DISSERTATIONES
INFORMATICAE
UNIVERSITATIS
TARTUENSIS
57

MUHAMMAD UZAIR

Soft decision making for agri-food 4.0

DISSERTATIONES INFORMATICAE UNIVERSITATIS TARTUENSIS

57

DISSERTATIONES INFORMATICAE UNIVERSITATIS TARTUENSIS
57

MUHAMMAD UZAIR

Soft decision making for agri-food 4.0

UNIVERSITY OF TARTU
Press
1632

Institute of Computer Science, Faculty of Science and Technology, University of Tartu, Estonia.

Dissertation has been accepted for the commencement of the degree of Doctor of Philosophy (PhD) in Computer Science on September 3, 2024 by the Council of the Institute of Computer Science, University of Tartu.

Supervisors

Prof.	Stefania Tomasiello Institute of Computer Science, University of Tartu, Estonia Department of Industrial Engineering, University of Salerno, Italy
Assoc. Prof.	Radwa Mohamed El Emam El Shawi Institute of Computer Science, University of Tartu, Estonia
Assoc. Prof.	Evelin Loit-Harro Institute of Agricultural and Environmental Sciences, University of Life Sciences, Estonia

Opponents

Prof.	Ferdinando Di Martino Department of Architecture, University of Naples Federico II, Italy
Prof.	Çağatay Çatal Department of Computer Science and Engineering, Qatar University, Qatar

The public defense will take place on October 15, 2024 at 10:15 in Narva Rd. 18-2046.

The publication of this dissertation was financed by the Institute of Computer Science, University of Tartu.

ISSN 2613-5906 (print) ISSN 2806-2345 (pdf)
ISBN 978-9916-27-666-2 (print) ISBN 978-9916-27-667-9 (pdf)

University of Tartu Press
http://www.tyk.ee/

To my family and friends

ABSTRACT

Sustainable agriculture is increasingly critical in the face of climate change and global food security challenges. Efficient management of water resources, accurate prediction of crop water requirements, and effective soil carbon management are essential for enhancing agricultural productivity and sustainability. Traditional methods often fall short due to their complexity, cost, and limited scalability. Therefore, there is a pressing need for innovative, data-driven approaches that leverage recent advancements in machine learning to address these challenges effectively.

To address the aforementioned challenges, we explored and evaluated various data-driven techniques for predicting evapotranspiration, soil organic carbon, wheat disease detection, and other critical parameters essential for sustainable agriculture. For evapotranspiration problem, we compared the performance of a recent variant of the Adaptive Neuro-Fuzzy System, namely ANFIS-T, with the state-of-the-art approaches. It was observed that ANFIS-T offers the highest interpretability with comparable accuracy and reduced training time compared to the state-of-the-art approaches. This highlights the importance of interpretable models in agricultural water management, providing actionable insights for better irrigation strategies. Secondly, we focused on the prediction of soil organic carbon (SOC) using ANFIS-T and Extreme Learning Machine (ELM) with fractional Tikhonov regularization (FTrELM). It was observed that ANFIS-T consistently achieves better accuracy compared to FTrELM, though FTrELM excels in training speed. The findings emphasize the potential of these advanced machine learning (ML) techniques in enhancing soil carbon management, which is crucial for mitigating climate change and improving soil health. Additionally, integrating these ML techniques into a cohesive framework for sustainable agriculture is discussed, in particular for early actions through data-driven ex-ante Life Cycle Assessment (LCA) for crop production. By combining different models, this research provides a comprehensive toolkit for stakeholders to make informed decisions, optimize resource use, and improve crop yields sustainably. Finally, we proposed a novel AutoML framework that incorporates contextual information to improve the accuracy and efficiency of disease detection in wheat crops. The study compares various ML models and demonstrates that incorporating context-aware features significantly enhances model performance.

Overall, the study addressed critical challenges in agriculture through the application of interpretable ML approaches and context-aware AutoML techniques. By enhancing the accuracy, efficiency, and interpretability of ML models in predicting evapotranspiration, SOC, and crop diseases, in addition to data-driven ex-ante LCA, this research contributes to the development of sustainable and resilient agri-food systems. The findings underscore the importance of integrating advanced ML with domain-specific knowledge to achieve practical and impactful solutions in agriculture.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANFIS-T	Adaptive Neuro-Fuzzy Inference System with Fractional Tikhonov Regularization
CANFIS	Coactive Neuro-Fuzzy Inference System
CANFIS-T	Co-Active Neuro-Fuzzy Inference System with Fractional Tikhonov Regularization
LCA	Life Cycle Assessment
SOC	Soil Organic Carbon
TPOT	Tree-based Pipeline Optimization Tool
UAV	Unmanned Aerial Vehicle
ML	Machine Learning
AutoML	Automated Machine Learning
GWP	Global Warming Potential
MF	Membership Function
FIS	Fuzzy Inference System
FT-rELM	Fourier Transformed Regularized Extreme Learning Machine
NSGA-II	Non-dominated Sorting Genetic Algorithm II
ARIMA	AutoRegressive Integrated Moving Average
CAAFE	Context-Aware Automated Feature Engineering
LLM	Large Language Model
RMSE	Root Mean Square Error
XGBoost	Extreme Gradient Boosting
CART	Classification and Regression Trees
GRNN	Generalized Regression Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression

LIST OF ORIGINAL PUBLICATIONS

Publications included in the thesis

I. **Uzair, M.**, Tomasiello, S., Loit, E. (2023). Interpretable Approaches to Predict Evapotranspiration. In: Abraham, A., Hanne, T., Gandhi, N., Manghirmalani Mishra, P., Bajaj, A., Siarry, P. (eds) **Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition. SoCPaR 2022. Lecture Notes in Networks and Systems**, vol 648. Springer, Cham. pages 275-284, https://doi.org/10.1007/978-3-031-27524-1_26
Author's contribution: (Leading Author) Contributed in identifying the problem, data pre-processing, conducted the experiments and some contributions in writing the paper.

II. **Uzair, M.**, Tomasiello, S., Loit, E., Wei-Lin, J., C. Predicting the soil organic carbon by recent machine learning algorithms, **2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress**, 2022, pp. 1-7, doi: 10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9928005
Author's contribution: (Leading Author) Identified the problem with co-authors, Data pre-processing, designed and conducted experiments, wrote some parts of the paper

III. Tomasiello, S., **Uzair, M.**, Liu, Y., Loit, E. (2023) Data-driven approaches for sustainable agri-food: coping with sustainability and interpretability. **Journal of Ambient Intelligence and Humanized Computing** 14, 16867–16878. https://doi.org/10.1007/s12652-023-04702-w
Author's contribution: I helped with data preparation and experiments

IV. **Uzair, M.**, El Shawi, R., Tomasiello, S., (2024) Context-Aware AutoML Approach to detect a wheat disease. **Submitted**
Author's contribution: (Leading Author) Identified the problem, proposed different solutions, acquired data, data preparation, experiments, and wrote some parts of the paper.

Publications not included in the thesis

V. Tomasiello S., **Uzair M.**, Loit E. (2021) ANFIS with fractional regularization for supply chains cost and return evaluation. **CEUR Workshop Proceedings, 13th International Workshop on Fuzzy Logic and Applications**, WILF. volume 3074. pages 1-9.

VI. Govardanan, Chemmalar, Y., Gokul, S., Gautam, M., Ramalingam, K., Dasaradharami, R., **Uzair, M.**, Gadekallu, T., Explainable AI for the Metaverse: A Short Survey, 2023 **International Conference on Intelligent Metaverse Technologies & Applications (iMETA)**,
pp. 1-6, doi: 10.1109/iMETA59369.2023.10294907.

PREFACE

1. INTRODUCTION

Modern agriculture and food production systems are facing increasing pressure from climate change, land and water availability, and, more recently, a pandemic. Building resilient and sustainable farming systems, to meet the challenges of agri-food production, is the ultimate concept behind Agriculture 4.0 (or more generally, Agri-Food 4.0). Agriculture 4.0 (by analogy with Industry 4.0) represents the fourth evolution in farming technology, employing and integrating different technologies, such as IoT, cloud computing, artificial intelligence. It aims at increasing productivity, allocating resources reasonably, adapting to climate change, and avoiding food waste. Although the discussion is moving towards Agriculture 5.0, which is a paradigm based on robots aiming to solve the workforce shortage in farms [SR20], there are still open issues in the context of Agri-Food 4.0, despite the fact that several decision support systems (DSS) have appeared over the last few years (e.g., for water resource management, farm operation scheduling, delivery plans) [Zha+20].

In addition to the above-mentioned agricultural objectives, there is also aim to decrease carbon emission within the EU Green Deal policy. This aspect is particularly important to optimize the farming supply chain with close respect to production sustainability. Hence, the choice of suitable techniques is critical.

Decision making deals with choosing an action from possible alternatives based on the input information. In the context of hard decision making, the problem is described by dividing the input space into a certain number of disjoint decision regions. This may lead to the choice of an inappropriate action near the decision boundary in the input space. Soft decision making, congruently with soft computing, uses some formalisms to take into account uncertainty and imprecision (e.g. fuzzy sets, rough sets, interval analysis) jointly with bioinspired computing schemes (neural networks, evolutionary techniques), in order to capture the hesitant idea of human beings in the decision-making process [IN00; Li13; Suo+20]. Investigating an interpretable framework, especially data-driven, is still in its infancy [Hom+20].

The innovativeness of this doctoral project is addressing the above-mentioned open issues, by formally deducing a multi-purpose soft decision making framework, which is interpretable. Interpretability has received attention from the European Parliament, whose General Data Protection Regulation recognizes any citizien's right to receive an explanation for algorithmic decisions [Eur16]. Interpretability of machine learning models is critical in agriculture due to the socioeconomic implications of agricultural decisions. Models must be understandable to stakeholders, including farmers, agronomists, and policymakers, to facilitate informed decision-making. In papers [Muh+22; MTL23], we discuss our proposed approach, which offers high interpretability. Our approach provides transparent decision-making processes that can be easily understood and trusted by users. The interpretability of our approach is particularly valuable in predicting complex agricultural phenomena like evapotranspiration and soil organic carbon (SOC). It must be pointed out that regarding SOC prediction, Estonian data was used.

