A Tool Combining Explainability and Benefits of Artificial Intelligence to Improve Trust in AI in Agriculture

TWAN BERG, University of Twente, The Netherlands

Artificial Intelligence is playing a growing role in the agricultural sector. The agricultural sector is facing various challenges impacting efficiency, such as diseases and pesticides, soil health management, and drainage and irrigation. The application of AI is considered part of a viable solution to the growing world population and the increased food consumption, by increasing the efficiency and yield in this sector. However, the adoption of AI in the agricultural sector is slow. Many AI systems are called black boxes, meaning that there is no way to explain the reasons behind predictions. Explainability of AI is important for building trust and understanding leading to an increased adoption rate. This research helps to bridge the gap between the potential benefits and the low comprehension and trust of an AI system in agriculture. A tool is developed combining the explanation and benefits of AI in agriculture tailored to the needs of farmers to ensure quicker adoption and more trust in AI systems. The effectiveness of the tool was evaluated through a focus group, demonstrating that the tool significantly increased the trust and understanding of users in the AI system, while also highlighting the need for a balance between simplicity and completeness of explanations.

Additional Key Words and Phrases: Artifical Intelligence, Machine Learning, Agriculture, Design Science

1 INTRODUCTION

In today's world AI is growing rapidly. Many new AI technologies emerge looking to improve efficiency, reduce costs, and help with innovation [3, 8]. The benefits of adopting AI in agriculture can be massive [26]. Alexandratos et al. [2] projected the global population to be approaching 10 billion by 2050. This will put upward pressure on the food demand of up to 70%. Despite the availability of AI tools supporting the agricultural sector, the adoption of technology remains low [5]. AI Adoption is the process of individuals, organizations, and industries integrating AI tools in their workflows, operations, and decision-making processes. Several factors can lay the groundwork for slow adoption, such as limited AI and technical skills and knowledge, lack of trust in AI systems, and financial reservations [8, 14, 29, 30].

An upcoming research topic and potential solution to the aforementioned problems with the adoption of AI in agriculture is Explainable AI [1, 35]. Explainable AI aims to turn modern AI systems into an understandable and transparent program by providing explanations of the internal processes and outputs. As described by Xu et al. [39] and Guidotti et al. [12], modern AI systems tend to operate as black boxes. Some even say these modern systems are a form of alchemy instead of a knowledge-based science. The field of Explainable AI is the research and development of tools and frameworks to create the opportunity to explain AI systems without significantly reducing the efficiency or accuracy of the AI model. This proves to be a challenge, Xu et al. [39] found a relation between the prediction accuracy and the explainability of an AI model which shows that as models become more complex they generally turn out to be more precise. This field can help propel the adoption of AI by increasing understanding and trust in AI systems through explanations [35].

The existing research in the field of AI in agriculture [20] and Explainable AI [23, 32] is lacking in the area of actual implementation and use of these systems in processes in the agricultural sector. This research aims to explore and demonstrate the potential of Explainable AI in the field of agriculture. Focusing on a weed detection and management tool as a case study [20, 27]. By the end of this research, we will have analyzed this AI tool on weed detection and management to be able to interpret the inputs, outputs, and internal processes of the tool as well as the possibilities and drawbacks of using this system. By combining this knowledge and research into the fields of AI in agriculture and Explainable AI in an interactive tool to explain this AI tool to farmers, we expect to contribute to the field of XAI by providing an accessible way for farmers to facilitate the adoption of this AI tool into their workflows by increasing the trust and understanding of this AI tool.

To achieve the goal of this research we will work with the following research question to lead the research:

In what ways can we help increase the awareness and understanding of an AI system for farmers, to increase the adoption of this AI system into the workflows of the agricultural industry, through an interactive explanatory tool?

Which we can answer through the following sub-questions:

(1) What are the important benefits, drawbacks, internal processes, and relations between input and output of the AI system on weed detection and management?
(2) What are the requirements for the explanatory tool's design and functionality considering the needs and challenges of farmers?
(3) To what extent do the explanations and visualizations provided in the explanatory tool increase the understanding and trust of farmers in the AI system?

This research analyzed the functionality, inner workings, benefits, and drawbacks of the specific AI tool on weed detection [27]. Based on the insights gained in this analysis an interactive explanatory tool was developed to explain the concepts of the AI tool to the target users, farmers. This tool provides interactive explanations and visualizations to improve the understanding and trust in the AI system. The tool's effectiveness was evaluated through a focus group. The results of this focus group demonstrate an increase in user understanding and trust in the AI system after following the explanation provided in the interactive explanatory tool. These results indicate that this and similar tools can contribute to the

TScIT 41, July 5, 2024, Enschede, The Netherlands

adoption of AI in agriculture by increasing trust and understanding of AI systems, leading to improved efficiency, sustainability, and innovation in this sector.

2 BACKGROUND

2.1 Artificial Intelligence in Agriculture

In terms of Artificial Intelligence and Machine Learning in the agricultural sector, there are many technologies and research on these technologies that aid in optimizing various aspects of agriculture, such as crop management, livestock monitoring, water, and soil management. For example, these AI systems can leverage machine learning algorithms to analyze satellite imagery or sensor data for precision agriculture [4, 9, 13, 18, 20, 26]. Reviews of the literature describe the models and algorithms used in these technologies, while shedding light on the benefits, such as reducing costs and pollution as well as improving efficiency, and challenges, including a lack of trust, financial obstacles, and the complexity of systems.

Despite the numerous existing AI solutions, the adoption of AI in agriculture remains low [5]. Showing a significant gap between new technologies and the practical implementation. One of the primary challenges is the lack of trust and understanding in AI systems. A proposed solution to this challenge is Explainable AI, aiming to increase the transparency of AI systems by providing interpretable explanations of AI models [1, 23, 32]. The combination of Explainable AI and existing AI technologies has been explored [21, 31, 36]. While the potential possibilities of this combination are showcased through practical applications, these technologies still provide technical and complex explanations, using methods tailored to the specific AI tools. This specificity makes it hard to generalize the concept to other AI tools, limiting their practical usability in real-world situations.

