Tiako P.F. (ed) *Intelligent and Sustainable Solutions*. Chronicle of Computing. OkIP. CAIF25#11
© 2025 Oklahoma International Publishing https://doi.org/ 10.55432/978-1-6692-0011-6_8

# Comparison between both methods of explanations: LIME and SHAP

Tarik Lahna[a], Bernard Kamsu Foguem[b]
[a]INPT, Toulouse, France
[b]UTTOP, Tarbes, France

*ABSTRACT*

Structural engineers can greatly benefit from understanding the reasons behind the behavior of machine learning algorithms in several key areas, such as feature engineering, model selection, confidence in predictions, taking action based on those predictions, and the development of more user-friendly interfaces. As a result, interpretability has become a central issue in deep learning, and research into interpretable models has gained significant attention from both industry and academia. Due to their transparency, these models are often preferred because they can achieve similar accuracy to non-interpretable models in certain applications. When interpretability is crucial, they may still be chosen even if their accuracy is slightly lower. However, limiting machine learning to models that are interpretable can often be a significant drawback. In this paper, we present a case study focused on predicting crack types using model-agnostic methods to explain deep learning predictions. These methods provide considerable flexibility in model selection, explanations, and representations by treating deep learning models as black-box functions

Keywords: LIME; SHAPE; STRUCTURAL ENGINEERING

## I. INTRODUCTION

Artificial Intelligence started its influence in the 1950s and it has been developed quickly in many fields starting from last decade. The amount of data increased exponentially in many areas with a set of improved algorithms and more powerful computer hardware. That is why, it becomes necessary to maintain the development of new technologies of AI for organizations (E. Brynjolfsson and A.N. McAfee 2017). Recently, artificial intelligence techniques such as machine learning and data science, are spreading in different fields (lahna et al., 2023). These techniques are also used in structural engineering to detect damages and cracks.

Cracking is generally detected based on visual examinations from an engineer and the main cracks in airport infrastructures are the ones present in airport pavements. However, computer-assisted detections for different types of cracking can be performed. This kind of assistance can improve maintenance services inside aerospace infrastructures to avoid different problems such as the safety of passengers during aircraft operations and the increased costs for maintenance departments. Plus, these infrastructures are vital economically for any region in the world (Peneda et al., 2011). Consequently, airport buildings need to modernize the management of maintenance by using artificial intelligence so that safety will be increased inside these kinds of buildings while costs allocated for maintenance inspections will be reduced. Predictive maintenance such as computer-assisted detections have an important role in airport infrastructures for the next decades. In this article, cracking that will be treated are cracking in structural elements and those in airport pavements. The study will focus on the comparison of two model-agnostic explanation approaches.

The purpose of the article is to define different two model-agnostic explanation approaches which are : LIME and SHAP. pre-trained networks. Also, these implementations allow to This will help to answer the following research questions:

What are the definitions of the main keywords of this article?

To answer all these research questions, the paper will focus on defining the main keywords of this article and developing the impact of the model-agnostic explanation approach based on different parameters.

First, the paper will draw partially on a literature review highlighting the explainable models described above. Second, a proposed method related to explainable models will be detailed.

Finally, limitations and a conclusion are presented.

## II. LITERATURE REVIEW

### 2.1. Artificial Neural Network

Artificial Intelligence (AI) first emerged in 1943 and has witnessed exponential growth, particularly since the early 2000s, a period now referred to as the "High-speed Development Period" (Zhang et al., 2019). This era saw significant advancements in both the theoretical foundations and practical applications of AI. At the Dartmouth Conference, as shown in Figure 1, AI was initially defined as the ability of a machine to mimic human intelligence, essentially allowing machines to exhibit behaviors that are similar to those of human cognitive processes, such as reasoning, learning, and problem-solving (Moor, 2006). However, Nils J. Nilsson offered a broader perspective, suggesting that AI is fundamentally a discipline focused on the representation and acquisition of knowledge, emphasizing the importance of how machines understand, store, and retrieve information (Kuipers et al., 2017). This approach stresses that AI is not just about replicating human behavior but about empowering machines to make informed decisions based on the knowledge they process

Presented at the 2025 OkIP International Conference
on Artificial Intelligence Frontiers (CAIF)
in Oklahoma City, OK, USA, and Online, on April 2, 2025

and learn over time. AI now encompasses a broad spectrum of technologies, including but not limited to computer vision, machine learning, virtual reality, and big data analytics (Zhuang et al., 2017). The growth in data across various fields, paired with the development of increasingly powerful computational hardware and sophisticated algorithms, has played a pivotal role in this technological explosion. As a result, AI has become indispensable in numerous sectors, necessitating continuous innovation to keep up with the increasing complexity and volume of data that organizations face (Brynjolfsson and McAfee, 2017). Recent advancements in AI techniques, particularly in machine learning and data science, have led to their widespread adoption across a variety of industries. In structural engineering, for example, these technologies are now employed to detect structural damages and cracks, enabling more accurate, efficient, and cost-effective methods of maintenance and safety monitoring (LeCun et al., 2015). The application of AI in this context allows engineers to analyze large datasets from sensors, images, and inspections, automatically identifying potential issues that would be difficult or time-consuming for humans to detect manually. This capability is transforming how we ensure the safety and longevity of critical infrastructure, making AI an invaluable tool in modern engineering practices.