Sustainability in agriculture is paramount due to the growing need to balance food production with environmental wellbeing. Sustainable agriculture practices aim to maintain productivity while minimizing negative environmental impacts. One key area where soft computing techniques contribute to sustainability is through Life Cycle Assessment (LCA). LCA evaluates the environmental effects of a product over its entire life cycle, thereby enhancing resource efficiency and reducing environmental liabilities. Traditional LCAs are mostly ex-post, conducted after the product lifecycle has ended, which limits their proactive use in improving sustainability during the development phase. We worked on the ex-ante LCA, allowing for predicting and mitigating environmental impacts early in the product development process, considering the wheat production case study. The proposed approach, which extends the one in papers [Muh+22; MTL23], is not only interpretable but sustainable itself, meeting the green AI requirement of limited energy consumption over the training.

Automated Machine Learning (AutoML) represents a significant advancement in the field of machine learning. AutoML aims to automate the process of model selection, hyperparameter tuning, and feature engineering, reducing the need for extensive human intervention and expertise. This is especially beneficial in agriculture, where the timely and accurate detection of issues such as crop diseases can significantly impact productivity and food security. We used AutoML in the detection of wheat stripe rust, a severe disease affecting wheat crops globally, discussed in paper IV. By integrating AutoML with context-aware feature engineering, we were able to outperform traditional deep-learning models like ResNet-18.

In conclusion, the integration of soft computing techniques in agriculture addresses critical issues of sustainability, interpretability, and efficiency. We were able to develop solutions that enhance soil health, optimize water use, predict environmental impacts, and improve disease detection. These solutions not only support sustainable agricultural practices but also ensure that the benefits of machine learning are accessible and understandable to all stakeholders involved. As the field continues to evolve, the ongoing refinement and application of these techniques will play a crucial role in meeting the global challenges of food security and environmental sustainability.

1.1. Research questions

- How to achieve interpretability and efficiency with a neuro-fuzzy system in comparison with different state-of-the-art machine learning algorithms?
- In what ways can data-driven approaches, and in particular, neuro-fuzzy systems, enhance sustainability in agri-food systems, and how do these methods compare with traditional techniques in terms of computational efficiency and interpretability?

- Does hand-crafted features from images work well with machine learning algorithms as compared to deep learning models?
- Can automated machine learning (AutoML) techniques generate model pipelines that achieve superior performance compared to traditional deep learning models in various predictive tasks? Can context-aware automated feature engineering improve the overall performance of machine learning models?
- Can automated machine learning (AutoML) techniques generate model pipelines that achieve superior performance compared to traditional deep learning models in various predictive tasks?

1.2. Aim and objectives

The aim of this research was to utilize soft computing techniques to address different challenges in agri-food 4.0. Fuzzy logic was the main component that was emphasized in soft computing techniques. The research objectives include interpretability of machine learning algorithms, sustainability in agri-food, and applying automated machine learning (AutoML) in agri-food problems. The agri-food challenges addressed during this research include prediction of soil organic carbon and evapotranspiration, ex-ante LCA of wheat production, and detection of a type of wheat disease using imagery data collected from different sources.

1.3. Thesis Organization

The thesis starts with the introduction, describing the problem and motivation,, research questions, and aims and objectives. We then explain the contributions in detail in Chapter II. Chapter III explains the dataset, methodology, results, and discussion, followed by conclusions in the last chapter.

2. PROBLEM DESCRIPTION

The agri-food sector faces significant challenges due to climate change and the increasing need for food production. Sustainable practices in this sector are crucial to ensure long-term productivity and environmental health. The accurate prediction of agricultural parameters such as evapotranspiration and soil organic carbon (SOC) is essential for effective water management, soil health, and overall crop yield.

Evapotranspiration, the total water loss from soil and crops, accounts for about 90% of agricultural water use. Accurate prediction of this parameter is vital for efficient irrigation strategies. Various machine learning techniques, including interpretable models like Decision Trees and Adaptive Neuro Fuzzy Inference Systems (ANFIS), are employed to predict evapotranspiration. These models balance accuracy and interpretability, making them useful for practical applications where understanding model decisions is as important as the predictions themselves.

Similarly, predicting SOC is crucial for maintaining soil health and achieving sustainability goals such as those set by the United Nations and the European Union. SOC influences soil properties and plays a role in mitigating climate change. Techniques like ANFIS and Extreme Learning Machines (ELM) with fractional Tikhonov regularization are effective in accurately predicting SOC and enhancing the interpretability of the classical ANFIS scheme. These models provide insights into the complex relationships between SOC, soil properties, and environmental factors, contributing to sustainable agricultural practices.

Automated Machine Learning (AutoML) is revolutionizing the field of agricultural disease detection. Traditional methods for detecting crop diseases like wheat stripe rust are labor-intensive and prone to human error. AutoML provides a robust and scalable framework for automating the entire machine learning pipeline, from data preprocessing to model selection and hyperparameter optimization.

In the context of wheat stripe rust detection, AutoML approaches integrate advanced feature engineering techniques to extract meaningful features from images of crops. These features are then used to train machine learning models that can accurately distinguish between healthy and diseased crops. For instance, the Tree-Based Pipeline Optimization Tool (TPOT) has been applied to optimize classification models, significantly outperforming traditional deep learning models like ResNet-18 in terms of accuracy and computational efficiency.

2.1. Sustainable and interpretable agri-food

2.1.1. Soil organic carbon prediction

SOC plays a crucial role in the global carbon cycle and is essential for maintaining soil health and productivity. It contributes to soil structure, water retention, and nutrient supply, which are critical for plant growth and agricultural sustainability

[Kan+22]. Soils are the largest terrestrial carbon reservoir, containing between 1,500 and 2,400 petagrams (Pg) of carbon within the one-meter depth [Dua+22]. Given this vast storage capacity, even small changes in SOC levels can significantly impact atmospheric carbon dioxide (CO_2) concentrations, thereby influencing global climate change [Xie+22].

The importance of SOC extends beyond its role in the carbon cycle. Enhancing SOC stocks is aligned with several Sustainable Development Goals (SDGs) of the United Nations, particularly those aimed at combating climate change and ensuring food security [UN 15]. The European Union's soil strategy for 2030 highlights the critical role of SOC in achieving net greenhouse gas removal targets, aiming for a reduction of 310 million tonnes of CO_2 equivalent annually.

A variety of techniques have been employed to predict SOC, incorporating different data sources and machine learning algorithms. For instance, in [Dua+22], researchers used the Random Forest (RF) algorithm combined with geospatial data from Google Earth Engine and soil samples to map SOC stocks in forest lands. This approach demonstrated the effectiveness of ensemble learning techniques in SOC prediction.

In [Ghe+21], the integrated Fourier Transform Infrared Spectroscopy (FTIR) data with environmental variables and tree-based machine learning models was used to predict SOC. The combination of FTIR, which provides detailed chemical information about soil samples, and advanced machine learning models significantly improved prediction accuracy. The Cubist model, optimized using a Bat algorithm (BA), yielded the best performance among several hybrid approaches.

Another study [Xie+22] has explored the use of remote sensing data, environmental covariates, and ensemble learning algorithms such as Gradient Boosted Decision Trees (GBDT) and Extreme Gradient Boosting (XGBoost) for SOC estimation. These methods underscore the importance of integrating diverse data sources and advanced algorithms to capture the complex factors influencing SOC.

2.1.2. Evapotranspiration prediction

This study investigates the critical significance of managing agricultural water resources effectively, especially given the challenges posed by climate change and the rising demand for food production. Agriculture, being a major consumer of water, faces significant losses through crop evapotranspiration (ET), which includes both evaporation from the soil and transpiration from plant surfaces. Accurate estimation of ET is paramount for devising efficient irrigation strategies that can optimize water usage and ensure sustainable agricultural practices [Sin88].

Traditional methods for measuring ET include direct techniques such as using evaporimeters and lysimeters [Er-+13]. These instruments, while providing reliable measurements, have practical limitations due to the necessity for daily data collection and challenges in deploying them across various regions. Consequently, indirect methods that utilize meteorological variables—such as sunshine hours, wind speed, relative humidity, rainfall, and temperature—have gained popularity. These methods often rely on empirical and semi-empirical models like the Penman–Monteith formula, which is widely used by the Food and Agriculture Organization (FAO) to estimate reference evapotranspiration [APS98].

In recent years, there has been a significant shift towards employing machine learning (ML) techniques for ET estimation. ML models, known for their ability to handle complex and nonlinear relationships, have improved the accuracy of ET predictions [LOL09; MA19; Bed21; SSJ22; SJS22]. However, a major concern with these models is their interpretability, particularly in applications with significant socioeconomic implications.

In the context of interpretability, Decision Trees (DTs) are highlighted for their simplicity and the ease with which their decision-making process can be understood [Bre+84]. DTs structure their predictions in a hierarchical manner, making it straightforward to trace how input variables influence the final output [Mol20]. However, DTs can become overly complex with deep trees, potentially compromising their interpretability [Bar+12; CG14].

Adaptive Neuro Fuzzy Inference System (ANFIS), a hybrid approach that combines neural networks with fuzzy logic principles is also discussed in this study along with DTs. ANFIS is designed to model fuzzy rules derived from data, providing a framework that is both flexible and interpretable. The introduction of fractional Tikhonov regularization to ANFIS (resulting in ANFIS-T) enhances its ability to produce interpretable models by controlling the complexity of the learned rules.

Support Vector Regression (SVR), a kernel-based technique [SBS00], is another model to consider. SVR is widely used in precision agriculture due to its robustness and accuracy. While not inherently interpretable, SVR can be adapted to offer insights into its predictions, making it a valuable addition to the comparative analysis.