2.2 Explainable AI

The primary aim of explainable AI (XAI) is to enhance the transparency and understandability of AI systems to improve the trust in these systems of non-technical users [1, 23, 32], especially in sectors like agriculture, where practical application and trust in these systems are crucial. In this research, the main goal is the design and development of an explanatory tool for the AI system on weed detection and management. To ensure the development of an effective tool, the field of Explainable AI discusses several important considerations to integrate with the tool.

A key consideration in XAI is the audience [23, 32]. The explanations need to be accessible and understandable, which can be accomplished through the simplification of technical terms as well as using visualizations and analogies to simplify complex concepts. This helps bridge the gap between non-technical users and the AI system.

The types and techniques of explanations significantly impact the effectiveness of the explanation. Both global and local explanations are essential [12, 23]. Local explanations focus on an individual prediction, ensuring interpretability of the reasoning behind a specific decision. Global explanations lead to an understanding of the entire model and reasoning for all possible outcomes. Interpretable models, including decision trees and linear models, are useful as they focus on the importance of input features and the rules used to make decisions [12]. Visualizations, like charts and heatmaps, used in the explanation can show what features have the most influence on the model's prediction, increasing the value of the explanation.

Visual explanations are considered a powerful way to explain complex processes and information [23, 32]. Incorporating intuitive visualizations representing the AI model's internal processes and workings as well as relations between the input objects and output classes can help users explore and understand these relations. Incorporating interactive elements, for example by allowing the user to adjust input variables and observe real-time outputs, further increases understanding.

The primary goal of XAI is building trust in AI systems. Therefore, the explanations need to be truthful and accurate, demonstrating and emphasizing the accuracy and reliability of the tool [12, 23]. Further trust can be built by eliminating doubts regarding factors like ethical considerations, fairness, and bias through explanations of how these risks are mitigated.

Finally, a crucial aspect of engagement with the tool and building an understanding is the contextual relevance of the explanations [23]. The explanations provided by the tool need to relate to practical situations in the agricultural sector, which increases the relevance of the explanation. The relevance and context make the benefits of adopting AI tools clearer.

However, the literature on Explainable AI identifies several challenges. There is a lack of consensus on the definitions of explainability and interpretability, leading to various standards and approaches. Existing techniques often have limited usability of existing explanation techniques in real-world situations due to negligence of the human side of explanations and interpretations. Further challenges include the scalability of the XAI concept to larger and more complex datasets, finding a balance in explanations between simplicity and completeness, and preventing potential bias or misinterpretation among users introduced by the explanations. It is essential to overcome these challenges to develop effective and usable XAI tools to improve the trust and adoption of AI in agriculture.

A clear research gap remains in effectively addressing the challenges of AI adoption in agriculture and Explainable AI by means of practical and usable solutions. Existing research lacks practical solutions that are simple, accessible, and trusted by non-technical users in the agricultural sector. This research aims to close this gap by focusing on the development of a solution in the form of an interactive explanatory tool that helps to increase trust, understanding, and therefore adoption of AI solutions. By providing a practical and accessible example of the application of Explainable AI in agriculture, this study aims to facilitate the adoption of AI in this sector.

3 CASE STUDY

In this research, we focus on a study by Pantazi et al. [27], which leverages Unmanned Aircraft Systems equipped with multispectral cameras to demonstrate the detection and mapping of the weed Silybum Marianum among other weed species. S. Marianum is responsible for major crop loss, particularly in the production of cereal and leguminous crops, and is hard to eradicate. This study provides a technical yet clear and detailed explanation of the developed tool, an important consideration before building upon this technology. Additionally, results found with this tool were easily available through several examples demonstrating the functionality of the tool. On top of that, the visual aspect of the generated maps provides a basis for clear and understandable visualizations to enhance the explanation in our tool. These points indicate that this tool is a suitable candidate as a case study for our explanatory tool, providing a strong foundation to build the explanatory tool.

The mapping of S. Marianum patches was performed using hierarchical self-organizing map classifiers including Counter-Propagation Artificial Neural Networks [17, 22, 40] and variants such as XY-Fused Networks and Supervised Kohonen Networks [6, 22]. These three techniques proved to be extremely effective, reaching an accuracy of 97.6% - 98.9% [27]. The three hierarchical maps were trained based on input objects consisting of 4 features, three spectral and a textural feature. The high-resolution images used to train and evaluate the models were captured by a multispectral camera, capturing green, red, and near-infrared spectral data, attached to an Unmanned Aircraft System. These three spectral bands together with a texture layer formed the input to the classifiers. This section discusses the techniques used in these classifiers.

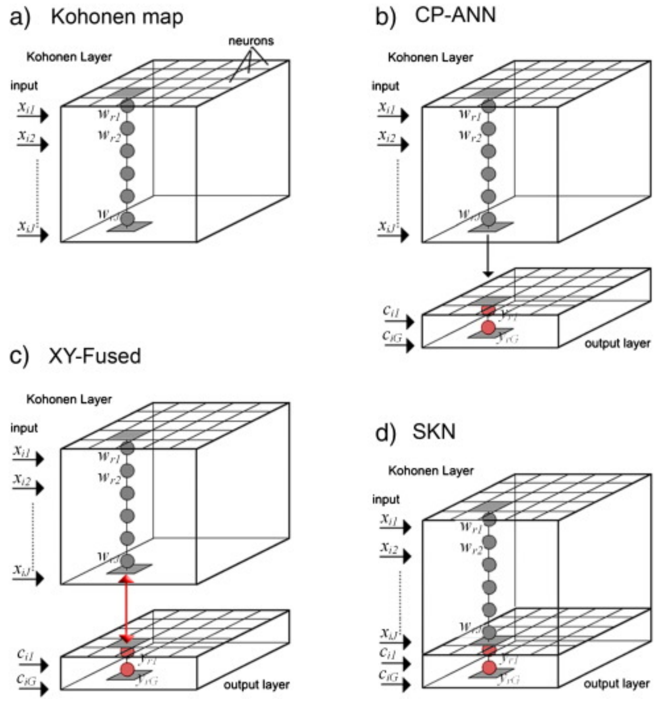

Fig. 1. Overview of the SOM (a), CP-ANN (b), XY-F (c), and SKN (d) consisting of J input variables and G output classes. x_{ij} is the j-th variable for the i-th sample. w_{rj} is the weight of the j-th variable for the r-th neuron. c_{ig} represents the class membership of the i-th sample to the g-th class. y_{rg} represents the weight of the g-th output for the r-th neuron. [6]