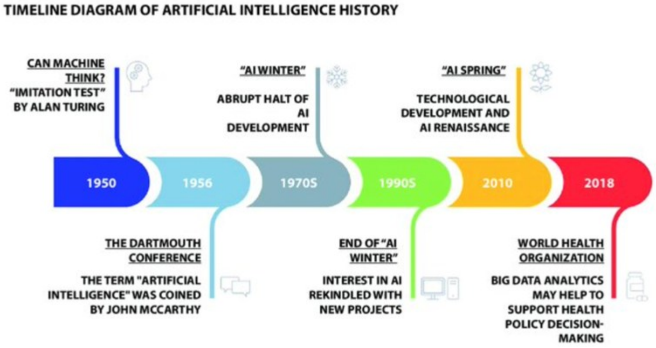

TIMELINE DIAGRAM OF ARTIFICIAL INTELLIGENCE HISTORY
CAN MACHINE THINK?
"IMITATION TEST" BY ALAN TURING
"AI WINTER"
ABRUPT HALT OF AI DEVELOPMENT
"AI SPRING"
TECHNOLOGICAL DEVELOPMENT AND AI RENAISSANCE
1950
1956
1970S
1990S
2010
2018
THE DARTMOUTH CONFERENCE
THE TERM "ARTIFICIAL INTELLIGENCE" WAS COINED BY JOHN MCCARTHY
END OF "AI WINTER"
INTEREST IN AI REKINDLED WITH NEW PROJECTS
WORLD HEALTH ORGANIZATION
BIG DATA ANALYTICS MAY HELP TO SUPPORT HEALTH POLICY DECISION-MAKING

**FIGURE 1:** Timeline diagram showing the evolution of AI from **(Bellini et al., 2022)**

### 2.2. Model-agnostic explanation approaches

#### A. LIME (Local Interpretable Model-Agnostic Explanation)

LIME is an interpretability method designed to explain the predictions of machine learning models, particularly complex ones like neural networks or ensemble models. It works by locally approximating a black-box model with a simpler, more interpretable model, such as linear regression or decision trees. By perturbing the input data and observing the changes in the model's output, LIME creates a local surrogate model that is easier for humans to understand. This method provides accessible explanations for specific predictions, which is crucial in fields where transparency in decision-making is essential, such as healthcare and finance (Ribeiro et al., 2016). One of the key strengths of LIME is that it is model-agnostic, meaning it can be applied to any machine learning model, regardless of its internal complexity.

Figure 2 describes several key principles of explainable methods:

**Transparency**: Ensuring that stakeholders can clearly understand the decision-making process of the models.

**Fairness**: Making sure the model's decisions are equitable for all individuals, including those from protected groups (such as race, religion, gender, disability, or ethnicity).

**Trust**: Evaluating how much confidence human users can place in the AI system's outputs and predictions.

**Robustness**: Ensuring the model is resilient to variations in input data or changes in parameters, maintaining reliable performance even when faced with uncertainty or unforeseen circumstances.

**Privacy**: Safeguarding sensitive user information and ensuring its confidentiality.

**Interpretability**: Offering explanations that are easily understood by humans regarding how the model arrives at its predictions and conclusions.

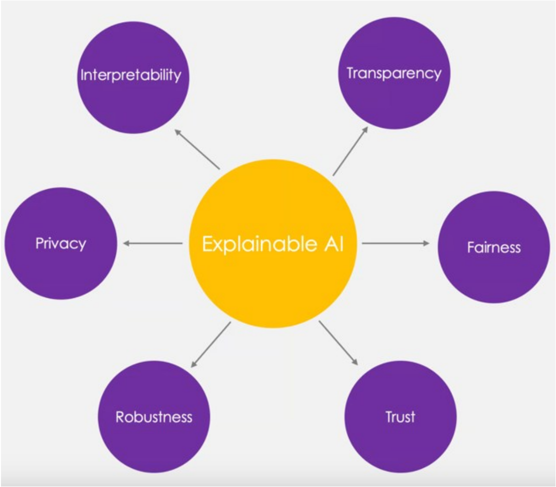


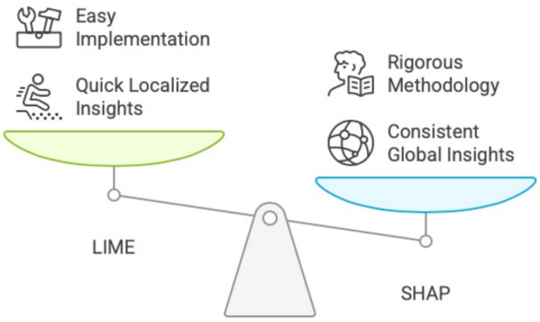
**FIGURE 2:** The definition of Explainable methods