2.1.3. Ex-ante Life Cycle Assessment for wheat production

The European Union (EU) is strongly committed to achieving the United Nations (UN) 2030 Agenda for Sustainable Development and its associated Sustainable Development Goals (SDGs). This commitment is reflected in the prioritization of sustainability issues across all sectors, including the agri-food industry. Two key methodologies frequently used to address sustainability in this context are Life Cycle Assessment (LCA) and Supply Chain Management (SCM) [GL07].

LCA is a comprehensive method for evaluating the environmental impacts of all stages of a product's life, from raw material extraction to production, use, and disposal. The primary aim of LCA is to increase resource use efficiency and minimize environmental liabilities. Traditionally, LCA is conducted ex-post, meaning after the product or process has been developed, to compare production systems or demonstrate compliance with environmental guidelines. However, ex-ante LCA, which evaluates potential environmental impacts at the research and development stage, is gaining traction. This approach allows developers to identify and mitigate environmental impacts early in the development process, potentially at lower costs. Despite its potential, ex-ante LCA outputs are indicative rather than definitive, as they rely on projections and assumptions [Van+20; Rea+08; TK22].

Several studies have highlighted the usefulness of ex-ante LCA in identifying environmental issues early in the development of various products, processes, or services, especially in the food and feed sectors [Ott+22]. For example, a study [Leo+22] applied ex-ante LCA to evaluate the impacts of biological nitrification inhibition in wheat production, assessing potential changes in fertilizer application rates and greenhouse gas emissions. The integration of machine learning (ML) techniques with LCA has been shown to enhance the interpretation and predictive capabilities of LCA results [Pri+22]. Formal frameworks combining ML with ex-ante LCA have been developed, utilizing techniques like Radial Basis Function Networks (RBFN) and Decision Trees (DT) to predict LCA indicators [KPK21].

SCM involves the planning, implementation, and control of supply chain operations to meet consumer demands effectively [GL07]. The Supply Chain Operations Reference (SCOR) model is a widely used framework for assessing supply chain processes, covering five main scopes: plan, source, make, deliver, and return [HSW04]. Each scope includes criteria for reliability, responsiveness, flexibility, cost, and asset management. Recent iterations of the SCOR model have incorporated sustainability metrics, addressing environmental concerns such as carbon emissions and waste generation.

Artificial Intelligence (AI) and digital technologies play a crucial role in advancing the SDGs, both by enhancing sustainability (AI for sustainability) and ensuring the sustainability of AI systems themselves [GL07]. The environmental impact of AI, particularly the energy consumption associated with training and deploying ML models, is a growing concern. Sustainable AI practices emphasize the need for both efficient algorithms and environmentally conscious AI development.

A new version of the Co-Active Neuro-Fuzzy Inference System (CANFIS) with fractional regularization, referred to as CANFIS-T, is introduced. CANFIS, a generalization of the Adaptive Neuro-Fuzzy Inference System (ANFIS), is designed to handle Multi-Input-Multiple-Output (MIMO) systems, unlike ANFIS, which is limited to Multi-Input-Single-Output (MISO) systems. Despite its robustness, CANFIS has often been misapplied in various studies [Aby+16; Hos13; Ayt09; PV16; MK15; MKP17; Gho+22], being used as a multi-input single-output system. Proper applications of CANFIS in the agri-food sector include predicting soil temperature, estimating evapotranspiration, modeling daily suspended sediment concentration, and forecasting irrigation depths [MKP17]. These studies demonstrate CANFIS's superior performance compared to traditional ANFIS and other ML models.

Ensemble ANFIS (e-ANFIS) models, which combine multiple ANFIS models to improve predictive accuracy, have been extensively used in the agri-food sector. For example, e-ANFIS has been applied to predict the output energy, economic productivity, and environmental emissions of canola production, complementing LCA analyses. Other applications include predicting the environmental indices of greenhouse production and assessing the sustainability of sugarcane production. However, the ensemble nature of e-ANFIS introduces computational complexity and potential error propagation, which can negatively impact its performance.

The concept of sustainable AI includes both the use of AI for sustainability purposes and the sustainability of AI systems themselves. This dual focus addresses the need for AI algorithms that are both effective in promoting sustainability and efficient in their energy consumption. Research has highlighted the significant environmental costs associated with training complex AI models, emphasizing the need for energy-efficient and interpretable AI approaches.

2.2. AutoML in agri-food

2.2.1. Rust stripes wheat disease detection

Detection and management of crop diseases is a pressing issue in global agriculture, particularly stripe rust in wheat. Wheat is a cornerstone of global food security, being the most widely cultivated crop worldwide [Ace+18]. According to the Food and Agriculture Organization (FAO), the global utilization of cereals, including wheat, is projected to see a 1% increase in 2024. However, wheat production is threatened by various diseases, with stripe rust, caused by the fungal pathogen *Puccinia striiformis* f.sp.tritici, being one of the most devastating. This disease can cause significant yield losses, surpassing the damage inflicted by other wheat diseases [Che20].

Traditional methods for detecting crop diseases rely heavily on manual inspection, which is not only time-consuming and labor-intensive but also prone to human error. This makes it impractical for large-scale agricultural applications. The limitations of manual inspection underscore the need for efficient, accurate, and scalable automated detection methods. Recent advancements in imaging technologies, particularly the use of Unmanned Aerial Vehicles (UAVs), offer promising alternatives. UAVs can capture high-resolution images over large agricultural areas, facilitating comprehensive disease surveillance.

The use of UAVs for plant disease detection has experienced considerable growth in recent years. UAVs offer a cost-effective and efficient means of acquiring high-resolution images over large agricultural areas, which, when combined with advanced image processing techniques, can significantly enhance disease detection and monitoring efforts.

Previous studies have explored various approaches for detecting diseases in different crops using features extracted from UAV-acquired images. For instance, researchers have developed methods for detecting narrow brown leaf spot in rice by extracting color features and vegetation indices from UAV images, followed by analysis using Pearson's correlation to identify influential features [Gu+23]. These features were then used in support vector regression to estimate disease severity.

Similarly, in mango leaf disease detection, researchers employed a systematic workflow involving preprocessing UAV imagery, converting it from RGB to HSI color space, and extracting features such as mean, variance, homogeneity, contrast, dissimilarity, entropy, and correlation [PM23]. Classification was performed using a variant of the Radial Basis Function Network (RBFN), showcasing the efficacy of the proposed approach. Another study on tomato plant diseases used deep learning models for classification after extracting features from RGB to HSI converted images [BPC23].

In the context of wheat disease detection, several studies have focused on different diseases such as powdery mildew and wheat Fusarium head blight.. These studies utilized texture features and vegetation indices, analyzed through regression models, to monitor disease presence and severity. For example, [Liu+23] focused on identifying powdery mildew in wheat by extracting textural features and using Partial Least Squares Regression (PLSR) for analysis. Another study on monitoring wheat Fusarium head blight used UAV remote sensing combined with vegetation indices and texture features to enhance predictive analysis using Support Vector Regression (SVR).

Expanding beyond wheat, other studies have applied UAV imagery for detecting diseases in crops like rubber trees and citrus plants. These studies highlight the versatility of UAV-based disease detection methodologies across different agricultural domains, emphasizing the critical role of advanced image processing techniques and ML algorithms in ensuring global food security.

The advent of AutoML techniques aims to automate the construction of ML pipelines, reducing the need for manual intervention. AutoML consists of various stages, including data preprocessing, feature engineering, model selection, hyperparameter optimization, and model evaluation. By using AutoML, users can streamline and expedite the model development process, significantly mitigating the need for extensive manual intervention.

This study integrates AutoML with context-aware feature engineering techniques to enhance the detection of stripe rust in wheat crops. The methodology involves extracting a diverse array of statistical features from UAV images and refining these features using Context-Aware Automated Feature Engineering. The resulting features are then used to train models using the Tree-Based Pipeline Optimization Tool (TPOT) [Ols+16], an open-source genetic programming-based AutoML system designed to optimize supervised classification models.

3. BACKGROUND

3.1. Preliminaries

Symmetric positive semi-definite matrix: A symmetric positive semi-definite matrix is a square matrix $A \in \mathbb{R}^{n \times n}$ that satisfies the following two properties:

1. **Symmetry**: The matrix A is symmetric, meaning $A = A^T$, where A^T is the transpose of A.
2. **Positive Semi-Definiteness**: For any non-zero vector $x \in \mathbb{R}^n$, the quadratic form $x^T A x \geq 0$. In other words, for all vectors x, the matrix produces a non-negative scalar when applying the quadratic form.

A matrix is *positive semi-definite* if all its eigenvalues are non-negative, and *positive definite* if all its eigenvalues are strictly positive.

Norm used in Section 3.2 and 3.3: The standard 2-norm has been used in the thesis unless explicitly defined differently.

Preliminaries sections 3.1.1 and 3.1.2 are taken from the book [TPL22] [1].

3.1.1. Fuzzy Sets

A *classical set* is a collection of distinct objects, where each object either fully belongs to the set or does not belong at all. This binary concept is represented by a characteristic function $\chi_A : X \to \{0,1\}$, where $\chi_A(x) = 1$ indicates that the element x is in the set A, and $\chi_A(x) = 0$ signifies that x is not in the set.

On the other hand, a *fuzzy set*, introduced by L.A. Zadeh in 1965, extends the notion of classical sets by allowing partial membership. In a fuzzy set B, each element x in the universe of discourse X is associated with a membership function $\mu_B : X \to [0,1]$. Here, $\mu_B(x)$ represents the degree to which x belongs to B. A value of $\mu_B(x) = 0$ means x does not belong to the set, while $\mu_B(x) = 1$ indicates full membership, and any value between 0 and 1 represents partial membership.

Example: Temperature Modeling. Consider the example of categorizing temperatures into the set of "warm" temperatures. In a classical set, the membership of a temperature might be defined as follows:

$$\chi_{\text{Warm}}(t) = \begin{cases} 1 & \text{if } 20°C \leq t \leq 30°C, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

In this scenario, any temperature strictly within the range of 20°C to 30°C is considered warm, while anything outside this range is not.