3.1 Hierarchical Self-Organizing Map Classifiers

3.1.1 Self-Organizing Maps. Self-Organizing Maps (SOM) are one of the most important Artificial Neural Network (ANN) architectures used for classification, mapping, and dimensionality reduction. Originally introduced by Kohonen [15], a SOM is an unsupervised machine learning technique to form a low-dimensional representation of a high-dimensional space, preserving the topological structure. Typically, a two-dimensional grid, either rectangular or hexagonal, is used for the low-dimensional representation (Fig. 1(a)).

The SOM is trained through competitive learning, where neurons compete to the right to respond to a subset of the input data. The low-dimensional representation consists of neurons equipped with a weight vector, representing the position in the input space. Training consists of several steps [15, 16, 22]. Each object in the training data is presented to the network. Subsequently, the weight vectors of the closest matching neurons are updated to closer represent the training data object, moderated by the neighborhood function and learning rate function. These functions gradually decrease, ensuring neurons quickly adapt to the input objects initially and later specialize to certain input objects. This defines the competitive learning aspect of this Artificial Neural Network.

3.1.2 Counter-Propagation Artificial Neural Networks. Counter Propagation Artificial Neural Networks (CP-ANN) extend SOMs by adding an additional output layer for an extended mapping mechanism [17, 22, 27, 40]. While a SOM consists of a single layer of neurons, a CP-ANN includes an input layer, the SOM layer, and an output layer that maps the clusters of the SOM to target classes (Fig. 1(b)). The training process is similar to that of a SOM but includes simultaneous training of the output layer. This results in a mapping between neurons in the SOM layer and neurons in the output layer, where the position of neurons in the SOM layer is used as a lookup for the corresponding output neuron, representing the predicted target class.

3.1.3 Supervised Kohonen Networks. Supervised Kohonen Networks (SKN) integrate supervised learning with SOMs [22, 27]. This addition allows for direct association of the clusters formed by the SOM layer and the target output classes, enhancing the classification capabilities of the network. Similar to CP-ANNs, SKNs have an input SOM layer and an output layer. However, in an SKN these maps are joined together to form a single input-output map, ensuring that the input object and corresponding output object are a combined input to the network in training (Fig. 1(d)). After training, the input-output map is decoupled, allowing a similar approach as CP-ANN for predictions.

3.1.4 XY-Fused Networks. The final technique used to train a model for the classification of S. marianum was the XY-Fused Network. XY-Fused Networks are another variation of CP-ANNs [22, 27]. Unlike previous techniques, where the BMU is determined solely by the SOM layer, XY-Fused Networks use a fused similarity measure based on a weighted combination of similarities between the input object and the input layer, and the corresponding output object and the output layer. The common winning unit is the neuron that gets updated in both maps is determined by this fused similarity measure (Fig. 1(c)). The rest of the procedures follow standard SOM processes.

4 METHODOLOGY

The main focus of this research is to design and prototype an explanatory tool for the AI systems on weed detection and management, to help explain the AI tool to potential new users in the agricultural sector. The research adopts the Design Science Research methodology as depicted in Figure 2 [28]. This methodology was developed to help understand and develop knowledge of how things can be constructed, arranged, or designed to achieve a set goal. This methodology guides the process of identification and motivation of the problem, defining the requirements of the proposed solution, designing and prototyping, and evaluating the solution.

The identification and motivation of the problem are outlined in previous sections. Next, the objectives of a solution are derived from the theory, through an in-depth analysis of the AI systems on weed detection and management. This involves the benefits, drawbacks, internal processes, and the influence of the input on the output of the system. This analysis is based on the literature describing or analyzing the AI system on weed detection [20, 27]. Using the knowledge gained from this review, I can develop a qualitative explanation of the AI tool. Furthermore, an understanding of UI design and the concept of Explainable AI through a literature review ensures that the explanatory tool accurately and effectively conveys the explanation of the AI system. Finally, the domain of this research is the agricultural sector. That means that all previous research will be done from the standpoint of this domain and with the necessary relevance to this domain to gain insights into domain-specific requirements.

Requirements relevant to the design and development of the tool are formulated based on the previous analysis. By means of these requirements, the design and prototype of the actual tool can be realized. The design should not only follow the requirements but should also ensure adherence to UI design principles. The resulting prototype will integrate the previous work into a product completing the research.

The theory of Explainable AI, the case study of the AI tool on weed detection and management, and the extracted requirements formed the basis for the design phase, with the goal to make a structured and intuitive tool, that effectively explains the concepts behind the AI tool to farmers. The first step in the design phase was to identify key features and concepts to include in the tool from the literature review and the analysis of our AI tool. Based on these key features a design for the explanatory tool was developed. With the supervisor of this project and peer students the design was evaluated, this feedback gave valuable insights on how to improve the design. This revised design formed the basis for the development of the tool.

The final step in the Design Science Research methodology that will be performed is an evaluation of the tool that was developed. To be able to draw conclusions about the effectiveness of the prototype, the system needs to be rigorously evaluated on the utility, quality, and efficacy of the system [28]. Venable et al. [37] proposed a framework with several approaches to evaluate the developed artifact. A summative approach to the evaluation using a Quick and Simple evaluation strategy is applied [37] in the form of a focus group [25].