#### B. SHAP (Shapley Additive Explanations)

SHAP is another popular method for explaining machine learning model predictions, based on Shapley values from cooperative game theory. SHAP values measure the contribution of each feature to a specific prediction, using the concept of fairly distributing a reward among players in a game according to their contributions. In machine learning, each Shapley value quantifies the importance of a feature in the model's outcome. This method ensures a consistent and fair attribution of importance to each feature, offering both local interpretability (explaining specific predictions) and global interpretability (providing insights into the overall behavior of the model). SHAP is particularly effective in explaining complex models like neural networks and tree-based methods (Lundberg & Lee, 2017).

## III. PROPOSED SYSTEM

While LIME and SHAP are both powerful explainability methods, they serve different purposes and can be complementary when used together. LIME is typically preferred when the goal is to gain quick, local explanations for individual predictions, making it useful in situations where users need to interpret specific outcomes. On the other hand, SHAP is ideal for a more consistent and fair attribution of feature importance across the entire dataset, providing a broader view of how features influence the model's predictions both locally and globally. Combining both methods allows for a comprehensive understanding of a model's behavior, with LIME providing fast, localized insights and SHAP offering more robust, mathematically grounded explanations (Ribeiro et al., 2016; Lundberg & Lee, 2017). This dual approach ensures that both local predictions and overall feature contributions are well-understood, which is especially valuable for model validation and building trust in AI systems. Choosing between LIME and SHAP largely depends on the specific needs of the user. LIME is well-suited for situations where fast and flexible local explanations are required, especially when computational efficiency is a priority. In contrast, SHAP is the better choice for cases where global interpretability and the fair distribution of feature importance are more critical. For example, SHAP is particularly useful in settings where the model needs to be explainable in a rigorous, mathematically consistent way, such as in high-stakes industries (Lundberg & Lee, 2017). Both methods can be used independently or together, depending on whether the goal is to prioritize quick, understandable explanations or a deeper, more thorough understanding of the model's behavior.



Choose LIME for speed; SHAP for depth.

**FIGURE 3:** A proposed method to differenciate between explainable methods

## IV. CONCLUSION

The best solution for explainable AI depends on your specific needs. LIME is ideal for fast, local explanations, while SHAP provides a more thorough, consistent approach for global and local interpretability. Using both methods together allows you to get a comprehensive picture of how the model makes decisions and the contributions of individual features. Whether you choose LIME, SHAP, or both depends on whether you need quick explanations or a deeper, more reliable understanding of the model's inner workings.

## REFERENCES

- Attaccalite, L., Di Mascio, P., Loprencipe, G., & Pandolfi, C. (2012). Risk Assessment Around Airport. *Procedia - Social and Behavioral Sciences*, *53*, 851–860.
- Baltaci, N., İpek, Ö., & Akbulut Yıldız, G. (2015). *The Relationship between Air Transport and Economic Growth in Turkey: Cross-Regional Panel Data Analysis Approach*. *7*, 89–100.
- Bellini, V., Cascella, M., Cutugno, F., Russo, M., Lanza, R., Compagnone, C., & Bignami, E. G. (2022). Understanding basic principles of Artificial Intelligence: a practical guide for intensivists. *Acta Bio-Medica : Atenei Parmensis*, *93*(5), e2022297.
- E. Brynjolfsson and A.N. McAfee. (2017). What's driving the Machine Learning explosion? *Harvard Business Review*, *18*.
- Fábio Celestino Pereira and Carlos Eduardo Pereira. Embedded image processing systems for automatic recognition of cracks using uavs. IFAC-PapersOnLine, 48(10):16–21, 2015.
- Kuipers, B., Feigenbaum, E. A., Hart, P. E., & Nilsson, N. J. (2017). Shakey: From Conception to History. *AI Magazine*, *38*(1), 88–103.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
- Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, *27*(4), 87.
- Peneda, M. J. A., Reis, V. D., & Macário, M. do R. M. R. (2011). Critical Factors for Development of Airport Cities. *Transportation Research Record*, *2214*(1), 1–9.
- Tarik Lahna, Bernard Kamsu-Foguem, Henry Fonbeyin Abanda,Maintenance in airport infrastructure: A bibliometric analysis and future research directions, Journal of Building Engineering, Volume 76, 2023, 106876
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.
- Zhang, X., Ming, X., Liu, Z., Yin, D., Chen, Z., & Chang, Y. (2019). A reference framework and overall planning of industrial artificial intelligence (I-AI) for new application scenarios. *The International Journal of Advanced Manufacturing Technology*, *101*(9), 2367–2389.
- Zhuang, Y., Wu, F., Chen, C., & Pan, Y. (2017). Challenges and opportunities: from big data to knowledge in AI 2.0. *Frontiers of Information Technology & Electronic Engineering*, *18*, 3–14.