In contrast, a fuzzy set approach allows for a gradual transition between "not warm" and "warm." The membership function for "warm" temperatures might be

[1] Some text has been rephrased with the help of LLMs such as ChatGPT

defined as:

$$\mu_{\text{Warm}}(t) = \begin{cases} 0 & \text{if } t \leq 15°C, \\ \frac{t-15}{10} & \text{if } 15°C < t < 25°C, \\ 1 & \text{if } t \geq 25°C. \end{cases} \tag{3.2}$$

Here, temperatures below 15°C are considered not warm, temperatures between 15°C and 25°C gradually increase in their degree of "warmness," and temperatures above 25°C are fully considered warm.

Membership Functions. The *membership function* (MF) is a critical concept in fuzzy set theory, as it defines how the degree of membership of each element in the universe is determined. The choice of MF is usually determined by consulting some experts related to the field. Several types of membership functions are commonly used:

- **Triangular MF**: This is a simple and widely used type of membership function, which is defined by three parameters: a lower limit a, an upper limit c, and a peak b. The function is expressed as:

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \leq a \text{ or } x \geq c, \\ \frac{x-a}{b-a} & \text{if } a < x \leq b, \\ \frac{c-x}{c-b} & \text{if } b < x < c. \end{cases} \tag{3.3}$$

- **Gaussian MF**: This membership function is characterized by a smooth, bell-shaped curve. It is defined by a central value m and a standard deviation $\sigma \geq 0$. The Gaussian MF is expressed as:

$$\mu_A(x) = \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \tag{3.4}$$

- **Generalized Bell-Shaped MF:** This function computes fuzzy membership values using a generalized bell-shaped membership function.

$$\mu_A(x) = \left(1 + \left|\frac{x-\overline{c}}{\overline{a}}\right|^{2\overline{b}}\right)^{-1}, \tag{3.5}$$

  where $\overline{a}, \overline{b}, \overline{c}$ are the function parameters. The values for function parameters are:

  - $\overline{a}$ - controls the spread of the curve and it should be $\overline{a}_{ir} > 0$
  - $\overline{b}$ - controls the steepness/shape of curve and it should be $\overline{b}_{ir} > 0$
  - $\overline{c}$ - this parameter determines the location of the peak of the curve and it can take any real value $\overline{c}_{ir} \in \mathbb{R}$

Example of Membership Functions. Consider the scenario where we want to describe the concept of a "medium-sized apartment" using a fuzzy set. Let X represent the size of an apartment in square meters.

A triangular membership function could define "medium-sized" as follows:

$$\mu_{\text{Medium}}(x) = \begin{cases} 0 & \text{if } x \leq 50 \text{ or } x \geq 100, \\ \frac{x-50}{25} & \text{if } 50 < x \leq 75, \\ \frac{100-x}{25} & \text{if } 75 < x < 100. \end{cases} \tag{3.6}$$

Alternatively, a Gaussian membership function might describe "medium-sized" apartments centered around 75 square meters with a standard deviation of 10 square meters:

$$\mu_{\text{Medium}}(x) = \exp\left(-\frac{(x-75)^2}{200}\right) \tag{3.7}$$

Both functions provide different perspectives on how "medium-sized" can be interpreted in the context of apartment sizes.

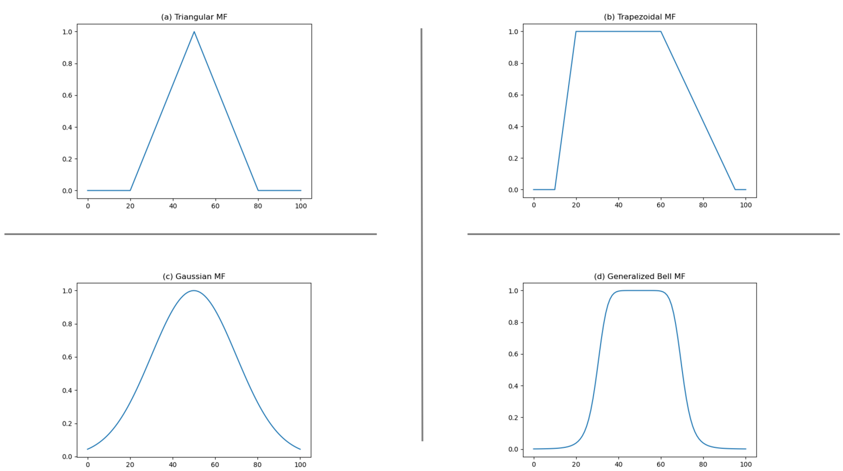

Figure 1. Examples of different membership functions

3.1.2. Fuzzy Inference Systems

A **Fuzzy Inference System (FIS)** is a framework that uses fuzzy logic to map inputs to outputs based on a set of fuzzy rules. These systems are widely used for decision-making processes where the information is uncertain or imprecise. An FIS works by taking inputs, processing them through fuzzy rules, and then producing an output, which can either be a fuzzy set or a specific numerical value after a process called defuzzification.

Components of a Fuzzy Inference System. A typical FIS consists of the following main components:

- **Fuzzification Module**: This module converts crisp inputs into degrees of membership by using membership functions. For each input value, the fuzzification module assigns a degree of membership between 0 and 1 in the relevant fuzzy sets.
- **Rule Base**: The rule base contains a collection of fuzzy IF-THEN rules. Each rule describes a relationship between the input variables and the output variables in the form of logical statements like: "IF *condition*, THEN *action*". For example:

  IF temperature IS low, THEN power IS high.

  These rules are created based on expert knowledge or experience.
- **Fuzzy Inference Engine**: This module processes the input information by applying the fuzzy rules from the rule base. It determines the degree to which each rule applies and combines these degrees to form a fuzzy output.
- **Defuzzification Module**: If a crisp (numerical) output is required, the defuzzification module converts the fuzzy output back into a single numerical value. Common defuzzification methods include the *centroid* method, which calculates the center of the area under the fuzzy output distribution.

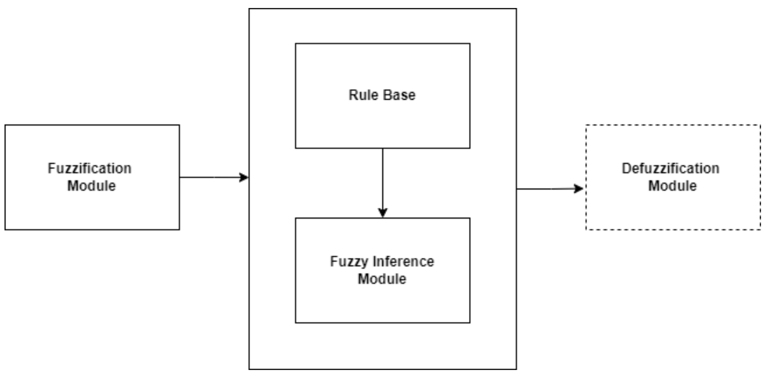

Figure 2. A general FIS scheme

Example of a Fuzzy Inference System. Consider a simple heating system controlled by an FIS. The inputs are the current temperature and the rate of change of temperature, and the output is the power setting for the heater. The fuzzy rule base might look like this:

- **Rule 1:** IF temperature IS low AND change of temperature IS negative small THEN power IS high.
- **Rule 2:** IF temperature IS comfortable AND change of temperature IS zero THEN power IS medium.

- **Rule 3:** IF temperature IS high AND change of temperature IS positive small THEN power IS low.

If the current temperature is 15°C and the temperature is decreasing slowly, the FIS might conclude, based on the fuzzy rules, that the heater should be set to a high power setting.

Defuzzification Example. Suppose the fuzzy output from the inference process indicates that the power should be somewhat medium and somewhat high. To decide on an exact power level, the FIS might use the centroid method to calculate the weighted average of the possible power levels, resulting in a specific numerical output like 75% power.

This process of moving from fuzzy sets back to a specific number is crucial for controlling systems where a precise action is needed.

Mamdani and Takagi-Sugeno-Kang Fuzzy Inference Systems.

Mamdani Fuzzy Inference System. The **Mamdani Fuzzy Inference System (FIS)** is one of the most widely used methods in fuzzy logic. It was first introduced by E.H. Mamdani in 1975 and is especially popular for its simplicity and ability to model expert knowledge.

In a Mamdani FIS, the rules are expressed in the form of fuzzy IF-THEN statements, such as:

$$\text{IF input1 is } A1 \text{ AND input2 is } A2, \text{ THEN output is } B.$$

Here's how the Mamdani FIS works:

1. **Fuzzification**: Convert the crisp input values into fuzzy values using membership functions.
2. **Rule Evaluation**: Apply the fuzzy rules to the fuzzy inputs to determine the degree to which each rule applies.
3. **Aggregation**: Combine the outputs of all the rules to form a single fuzzy set.
4. **Defuzzification**: Convert the fuzzy output into a crisp value. Common defuzzification methods include the centroid or center of gravity method, which calculates the center of the area under the aggregated fuzzy set.

For example, consider a simple heating system where the input variables are *temperature* and *rate of temperature change*, and the output is the *heater power*. A rule might be:

IF temperature is low AND rate of change is slow, THEN heater power is high.

This system would evaluate the degree to which the current conditions match this rule and then aggregate the results to decide on the heater power.

Takagi-Sugeno-Kang Fuzzy Inference System. The **Takagi-Sugeno-Kang (TSK) Fuzzy Inference System** differs from the Mamdani FIS in how it handles the output of each rule. Instead of using fuzzy sets for the output, TSK uses mathematical functions of the input variables.

A typical rule in a TSK system is written as:

$$\text{IF input1 is } A1 \text{ AND input2 is } A2, \text{ THEN output } y = f(input1, input2).$$

Where $f(input1, input2)$ is a linear or nonlinear function, often a linear combination of the inputs.

Here's how the TSK FIS works:

1. **Fuzzification**: Similar to Mamdani, convert the crisp inputs into fuzzy values.
2. **Rule Evaluation**: Apply the fuzzy rules, but instead of fuzzy outputs, calculate a function of the inputs for each rule.
3. **Weighted Average**: Combine the outputs of all the rules using a weighted average, where the weights are the degrees to which each rule applies.