The focus group was done in two phases: a test run and the main evaluation. The test run involved two participants while the main evaluation phase involved six participants, all representative of the target user group in the sense that they had limited technical knowledge and no previous experience with AI systems. After a brief introduction, all participants were asked to work their way through the tool independently, asking for help when needed. Then there was a group discussion about several aspects important to the goal of the evaluation phase, helping to gather insights into the experience with the tool. These discussions were recorded and transcribed for easier analysis of discussed points. The goal of the test run was to get feedback useful for identifying major issues and improving the tool, areas of focus included the usability of the tool, clarity of explanations, and the user experience. The main evaluation phase aimed to gather comprehensive and rigorous feedback on the effectiveness, usability, and educational benefits of the tool.

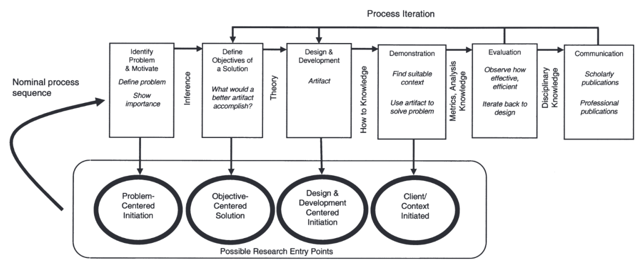

Fig. 2. DSRM Process Model. [28]

5 REQUIREMENTS

To ensure an effective design and development of the explanatory tool, several requirements were considered. The requirements are based on the principles of Explainable AI and tailored to the needs of the target users, farmers. In this section, the essential requirements for our tool will be discussed.

5.1 Functional Requirements

The functional requirements (FR), as listed in Table 1, describe the specific behaviors and features the tool should have. These focus on the types, methods, and structure of the explanations provided by the tool. The requirements are based on the theory and previous research on effective methods of explaining complex concepts. Functional requirements are the core features of the tool, essential for providing effective explanations. These requirements are directly related to the objectives of the AI tool on weed detection and mapping.

5.2 Non-Functional Requirements

Non-functional requirements (NFR), as listed in Table 2, specify the general structure and outline of the tool, without going in-depth on specific features or functions. These requirements are used to evaluate the effectiveness of the tool as a whole, rather than its individual elements. They include aspects such as performance, usability, and efficiency, ensuring the tool is effective and usable in real-world applications.

Table 1. Functional Requirements for the explanatory tool

	Requirement	Description	Related Works
FR1	System Metrics	The tool should include system metrics, like accuracy and performance scores. The accuracy and reliability of the AI tool are important aspects to build trust in the AI system	[10, 19]
FR2	Causality	The causal relation between input features and output predictions should be explained. This helps users understand why predictions are made, increasing trust and the ability to make informed decisions	[1, 11, 12, 24, 32]
FR3	Visualizations	Visual aids in the tool make complex concepts easier to follow, providing a better understanding and interpretation. Making the explanations more effective	[1, 19]
FR4	Interactive elements	Interactive elements, like sliders and inputs to control variables and observe the effects, and quizzes, help users engage with the tool and provides hands-on experience, enhancing the learning process	[1, 7]
FR5	Examples	The tool should include real-world examples to illustrate concepts. Examples help to place the information in context, making abstract concepts more concrete	[33]
FR6	Abstractions	Using abstraction of complex concepts by excluding complicated details turns the explanation into a more manageable form, increasing the understanding of the explanations	[32, 33]
FR7	Diagrams and Graphs	Including diagrams and graphs to explain data helps to make this data more interpretable, increasing the understandability of the explanations	[11]
FR8	Progressive Disclosure	More detailed information should be revealed progressively. By presenting information in manageable parts we improve the focus and attention by managing the amount of information provided at once	[11]
FR9	Two-way Communication	Interactive communication between the tool and the user should be available, enabling users to engage actively with the tool. The increased engagement facilitates the learning process	[7, 19]
FR10	Context-specific	Context-specific explanations should be provided from within the context of specific recommendations and reasoning, leading to increased learning	[11]

Table 2. Non-Functional Requirements for the explanatory tool

	Requirement	Description	Related Works
NFR1	Target audience	The tool should be designed with the target audience, farmers, in mind. These users often have limited technical knowledge, so the tool should be accessible and intuitive. The ease of use of the system will help adoption	[12, 23, 32]
NFR2	Step-by-step process	The tool should guide the user through the explanation process step-by-step. This ensures that users do not feel overwhelmed and helps with building a gradual understanding making it easier to follow the explanation and allowing users to drop out at a point where the technical details become irrelevant	[19]
NFR3	Clear, concise, and accurate explanations	Explanations should be free of technical terms and easily understandable to non-technical users. This increases the ability to understand the process, allowing easier adoption	[12, 33]

6 THE EXPLANATORY TOOL

As described in the Methodology section the design and development of the tool was based on the theory of Explainable AI, the case study of the AI tool, and the extracted requirements, aiming to make an effective explanation of the concepts behind the AI tool accessible to farmers.

The realized tool uses a chapter-based approach, where each chapter introduces a different concept at varying levels of complexity. The tool is structured into four main chapters: Introduction, Input-Output Maps, Self-Organizing Maps, and Conclusion, to develop a step-by-step explanation from a high-level overview to a technical explanation. This approach, starting the explanation with an overview of the relationship between input features and the output map before discussing the more technical details of Self-Organizing Maps, allows a user to drop out of the explanation when technical details are considered irrelevant. Additionally, developing a high-level understanding before introducing complex details can help maintain user engagement [FR8, NFR2].

The introduction provides an overview of the purpose, features, and benefits of the AI tool. It starts with an example of the tool, which shows how the AI uses an image of a field as input and produces a map indicating the distribution of S. Marianum on that field. Then, the introduction describes the impact and problems of weeds in agriculture, the need for effective weed management, and how the AI tool can help with this, and it introduces how the AI tool works. The introduction concludes with a quiz testing whether the user has understood the basic concepts [FR3,4,5,9,10].

