Deep Feature Extraction from Brain MRIs in Explainable Mode: DGXAINet. *Diagnostics*, *13*(5), 859. https://doi.org/10.3390/diagnostics13050859

Tian, Z., Cheng, Y., Zhao, S., Li, R., Zhou, J., Sun, Q., & Wang, D. (2024). Deep learning radiomics-based prediction model of metachronous distant metastasis following curative resection for retroperitoneal leiomyosarcoma: a bicentric study. *Cancer Imaging*, 24(1), 1-13. https://doi.org/10.1186/s40644-024-00697-5