For example, in a TSK system controlling the same heating system, a rule might be:

$$\begin{gathered}\text{IF temperature is low AND rate of change is slow,}\\ \text{THEN heater power } y = 0.5 \times \text{temperature}\\ +0.8 \times \text{rate of change} + 2.\end{gathered}$$

The final output is a weighted average of all the rules' outputs, making TSK FIS particularly useful in systems requiring precise, numerical outputs.

3.2. Adaptive Neuro-Fuzzy Inference System with Fractional Tikhonov Regularization (ANFIS-T)

Jang [Jan93] introduced a five-layered network architecture called Adaptive Network-based Fuzzy Inference System or Adaptive Neuro-Fuzzy Inference System (ANFIS) to represent the Takagi-Sugeno fuzzy inference system. The network contains fixed and adaptive nodes (i.e. the outputs of the latter are dependent on the parameters of the node).

ANFIS uses a hybrid learning approach which includes gradient descent with backpropagation and least-squares (LS) method. Learning is the process to tune parameters in a way to achieve better performance. The standard learning algorithm in ANFIS aims at adjusting both the antecedents and the consequent parameters of the fuzzy rules.

The five layers of ANFIS can be described as fuzzification layer (L1), product layer (L2), normalization layer (L3), consequent functions layer (L4), and the

output layer (L5).
In the first layer, the input is processed by means of membership functions (MFs), such as the generalized-bell shaped membership function or the Gaussian one.

The formulas for each layer in general form are:

- $L1: \quad O_i = \mu_{A_{ir}}(x_i);$
- $L2: \quad w_r = \Pi_{i=1}^{n} O_i;$
- $L3: \quad \overline{w}_r = w_r / \sum_{j=1}^{R} w_j;$
- $L4: \quad o_r = \overline{w}_r f_r = \overline{w}_r(\theta_{0r} + \theta_{1r}x_1 + \ldots + \theta_{nr}x_n);$
- $L5: \quad f = \sum_{r=1}^{R} o_r$

where $\mu_{A_{ir}}$ represents the membership function, and A_{ir} are fuzzy sets, representing linguistic attributes of the input x_i, $i = 1, 2, ..., n$, in the r-th rule $(r = 1, 2, ..., R)$. The values θ_{ir} are the parameter set and they are referred to as *consequent parameters*. The parameters θ_{ir} are also called the unknown parameters which are determined by the least squares method.

In general, ANFIS implements the rules of the form

IF x_1 is A_{1r} AND ... AND x_n is A_{nr}
THEN $f_r = \theta_{0r} + \theta_{1r} + \ldots + \theta_{nr}$

An example of ANFIS scheme is depicted in Figure 3.

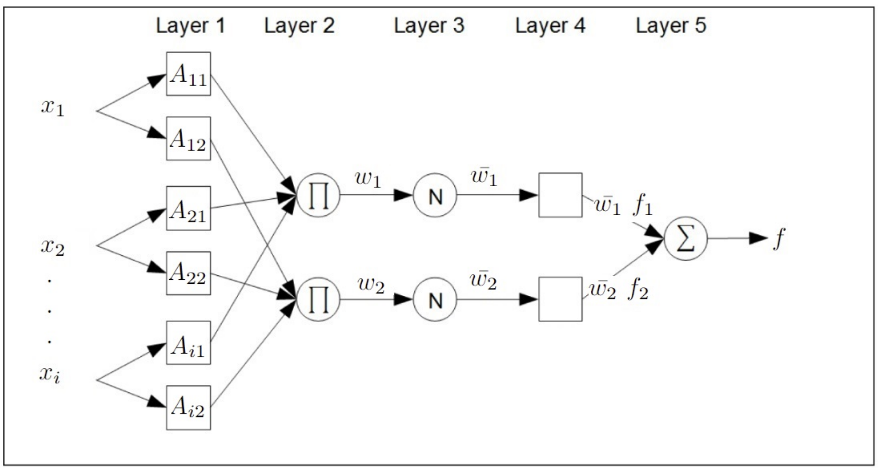

Figure 3. ANFIS architecture [Muh+22]

As mentioned before, ANFIS uses a hybrid learning algorithm based both on backpropagation and LS method. The node outputs progress to layer L4, where the consequent parameters are determined using the LS approach for fixed antecedent parameters. A matrix equation is obtained using the training data such as [TPL22]

$$\mathbf{H}\theta = \mathbf{o}, \tag{3.8}$$

where matrix $\mathbf{H}$ represents the input, θ is a vector collecting the unknown parameters and $\mathbf{o}$ is the vector of the target values.

The LS method is formulated as [2]:

$$\min_{\theta} \|\mathbf{H}\theta - \mathbf{o}\|^2 \tag{3.9}$$

with the solution

$$\theta* = \overline{\mathbf{H}}\mathbf{o}, \tag{3.10}$$

where $\overline{\mathbf{H}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the pseudoinverse of $\mathbf{H}$. In the standard learning algorithm for ANFIS, the antecedent parameters are then adjusted using back-propagation.

In ANFIS-T, the learning algorithm is based only on LS with fractional Tikhonov regularization.

The fractional Tikhonov method formalizes the following minimization problem

$$\min_{\theta} \|\mathbf{H}\theta - \mathbf{o}\|_P^2 + \lambda\|\mathbf{v}\|^2, \tag{3.11}$$

where $\|\theta\|_P$[3] $= (\theta^T\mathbf{P}\theta)^{0.5}$ and $\mathbf{P}$ is a symmetric positive semi-definite matrix defined as

$$\mathbf{P} = (\mathbf{H}^T\mathbf{H})^{\frac{\alpha-1}{2}}, \tag{3.12}$$

where $\alpha \in (0,1)$ is the fractional regularization parameter and λ a general regularization parameter ($\lambda \in R_+$).

The solution is [TPL22]:

$$\theta* = (\mathbf{M}^{\frac{\alpha+1}{2}} + \lambda\mathbf{I})^{-1}\mathbf{M}^{\frac{\alpha-1}{2}}\mathbf{H}^T\mathbf{o}, \tag{3.13}$$

where $\mathbf{M} = \mathbf{H}^T\mathbf{H}$, and red$\mathbf{I}$ is an identity matrix. The o_r represents partial output of a node in $L4$ and final output is calculated at $L5$ by $\sum_{r=1}^{R} o_r$. However, the bold $\mathbf{o}$ in Eq 3.8 is the vector of target values. The values θ_{ir} represents the unknown parameters of a node in $L4$ and θ is the vector that contains all the unknown parameters (Eq 3.8).

In [TPL22], it has been formally proved that ANFIS-T can achieve a high accuracy by a small number of rules, with the number of rules equal to the number of terms. In this way, the interpretability is highly improved. For instance, if there are n input variables x_i and the number of rules is set to 2 (implying 2 terms), the following rules can be extracted from ANFIS-T:

IF x_1 is SMALL AND … x_n is SMALL THEN o_1,
IF x_1 is LARGE AND … x_n is LARGE THEN o_2.

[2] Standard $l2-norm$ is used in Eq. 3.9

[3] Standard $2-norm$ is used for $\|\theta\|$

3.3. Co-Active Neuro-Fuzzy Inference System with Fractional Tikhonov Regularization (CANFIS-T)

The Co-Active Neuro-Fuzzy Inference System (CANFIS) is a generalization of the Adaptive Neuro-Fuzzy Inference System (ANFIS) [MJ95]. Like ANFIS [Jan93], it presents a multi-layered network architecture to describe the Takagi-Sugeno fuzzy inference system, but unlike ANFIS, it allows for multiple outputs. Hence, while ANFIS models Multi-Input-Single-Output (MISO) systems, CANFIS models Multi-Input-Multiple-Output (MIMO) systems. In CANFIS, the fuzzy rules are constructed with shared membership values to take into account any possible correlation among outputs.

Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be the input vector with n attributes. Let L_i denote the ith layer of this network. The operations performed through the different layers can be summarized as follows:

- $L1: \quad u_{ir} = \mu_{A_{ir}}(x_i);$
- $L2: \quad w_r = \Pi_{i=1}^{n} u_{ir};$
- $L3: \quad f_{rk} = w_r C_{rk}(\mathbf{x});$
- $L4: \quad \overline{o}_k = \sum_{r=1}^{R} f_{rk},$
- $L5: \quad o_k = \overline{o}_k \left(\sum_{j=1}^{R} w_j\right)^{-1}.$

where $\mu_{A_{ir}}$ represents the membership function, and A_{ir} are fuzzy sets representing linguistic attributes of the input x_i in the r-th rule $(r = 1, 2, ..., R)$, and o_k, with $k = 1, \ldots, p$, are the p computed outputs. The linear functions $C_{rk}(\mathbf{x})$ are a linear combinations of x_i through $nRp \times p$ unknown parameters.

The fuzzy sets are uniquely identified by means of membership functions (MFs), here assumed to be parameterized functions such as the gaussian function (Eq 3.4) or generalized bell-shaped function (Eq 3.5).

A CANFIS scheme is depicted in Figure 4. Like ANFIS, the standard CANFIS uses a hybrid learning approach, including backpropagation and least-squares (LS) method.

Given N training samples, the following matrix equation is obtained using the training data:

$$\mathbf{H}\theta = \mathbf{O}, \tag{3.14}$$

where $\mathbf{O}$ is the $N \times p$ output matrix , θ is the matrix of the unknown parameters with dimension $nRp \times p$, $\mathbf{H} = [\overline{\mathbf{H}}_1 \ldots, \overline{\mathbf{H}}_p]$ is the block matrix consisting of $N \times nR$ matrices $\overline{\mathbf{H}}_i$, $i = 1, \ldots, p$.