The input and output maps are explained in the next chapter, it starts with an explanation of AI algorithms in general and how that is related to the weed detection tool. Then, it describes how the tool uses the four input features (Green, Red, Near-infrared, and Texture) extracted from an input image to generate a map showing patches of S. Marianum. Through an interactive and simplified example, the user can learn how the input features influence the output predictions of the AI. As we can see in Figure 3 this part of the explanation indicates groups of neurons on the input and output maps with colored circles, while explaining on the right side of the screen what the relation between the input and output is in those areas. This example is a step-by-step explanation controlled by the user, gradually revealing more details, which improves their learning [FR2,3,5,8].

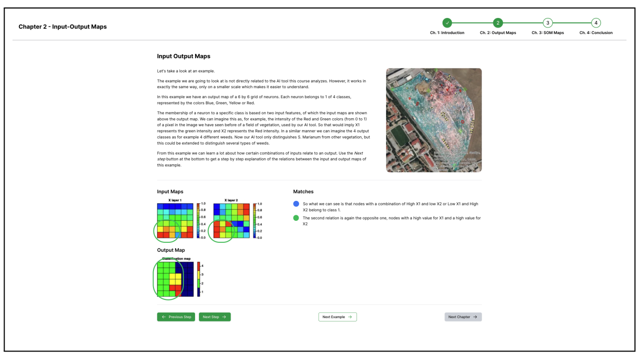

Fig. 3. Explanation of the relation between input features and output predictions

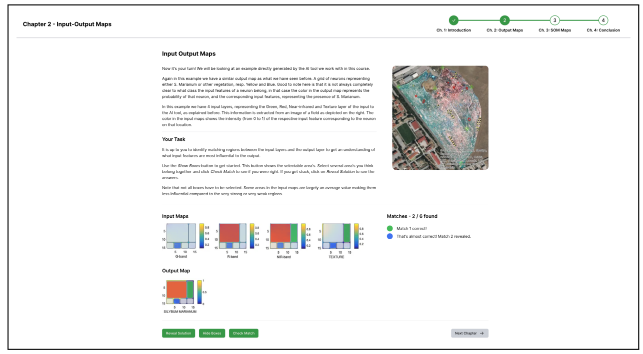

Fig. 4. Interactive exercise for identifying matching input and output regions

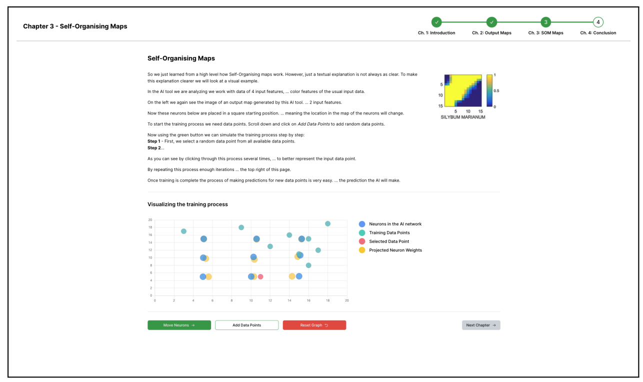

Fig. 5. Interactive demonstration of the training process of a SOM

Finally, in this chapter the user can do an interactive challenge, here the user is asked to find matching regions between input features and the output map, testing whether they understand the relationship between input and output. As seen in Figure 4 this interactive example shows an overlay of selectable boxes on the input and output maps, while asking the user to match the most influential input regions with the correct output region [FR2,3,4,5,9]. The input and output maps are maps created by the AI tool on weed detection [FR10].

Once the user understands these relationships, the tool moves on to the underlying algorithm mapping input features to output predictions: Self-Organizing Maps. This is one of the key concepts identified to be explained by the tool. Analyzing several existing explanations of Self-Organizing Maps [15, 16, 22, 38] provided insights into intuitive ways to explain this concept. This chapter introduces the training process of a SOM, it explains how data points with the four input features train a grid of neurons so it can be used for predictions. With a 2-dimensional graph, the training algorithm is visualized to the user, making this abstract concept easier to understand, as seen in Figure 5 [FR7]. Each step of the training algorithm, including selecting a data point and adjusting the neuron weights, is controlled by the user at their own pace. Some abstractions have been made to this explanation to reduce complexity, details such as the neighborhood function and learning rate function are excluded from the explanation to make the overall process clearer [FR3,4,6,8].

The last chapter of the tool is the conclusion, with a summary tying the key concepts, benefits, accuracy, and next steps together into a short overview and a final quiz the user can take to test their overall understanding of the material and the concepts learned. A closing remark concludes the explanation provided in the explanatory tool [FR1,5,9,10].

This approach, explaining the concepts from a high-level overview to a detailed technical explanation through several chapters, aims to make AI concepts understandable and accessible for farmers, increasing their trust and helping them to effectively adopt the AI tool for weed detection.

7 EVALUATION

7.1 Requirements

Looking at the structure, features, elements, and concepts included in the explanatory tool it can be seen that the functional requirements are satisfied by the tool. Identifying the non-functional requirements proves to be more difficult, the evaluation through a focus group will provide more insights into the achievement of these requirements. Additionally, evaluating an early iteration of the developed tool with the supervisor of this research and two peer students provided insights into needed improvements like the intermediate explanations and links between chapters, the relation of each chapter to the bigger picture of detecting and mapping S. Marianum and to add a quiz at the end testing the knowledge of the user.

7.2 Focus Group

An evaluation in the form of a focus group was used to get more insights into the effectiveness and usefulness of the tool. In the main evaluation, key aspects, such as the understanding of the concepts of the AI tool, the engagement and interest in the material, and the practical applicability of the tool in real-world scenarios, were discussed. This evaluation through a focus group provided valuable insights, as discussed below, into the benefits and drawbacks of the explanatory tool, areas for further improvement, and the effectiveness of the explanations to enhance the understanding and trust of users.

7.2.1 General Impressions. The general impression among the participants was two-sided. On the one hand, the overall feel, look, and structure of the tool were appreciated and very clear, and the step-by-step approach was positively received. On the other hand, as the participants got further into the tool they started having more trouble understanding all the material. A rather obvious but important observation was the added difficulty of the language of the tool, the tool was in English while all the participants were Dutch. Most participants expressed confidence in their level of English, however, one participant mentioned, "Even though I think I understand English well, it makes these complex concepts even more difficult." Which was supported by the other participants

7.2.2 Understanding of concepts. The first difficulties were encountered in the second chapter on the input and output maps, not all concepts and relations were immediately clear. However, with a little extra effort and explanation all the participants managed to grasp this part of the explanation. "The input-output maps clarify how certain inputs influence the output and what combinations of input produce an output." The most difficulty was experienced in the third chapter. This was to be expected as this explains the mathematical background of the AI and is the most complex concept explained in the tool.

Although these explanations and concepts proved to be difficult to understand, all participants indicated that at the end of the course, they understood the general overview and bigger picture of the AI system. Participants were confident in their general understanding of the processes behind the AI system, while some of the technical details might still be unclear. The quiz questions in the introduction and conclusion indicated the general understanding of the participants, as all of them had all the questions right the first time.

An interesting issue addressed by the participants is the level of interest needed to fully grasp the concepts. As one participant described, "It is enough for me to learn about the 3 colors and texture layer as input that determine whether it is S. Marianum or not, while not knowing about the details of the SOM." The other participants agreed with this statement and indicated that that knowledge together with some examples would be enough for them. However, one of the participants, supported by the others, mentioned "It is impossible to know how much information a user might want, some users might want to know more about the technical details."

7.2.3 Clarity of explanations. While discussing the general clarity of the explanations without addressing the complexity of concepts, the impression among the participants was that it was still quite technical. Although already simplified, the explanations still included many new terms and concepts for which more basic knowledge is needed than most non-technical users have.

One of the participants noted, "Reading explanations twice helps to make the concepts land in my mind, helping with understanding certain concepts". Another participant mentioned, "It would be helpful to provide several additional examples of the same concepts to enhance understanding." Another one of the participants mentioned, "The concluding sentences of each piece of text help me follow the bigger picture of the explanation, similar to the effect of the summary in the concluding chapter" Another participant said, "The concluding chapter helped tie everything together for me. " Finally, a suggestion was made to add a summary of learned concepts at the end of each chapter, to have a small overview of the most important points.

7.2.4 Trust. The main goal of this research is to investigate whether the developed explanatory tool can help build trust in an AI system among users, to improve the adoption of AI. The consensus was that, although none of the participants fully understood the complex and technical details of the AI system, they all understood the general overview of the process behind the AI. One participant noted, supported by the others, "Even though I don't fully understand the technical details, I trust the system a lot more now that I know the general process behind it."

The knowledge about how the 4 input features of input images were processed to make a prediction is enough to understand and believe the prediction made by the AI system. That together with some examples of the AI at work provided plenty of information to help build the trust of the participants in the AI system. One participant mentioned, "Just knowing a well-thought-out mathematical model supports the AI system is enough to trust the system, even without understanding the mathematical model." The fact that "it isn't just the press of a button and magically an output appears" is sufficient knowledge, indicated another participant.

8 DISCUSSION

An important question raised during the focus group was whether it is possible to simplify the explanations in the tool even further to make them accessible to anyone. Although the general overview of the AI system was clear, the participants struggled with the more technical details such as the Self-Organizing Maps. This concept had already been simplified and excluded several details, but the idea remained complex. There needs to be a balance between further simplification and accuracy of the explanation. By removing even more details we risk losing important aspects, resulting in an incomplete understanding of the processes behind the AI system. Researchers like Miller [24] discussed that simplifying explanations is important, but that this should not be done at the expense of the completeness of an explanation. Participant feedback suggested that the provided explanations developed a high-level understanding while technical details still need more extensive explanations. In future work, adding more examples, analogies, and animations can help to break the concepts up into smaller parts, providing a more understandable explanation without removing essential details.

Another significant point was whether it is necessary to fully understand all the details before you can trust and use the AI tool. Technical details are often too complex and uninteresting to the user [12, 23]. This consideration is especially important in fields like agriculture, as the non-technical background of most users in this area requires much more effort to understand all the complex concepts [12, 32]. Our findings suggest that an explanation that provides users with enough information to understand the benefits and a high-level overview of how the AI works is sufficient for an increased understanding and more trust in the AI system. However, ensuring access to the technical details should still be supported for users who are interested or have a more technical background.

The findings of this research suggest that with some small improvements and additions, the developed tool can be effectively used for training farmers in the usage of this AI system. Several chapters of the tool provide clear and understandable explanations of important concepts surrounding AI systems, such as the benefits, goals, and inner workings. Although not all the complex concepts of the mathematical model of an AI system are easy to understand, the general picture of the process of the AI is clear. This research shows that this understanding is beneficial to the trust users have in the system and confirms that increased transparency into the AI can help build trust among the users. Through explanations and descriptions of the processes it becomes clear that the AI is not a "magic box" but a mathematical model, helping users build trust in the AI system [12, 34, 39].

These findings suggest that over time tools like this can lead to bigger changes in the agricultural sector. The tool can be extended to support several different AI systems, providing solutions to various problems in the sector, such as crop management, livestock monitoring, and water and soil management [4, 20]. As farmers increase their understanding of these AI systems through this tool, start to build trust in AI, and see the benefits of using AI, the integration of AI tools into their workflows will accelerate. Driving a change in the sector towards more efficient and data-driven practices.

The primary challenge identified through this research is to create explanations simple and accessible enough for non-technical users without losing important details of the AI tool in the explanation. The challenge of making complex AI systems understandable to a wide audience and the limited use found in this research for a deep understanding questions the arguments made in existing research on explainable AI, suggesting to aim for a broad and deep understanding of all the complex concepts [1, 12, 23, 32, 35].

9 CONCLUSION

In this research we investigated how we can increase the understanding and trust of farmers in AI systems through an explanatory tool, aiming to facilitate the adoption of AI in agriculture. This study addresses the need for Explainable AI in agriculture, helping farmers to use AI systems effectively in their practices.