The LS method is formulated as [4]

$$\min_{\theta} ||\mathbf{H}\theta - \mathbf{O}||^2 \tag{3.15}$$

[4] Standard $2 - norm$ is used in Eq. 3.15

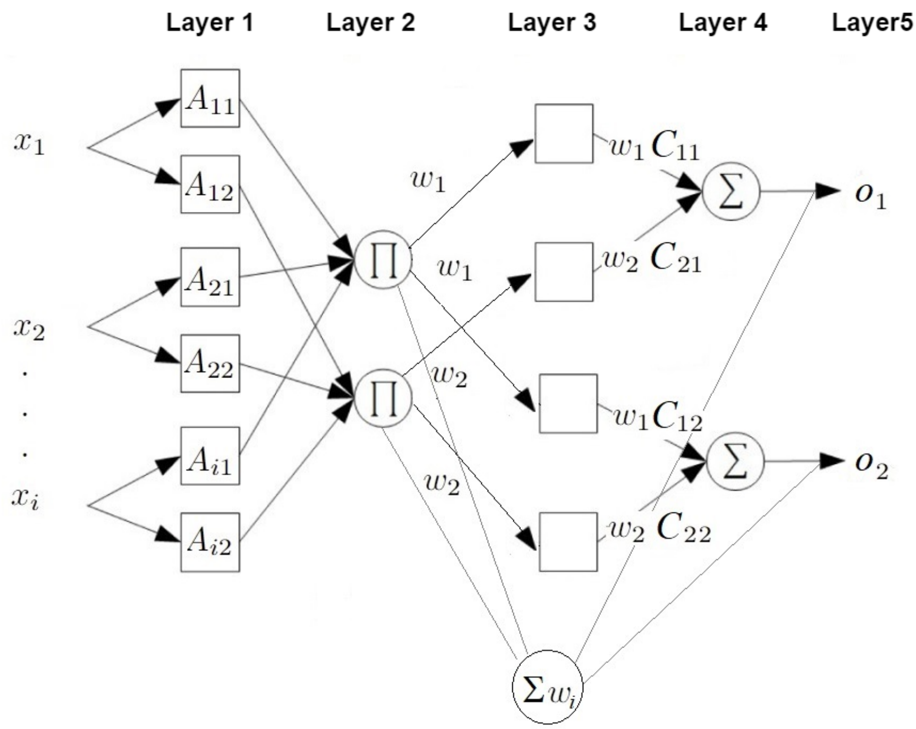

Figure 4. A CANFIS architecture [Tom+23]

with the solution

$$\theta^* = \mathbf{H}^*\mathbf{O}, \tag{3.16}$$

where $\mathbf{H}^* = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the pseudoinverse of $\mathbf{H}$. The fractional Tikhonov method represents a generalization of the standard LS method through the following minimization problem

$$\min_{\theta} \|\mathbf{H}\theta - \mathbf{O}\|_P^2 + \lambda\|\theta\|^2, \tag{3.17}$$

where $\|\theta\|_P$[5] $= (\theta^T\mathbf{P}\theta)^{\frac{1}{2}}$ and $\mathbf{P}$ is a symmetric positive semi-definite matrix defined as

$$\mathbf{P} = (\mathbf{H}^T\mathbf{H})^{\frac{\alpha-1}{2}}, \tag{3.18}$$

where $\alpha \in (0,1)$ is the fractional regularization parameter and λ a general regularization parameter ($\lambda \in R_+$).

Following [TPL22], the solution is:

$$\Theta^* = (\mathbf{M}^{\frac{\alpha+1}{2}} + \lambda\mathbf{I})^{-1}\mathbf{M}^{\frac{\alpha-1}{2}}\mathbf{H}^T\mathbf{O}, \tag{3.19}$$

where $\mathbf{M} = \mathbf{H}^T\mathbf{H}$ and $\mathbf{I}$ is an identity matrix.

In [TPL22], the accuracy of ANFIS with fractional regularization was formally proved. It is possible to mimic the same proofs to prove the accuracy in the case of multiple outputs in CANFIS.

The jth rule that is possible to extract from CANFIS is

$$\text{IF } x_1 \text{ is } A_{1j} \ldots \text{ AND } x_n \text{ is } A_{nj} \text{ THEN } \{C_{j1}, \ldots, C_{jp}\}.$$

It is important to note that in order to ensure interpretability in fuzzy systems, there should be a small number of rules allowing easy reading and understanding. This also means that the number of terms should be small enough (e.g. 3-5) [Men13]. This has motivated the adoption of fractional regularization in ANFIS first [TPL22] and then in CANFIS.

3.4. Automated Machine Learning (AutoML)

Automated Machine Learning (AutoML) is an emerging field that aims to simplify the application of machine learning models by automating the process of model selection and hyperparameter optimization. The primary goal of AutoML is to make machine learning accessible to non-experts and to save time for experienced practitioners. One of the key challenges in AutoML is the Combined Algorithm Selection and Hyperparameter Optimization (CASH) problem, which involves simultaneously selecting the best machine learning algorithm and its optimal hyperparameters for a given dataset.

[5] Standard $l2 - norm$ is used for $\|\theta\|$

Machine learning algorithms often come with numerous hyperparameters that significantly influence their performance. Additionally, different algorithms can vary greatly in their suitability for different types of data and problems. The combination of these factors results in an enormous search space, making manual optimization impractical. The CASH problem is a formal approach to address this complexity by treating algorithm selection and hyperparameter tuning as a unified optimization problem [Tho+13].

The CASH Problem: The term CASH was introduced by Thornton et al. [Tho+13] to describe the task of finding the best combination of learning algorithm and hyperparameters to minimize the error on a given dataset. The CASH problem is formally defined as follows:

$$A^*_{\lambda^*} \in \underset{A^{(j)} \in A, \lambda \in \Lambda^{(j)}}{\operatorname{argmin}} \frac{1}{k} \sum_{i=1}^{k} L(A^{(j)}_{\lambda}, D^{(i)}_{\text{train}}, D^{(i)}_{\text{valid}}) \tag{3.20}$$

where A denotes the set of all possible algorithms, $\Lambda^{(j)}$ represents the hyperparameter space for algorithm $A^{(j)}$, and L is the loss function used to measure performance. The objective is to identify the algorithm A^* and hyperparameters λ^* that achieve the lowest cross-validation error using the training set $D^{(i)}_{\text{train}}$ and validation set $D^{(i)}_{\text{valid}}$. Usually, k-fold cross-validation is used which splits the training data into k equal sized partitions $D^{(1)}_{\text{valid}}, \ldots, D^{(k)}_{\text{valid}}$, and sets $D^{(i)}_{\text{train}} = D \backslash D^{(i)}_{\text{valid}}$ for $i = 1, \ldots, k$.

The complexity of the CASH problem arises from the vast and high-dimensional search space, which includes a wide range of algorithms and their respective hyperparameters. This search space is characterized by both continuous and categorical variables, often with hierarchical dependencies. Traditional approaches that treat algorithm selection and hyperparameter optimization as separate tasks fail to address the complexity of the combined problem effectively.

Interpretability: Interpretability in machine learning is the ability to explain or to present in understandable terms to a human the workings and decisions of a machine learning model. As machine learning systems are increasingly employed in critical domains such as healthcare, finance, criminal justice, and autonomous driving, the need for interpretability has become paramount. Without interpretability, users and stakeholders may find it challenging to trust and adopt models, especially when these models are complex and their decisions have significant consequences [Lip18].

The interpretability of machine learning models can be broadly categorized into two approaches: *intrinsic interpretability* and *post-hoc interpretability*. Intrinsically interpretable models are those that are inherently understandable by humans, such as linear regression models, decision trees, and rule-based classifiers. These models provide clear insights into how input features are used to generate predictions, making them more transparent but sometimes at the cost of reduced predictive accuracy, especially on complex tasks [Rud19].

On the other hand, post-hoc interpretability techniques are applied to complex, often black-box models, such as deep neural networks, ensemble models, and support vector machines. These methods aim to provide explanations after the model has made its predictions. Techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) are widely used for this purpose. LIME works by approximating the model locally with an interpretable model around the prediction of interest, allowing users to understand what features contributed most to a particular decision [RSG16]. SHAP values, grounded in cooperative game theory, provide a unified measure of feature importance across different predictions, helping to explain both individual predictions and the model as a whole [LL17].

Moreover, interpretability is closely linked to issues of fairness, accountability, and transparency in AI systems. The ability to interpret model decisions allows stakeholders to detect and correct biases, ensuring that models do not perpetuate or exacerbate unfair outcomes. In regulated industries, interpretability is not only desirable but often legally required to ensure that decisions can be audited and justified [DK17].

In summary, while there is often a trade-off between interpretability and accuracy, the development of interpretability techniques is crucial for the responsible deployment of machine learning systems, enabling users to trust and effectively manage these powerful tools.

3.4.1. TPOT: A Tree-based Pipeline Optimization Tool for Automating Machine Learning

Machine learning is commonly described as a "field of study that gives computers the ability to learn without being explicitly programmed" [Sim13]. Despite this common claim, machine learning practitioners know that designing effective machine learning pipelines is often a tedious endeavor and typically requires considerable experience with machine learning algorithms, expert knowledge of the problem domain, and brute force search to accomplish [Ols+16b]. Thus, contrary to what machine learning enthusiasts would have us believe, machine learning still requires considerable explicit programming.

In response to this challenge, several automated machine learning methods have been developed over the years [HLS15]. Tree-based Pipeline Optimization Tool (TPOT) was developed by [Ols+16a] that automatically designs and optimizes machine learning pipelines for a given problem domain, without any need for human intervention. In short, TPOT optimizes machine learning pipelines using a version of genetic programming (GP), a well-known evolutionary computation technique for automatically constructing computer programs [Ban+98].

Machine Learning Pipeline Operators: At its core, TPOT is a wrapper for the Python machine learning package, scikit-learn [Ped+12]. Thus, each machine learning pipeline operator (i.e., GP primitive) in TPOT corresponds to a machine learning algorithm, such as a supervised classification model or standard feature scaler. All implementations of the machine learning algorithms listed in Table 1 are from scikit-learn (except XGBoost).

Constructing Tree-based Pipelines: To combine these operators into a machine learning pipeline, they are treated as GP primitives and construct GP trees from them. Figure 5 shows an example tree-based pipeline, where two copies of the data set are provided to the pipeline, modified in a successive manner by each operator, combined into a single data set, and finally used to make classifications. Because all operators receive a data set as input and return the modified data set as output, it is possible to construct arbitrarily shaped machine learning pipelines that can act on multiple copies of the data set. Thus, GP trees provide an inherently flexible representation of machine learning pipelines. In order for these tree-based pipelines to operate, three additional variables are stored for each record in the data set. The "class" variable indicates the true label for each record and is used when evaluating the accuracy of each pipeline. The "guess" variable indicates the pipeline's latest guess for each record, where the classifications from the last classification operator in the pipeline are stored as the "guess". Finally, the "group" variable indicates whether the record is to be used as a part of the internal training or testing set, such that the tree-based pipelines are only trained on the training data and evaluated on the testing data [OM19]).

Optimizing Tree-based Pipelines: To automatically generate and optimize these tree-based pipelines, a GP algorithm [Ban+98] is implemented in the Python package DEAP [For+12]. The TPOT GP algorithm follows a standard GP process: To begin, the GP algorithm generates 100 random tree-based pipelines and evaluates their balanced cross-validation accuracy on the data set. For every generation of the GP algorithm, the algorithm selects the top 20 pipelines in the population according to the NSGA-II selection scheme [Deb+02], where pipelines are selected to simultaneously maximize classification accuracy on the data set while minimizing the number of operators in the pipeline. Each of the top 20 selected pipelines produces five copies (i.e., offspring) into the next generation's population, 5% of those offspring cross over with another offspring using one-point crossover, then 90% of the remaining unaffected offspring are randomly changed by a point, insert, or shrink mutation (1/3 chance of each). Every generation, the algorithm updates a Pareto front of the non-dominated solutions [Deb+02] discovered at any point in the GP run. The algorithm repeats this evaluate-select-crossover-mutate process for 100 generations—adding and tuning pipeline operators that improve classification accuracy and pruning operators that degrade classification accuracy—at which point the algorithm selects the highest-accuracy pipeline from the Pareto front as the representative "best" pipeline from the run.

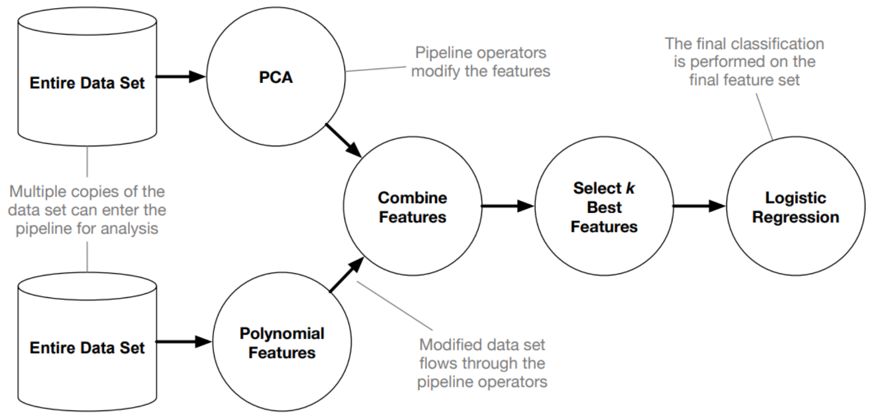

Figure 5. An example tree-based pipeline from TPOT (taken from [OM19])

Section	**Operators**	**Description**
Supervised Classification Operators	DecisionTree, RandomForest, eXtreme Gradient Boosting Classifier (from XGBoost [CG16]), LogisticRegression, and KNearestNeighborClassifier.	Classification operators store the classifier's predictions as a new feature as well as the classification for the pipeline.
Feature Preprocessing Operators	StandardScaler, RobustScaler, MinMaxScaler, MaxAbsScaler, RandomizedPCA [MRT11], Binarizer, and PolynomialFeatures.	Preprocessing operators modify the data set in some way and return the modified data set.
Feature Selection Operators	VarianceThreshold, SelectKBest, SelectPercentile, SelectFwe, and Recursive Feature Elimination (RFE).	Feature selection operators reduce the number of features in the data set using some criteria and return the modified data set.

Table 1. Machine learning pipeline operators in TPOT

4. EXPERIMENTAL RESULTS

In this section, we will provide details regarding the numerical experiments and results. Some common settings for case studies are also presented. Detailed explanations can be found under the respective case study.

Preprocessing:. A preprocessing step to normalize data was applied to the experiments for case studies I, II, and III. The data was normalized using min-max normalization, ensuring the data range is $[0,1]$. The general formula for min-max normalization is:

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)}$$

where x' is the normalized value and x is the original value where $x \in X$.

Performance metric:. Root Mean Squared Error (RMSE) was used as the performance metric for experiments in case studies I, II, and III, because we were dealing with regression problems in those case studies. Case study IV introduces performance metrics related to the classification problem.

The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where

- y_i represents the actual values,
- $\hat{y}_i$ represents the predicted values,
- n is the total number of data points.

Choice of membership function:. The choice of membership functions for ANFIS-T and CANFIS-T was made after conducting preliminary experiments specific to the case study. Also, several experiments were conducted with varying numbers of rules and types of membership function and only the results with the best scheme are presented here. The commonly used MF was Gaussian membership function.

Choice of Fractional Regularization parameters α, λ:. There are two parameters that controls the regularization in ANFIS-T and CANFIS-T. α controls the *fractional* part which makes the learning part more smooth and λ is the general regularization parameter. The range for α was chosen between $\alpha \in \{0.1, 0.2, \ldots, 0.9, 1\}$. It was also considered that the value of α should be less than 1, otherwise the *fractional* part from the regularization is removed. For λ, there were different values chosen $\lambda \in \{10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3}\}$. It is also to be noted that if the value of λ is 0, then the standard regularization does not apply.

Choice of AutoML framework:. There are several AutoML frameworks available (see Table 2). The choice of AutoML framework was made based on the performance of the tabular dataset and also based on the literature. For example, a benchmark study was conducted in [Eld+24], and it was observed that TPOT performed really well on most of the datasets. We also conducted several experiments and TPOT provided the pipelines that showed overall better performance.

Framework	Popularity (stars on Github)	Optimization technique
AutoWeka	314	Bayesian optimization
AutoSKlearn	7k	Bayesian optimization
TPOT	9.1k	Evolutionary optimization
Recipe	50	Grammar based genetic algorithm
ATM	522	Distributed Random search & Tree-Parzen estimators
SmartML	23	Bayesian optimization

Table 2. AutoML frameworks available [Eld+24]

Hardware:. For all experiments, the approaches were implemented in Scilab and executed on a Dell XPS with a Core-i9 processor (2.4GHz, 16 CPUs) and 32GB of RAM.

4.1. Case study I: Soil organic carbon prediction

4.1.1. Preliminary Experiments

Dataset: Datasets used in the preliminary experiments are publicly available and a general overview of the datasets is presented in Table 3.

Dataset	Abbreviation	Instances	Features
Airfoil Noise [BPM14]	AIR	1503	5
Abalone [Nas+95]	ABA	4176	8
Pole Telecommunications [Alc+10]	POLE	14998	26
Pumadyn [Alc+10]	PUMA	8192	32

Table 3. Benchmark Datasets for Preliminary Experiments [Muh+22]

Experiment Details: Initial experiments were conducted on different publicly available datasets (see Table 3), for an initial comparison between the two approaches. For a fair comparison against the results in [TPL22], a 5-fold cross-validation was used for the AIR and ABA datasets, and a 10-fold cross-validation for the remaining datasets in Table 3. ANFIS-T achieved the highest accuracy across all datasets, while FT-rELM demonstrated faster training time. The results are shown in Table 4.

Dataset	Approach	Rules	RMSE	Time (s)
ABA	ANFIS-T ($\gamma = 1$, $\alpha = 0.5$) [TPL22]	2	1.478 ± 0.1825	2.37
ABA	FT-rELM (L=50, $\gamma = 0.0001$, $\alpha = 0.1$)	-	2.0629 ± 0.2051	0.091
AIR	ANFIS-T ($\gamma = 0.01$, $\alpha = 0.1$) [TPL22]	4	4.9726 ± 0.5354	2.96
AIR	FT-rELM (L=20, $\gamma = 0.01$, $\alpha = 1$)	-	6.8240 ± 0.3689	0.052
POLE	ANFIS-T ($\gamma = 0.1$, $\alpha = 0.1$) [TPL22]	2	15.2477	177.68
POLE	FT-rELM (L=80, $\gamma = 0.1$, $\alpha = 0.9$)	-	41.7306	0.836
PUMA	ANFIS-T ($\gamma = 0.1$, $\alpha = 0.1$) [TPL22]	2	0.00084	93.55
PUMA	FT-rELM (L=100, $\gamma = 0.1$, $\alpha = 0.1$)	-	0.0280	0.811

Table 4. Preliminary Experiments: Comparison with Test Results from [TPL22]

4.1.2. EstSoil-EH: High-Resolution Eco-Hydrological Modeling Parameters Dataset for Estonia

For our experiments, we used the EstSoil-EH dataset [Kmo+21], which contains 8198 samples. This dataset is described in Table 5. Out of 125 features, only 6 were selected for predicting soil organic carbon (SOC), based on their importance as determined by Random Forest (RF) in [Kmo+21]. RF was preferred over Neural Networks due to its robustness to noisy data and ability to identify important predictors. The data was split into 60% training and 40% validation. The reference study reported an RMSE of 0.0295 for SOC prediction [Kmo+21].

To compare performance, we split the data into 60-40 and 70-30 training-test ratios and performed 5 and 10-fold cross-validation. The 60-40 split was used for fair comparison with the reference paper, while the 70-30 split was used to evaluate ANFIS-T's performance. The results are reported separately for both splits.