Requirements for the tool were identified through a literature review of the theory on Explainable AI and a case study of the AI tool on weed detection. The design and development of the tool based on these requirements resulted in a chapter-based explanation, progressively introducing more complex details and key aspects of the analyzed AI tool such as the relation between input and output and the underlying mathematical model: Self-Organizing Maps. This explanatory tool was then evaluated through a focus group session to gain insights and feedback regarding the effectiveness of the tool.

The explanatory tool provides a clear and structured approach to explaining the benefits, drawbacks, and processes of the AI system. The evaluation through the focus group proved that the tool increased the understanding of participants in the AI system, significantly increasing their trust. However, this evaluation also indicated that complex concepts like the technical details of a SOM require further simplification to be understandable to a broader audience. On the other hand, this research shows that even without a deep technical understanding the trust can be significantly increased through a high-level overview. An understanding of the process and the notion that there exists a mathematical background to the AI tool is sufficient to build trust among the general audience.

The findings of this research show that an explanatory tool providing simple and accessible explanations enhances understanding and facilitates building trust in an AI system. Which can increase the adoption of AI, allowing farmers to make more informed decisions by utilizing AI systems. Accelerating the integration of AI into the agricultural sector, leading towards increased efficiency and sustainability in agricultural practices. This demonstrates that the explanatory tool can help bridge the gap between complex AI concepts and the target user, farmers. This research adds to the existing theory, by providing a practical example of using explainable AI for a real-world agricultural AI system.

Limitations of this research and its findings include the small number of participants for the evaluation and the focus of the explanatory tool on a very specific AI system. Future research should explore the use of this explanatory tool and variations, such as further simplification or omitting technical details, for several other AI applications and across the agricultural sector as well as other domains. Investigating the results of similar experiments with variants of the tool, larger participant pools, and different AI systems would be beneficial to the generalizability of the findings.

ACKNOWLEDGMENTS

I would like to thank my supervisor Joschka Hüllmann for guiding me through the research process and answering any questions I came across.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–18. https://doi.org/10.1145/3173574.3174156

[2] Nikos Alexandratos and Jelle Bruinsma (Eds.). 2012. *World agriculture towards 2030/2050: the 2012 revision.* https://doi.org/10.22004/ag.econ.288998 ISSN: 2521-1838.

[3] Sulaiman Alsheibani, Chris Messom, and Yen Cheung. 2020. *Re-thinking the Competitive Landscape of Artificial Intelligence.* http://hdl.handle.net/10125/64460

[4] Ishana Attri, Lalit Kumar Awasthi, and Teek Parval Sharma. 2024. Machine learning in agriculture: a review of crop management applications. *Multimedia Tools and Applications* 83, 5 (Feb. 2024), 12875–12915. https://doi.org/10.1007/s11042-023-16105-2

[5] Benoit A. Aubert, Andreas Schroeder, and Jonathan Grimaudo. 2012. IT as enabler of sustainable farming: An empirical analysis of farmers' adoption decision of precision agriculture technology. *Decision Support Systems* 54, 1 (Dec. 2012), 510–520. https://doi.org/10.1016/j.dss.2012.07.002

[6] Davide Ballabio and Mahdi Vasighi. 2012. A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies. *Chemometrics and Intelligent Laboratory Systems* 118 (Aug. 2012), 24–32. https://doi.org/10.1016/j.chemolab.2012.07.005

[7] Astrid Bertrand, Tiphaine Viard, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2023. On Selective, Mutable and Dialogic XAI: a Review of What Users Say about Different Types of Interactive Explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3544548.3581314

[8] Marija Cubric. 2020. Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. *Technology in Society* 62 (Aug. 2020), 101257. https://doi.org/10.1016/j.techsoc.2020.101257

[9] Swarup Das, Indrajit Ghosh, Gouravmoy Banerjee, and Uditendu Sarkar. 2018. Artificial Intelligence in Agriculture: A Literature Survey. (May 2018).

[10] Anupam Datta, Shayak Sen, and Yair Zick. 2016. Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. In *2016 IEEE Symposium on Security and Privacy (SP)*. 598–617. https://doi.org/10.1109/SP.2016.42 ISSN: 2375-1207.

[11] Shirley Gregor and Izak Benbasat. 1999. Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly* 23, 4 (1999), 497–530. https://doi.org/10.2307/249487 Publisher: Management Information Systems Research Center, University of Minnesota.

[12] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A Survey of Methods for Explaining Black Box Models. *Comput. Surveys* 51, 5 (Aug. 2018), 93:1–93:42. https://doi.org/10.1145/3236009

[13] Moon Halder, Ayon Datta, Md Kamrul Hossain Siam, Shakik Mahmud, Md. Saem Sarkar, and Md. Masud Rana. 2023. A Systematic Review on Crop Yield Prediction Using Machine Learning. In *Intelligent Systems and Networks*, Thi Dieu Linh Nguyen, Elena Verdú, Anh Ngoc Le, and Maria Ganzha (Eds.). Springer Nature, Singapore, 658–667. https://doi.org/10.1007/978-981-99-4725-6_77

[14] IBM. 2024. Data Suggests Growth in Enterprise Adoption of AI is Due to Widespread Deployment by Early Adopters. https://newsroom.ibm.com/2024-01-10-Data-Suggests-Growth-in-Enterprise-Adoption-of-AI-is-Due-to-Widespread-Deployment-by-Early-Adopters

[15] Teuvo Kohonen. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43, 1 (1982), 59–69. https://doi.org/10.1007/BF00337288

[16] Teuvo Kohonen and Timo Honkela. 2007. Kohonen network. *Scholarpedia* 2, 1 (Jan. 2007), 1568. https://doi.org/10.4249/scholarpedia.1568

[17] Igor Kuzmanovski and Marjana Novič. 2008. Counter-propagation neural networks in Matlab. *Chemometrics and Intelligent Laboratory Systems* 90, 1 (Jan. 2008), 84–91. https://doi.org/10.1016/j.chemolab.2007.07.003

[18] Vijaya Lakshmi and Jacqueline Corbett. 2023. Using AI to Improve Sustainable Agricultural Practices: A Literature Review and Research Agenda. *Communications of the Association for Information Systems* 53, 1 (2023), 96–137. https://doi.org/10.17705/1CAIS.05305

[19] Christine P. Lee, Min Kyung Lee, and Bilge Mutlu. 2024. The AI-DEC: A Card-based Design Method for User-centered AI Explanations. https://doi.org/10.1145/3643834.3661576 arXiv:2405.16711 [cs].

[20] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. 2018. Machine Learning in Agriculture: A Review. *Sensors* 18, 8 (Aug. 2018), 2674. https://doi.org/10.3390/s18082674 Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[21] Eudes Smith M. Linheiro, Gitanjali R. Shinde, ParikShit N. Mahalle, and Riddhi Mirajkar. 2023. Explainable AI (XAI) for Agriculture. In *Industry 4.0 Convergence with AI, IoT, Big Data and Cloud Computing: Fundamentals, Challenges and Applications*. Vol. 4. Bentham Science Publishers, 196. Google-Books-ID: a7brEAAAQBAJ.

[22] Willem Melssen, Ron Wehrens, and Lutgarde Buydens. 2006. Supervised Kohonen networks for classification problems. *Chemometrics and Intelligent Laboratory Systems* 83, 2 (Sept. 2006), 99–113. https://doi.org/10.1016/j.chemolab.2006.02.003

[23] Christian Meske, Enrico Bunde, Johannes Schneider, and Martin Gersch. 2022. Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management* 39, 1 (Jan. 2022), 53–63. https://doi.org/10.1080/10580530.2020.1849465 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10580530.2020.1849465.

[24] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (Feb. 2019), 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[25] David L. Morgan and Margaret T. Spanish. 1984. Focus groups: A new tool for qualitative research. *Qualitative Sociology* 7, 3 (Sept. 1984), 253–270. https://doi.org/10.1007/BF00987314

[26] Rosana Cavalcante de Oliveira and Rogério Diogne de Souza e Silva. 2023. Artificial Intelligence in Agriculture: Benefits, Challenges, and Trends. *Applied Sciences* 13, 13 (Jan. 2023), 7405. https://doi.org/10.3390/app13137405 Number: 13 Publisher: Multidisciplinary Digital Publishing Institute.

[27] X. E. Pantazi, A. A. Tamouridou, T. K. Alexandridis, A. L. Lagopodi, J. Kashefi, and D. Moshou. 2017. Evaluation of hierarchical self-organising maps for weed mapping using UAS multispectral imagery. *Computers and Electronics in Agriculture* 139 (June 2017), 224–230. https://doi.org/10.1016/j.compag.2017.05.026

[28] Ken Peffers, Tuure Tuunanen, Marcus Rothenberger, and S. Chatterjee. 2007. A design science research methodology for information systems research. *Journal of Management Information Systems* 24 (Jan. 2007), 45–77.

[29] Jayanthi Radhakrishnan and Manojit Chattopadhyay. 2020. Determinants and Barriers of Artificial Intelligence Adoption – A Literature Review. In *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation*, Sujeet K. Sharma, Yogesh K. Dwivedi, Bhimaraya Metri, and Nripendra P. Rana (Eds.). Springer International Publishing, Cham, 89–99. https://doi.org/10.1007/978-3-030-64849-7_9

[30] Amal Ben Rjab, Sehl Mellouli, and Jacqueline Corbett. 2023. Barriers to artificial intelligence adoption in smart cities: A systematic literature review and research agenda. *Government Information Quarterly* 40, 3 (June 2023), 101814. https://doi.org/10.1016/j.giq.2023.101814

[31] Masahiro Ryo. 2022. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture* 6 (Jan. 2022), 257–265. https://doi.org/10.1016/j.aiia.2022.11.003

[32] Waddah Saeed and Christian Omlin. 2023. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems* 263 (March 2023), 110273. https://doi.org/10.1016/j.knosys.2023.110273

[33] Johanes Schneider and Joshua Handali. 2019. Personalized explanation in machine learning: A conceptualization. https://doi.org/10.48550/arXiv.1901.00770 arXiv:1901.00770 [cs, stat].

[34] Julian Senoner, Simon Schallmoser, Bernhard Kratzwald, Stefan Feuerriegel, and Torbjørn Netland. 2024. Explainable AI improves task performance in human-AI collaboration. https://doi.org/10.48550/arXiv.2406.08271 arXiv:2406.08271 [cs].

[35] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (Feb. 2021), 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

[36] Nikolaos L. Tsakiridis, Themistoklis Diamantopoulos, Andreas L. Symeonidis, John B. Theocharis, Athanasios Iossifides, Periklis Chatzimisios, George Pratos, and Dimitris Kouvas. 2020. Versatile Internet of Things for Agriculture: An eXplainable AI Approach. In *Artificial Intelligence Applications and Innovations*, Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis (Eds.). Springer International Publishing, Cham, 180–191. https://doi.org/10.1007/978-3-030-49186-4_16

[37] John Venable, Jan Pries-Heje, and Richard Baskerville. 2016. FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems* 25, 1 (Jan. 2016), 77–89. https://doi.org/10.1057/ejis.2014.36

[38] Wikipedia. 2024. Self-organizing map. https://en.wikipedia.org/w/index.php?title=Self-organizing_map&oldid=1228749485 Page Version ID: 1228749485.

[39] Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. 2019. Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In *Natural Language Processing and Chinese Computing*, Jie Tang, Min-Yen Kan, Dongyan Zhao, Sujian Li, and Hongying Zan (Eds.). Springer International Publishing, Cham, 563–574. https://doi.org/10.1007/978-3-030-32236-6_51

[40] Jure Zupan, Marjana Novič, and Johann Gasteiger. 1995. Neural networks with counter-propagation learning strategy used for modelling. *Chemometrics and Intelligent Laboratory Systems* 27, 2 (Feb. 1995), 175–187. https://doi.org/10.1016/0169-7439(95)80022-2