Name	Description	Type
CLAY1	Clay content	Continuous
tri_stdev	Terrain roughness index, standard deviation	Continuous
SAND1	Sand content	Continuous
ls_median	LS factor, median	Continuous
drain_pct	Area under drainage in percent	Continuous
ROCK1	Coarse fragments rock content	Continuous

Table 5. SOC Dataset Attributes [Muh+22]

4.1.3. Results

First Split (60% Training and 40% Test): Table 6 shows the approach, cross-validation, RMSE, and training time in seconds. ANFIS-T without cross-validation has a shorter training time but higher RMSE. FT-rELM shows similar performance to ANFIS-T without cross-validation but with a shorter training time. ANFIS-T with 5-fold cross-validation has RMSE close to 10-fold cross-validation but with shorter training time. FT-rELM demonstrates significantly shorter training times than ANFIS-T across all setups.

Approach	**Cross-Validation**	**RMSE**	**Time (s)**
ANFIS-T (3 terms, $\lambda = 0.001$, $\alpha = 0.1$)	No	0.1493	2.40
ANFIS-T (2 terms, $\lambda = 0.1$, $\alpha = 1$)	5-fold	0.0179 ± 0.0020	4.37
ANFIS-T (2 terms, $\lambda = 0.1$, $\alpha = 1$)	10-fold	0.0164 ± 0.0026	9.99
FT-rELM ($\lambda = 0.1$, $\alpha = 0.9$)	No	0.1555	0.0012
FT-rELM ($\lambda = 0.001$, $\alpha = 0.1$)	5-fold	0.0959 ± 0.0614	0.038
FT-rELM ($\lambda = 1000$, $\alpha = 0.3$)	10-fold	0.0897 ± 0.0447	0.062

Table 6. Results for First Split [Muh+22]

The variation of RMSE with different λ and α values for ANFIS-T and FT-rELM is experimented. For ANFIS-T, RMSE increases with higher λ and α values when there are 3 terms without cross-validation. When there are 2 terms with 5 and 10-fold cross-validation, RMSE decreases with larger α values, while λ has less impact. In the case of FT-rELM, RMSE fluctuates with varying λ and α, achieving the best results with smaller values.

(a) (b)

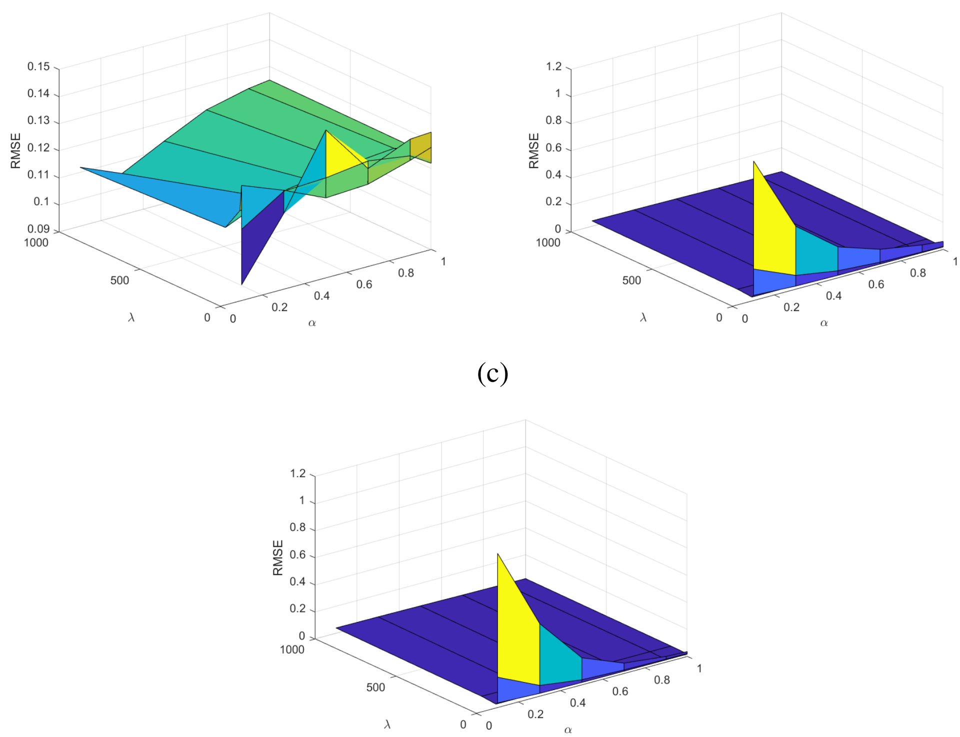

Figure 6. ANFIS-T (60-40 Split) - RMSE vs (λ, α): (a) Without Cross-Validation, (b) 5-fold Cross-Validation, (c) 10-fold Cross-Validation.

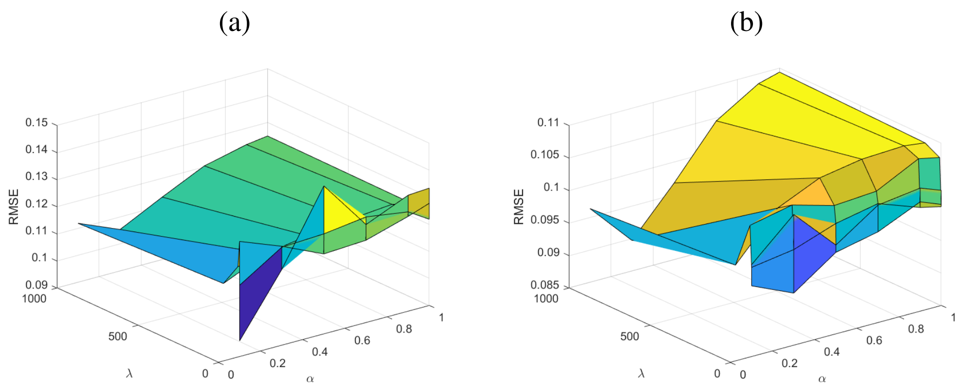

Figure 7. FT-rELM (60-40 Split) - RMSE vs (λ, α): (a) 5-fold Cross-Validation, (b) 10-fold Cross-Validation.

Second Split (70% Training and 30% Test): Table 19 presents the results for the second split. ANFIS-T with 5 and 10-fold cross-validation again shows similar RMSE, but the training time difference increases. Both ANFIS-T and FT-rELM without cross-validation perform poorly in terms of RMSE compared to their cross-validation results. There is little difference in RMSE and training time for FT-rELM with 5 and 10-fold cross-validation.

Approach	Cross-Validation	RMSE	Time (s)
ANFIS-T (3 terms, $\lambda = 0.001$, $\alpha = 0.1$)	No	0.1688	2.61
ANFIS-T (2 terms, $\lambda = 10$, $\alpha = 0.1$)	5-fold	0.0184 ± 0.0016	5.60
ANFIS-T (2 terms, $\lambda = 0.1$, $\alpha = 0.5$)	10-fold	0.0179 ± 0.0020	12.08
FT-rELM ($\lambda = 0.1$, $\alpha = 0.9$)	No	0.1787	0.0094
FT-rELM ($\lambda = 0.001$, $\alpha = 0.3$)	5-fold	0.0943 ± 0.0217	0.025
FT-rELM ($\lambda = 0.001$, $\alpha = 1$)	10-fold	0.0835 ± 0.0170	0.034

Table 7. Results for Second Split [Muh+22]

For ANFIS-T with 3 terms and no cross-validation, RMSE increases with higher λ values and improves slightly with higher α values. For ANFIS-T with 2 terms and 5 or 10-fold cross-validation, RMSE rises with larger λ values and smaller α values. In the case of FT-rELM, RMSE remains mostly unchanged but shows a slight decrease with smaller α values.

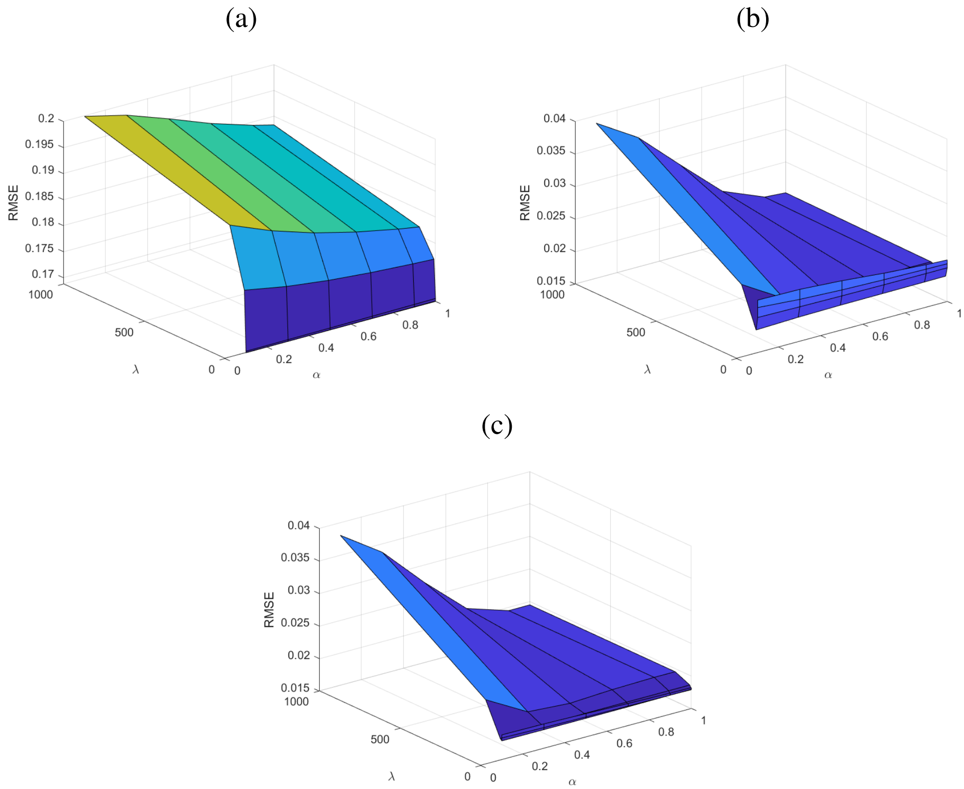

Figure 8. ANFIS-T (60-40 Split) - RMSE vs (λ, α): (a) Without Cross-Validation, (b) 5-fold Cross-Validation, (c) 10-fold Cross-Validation.

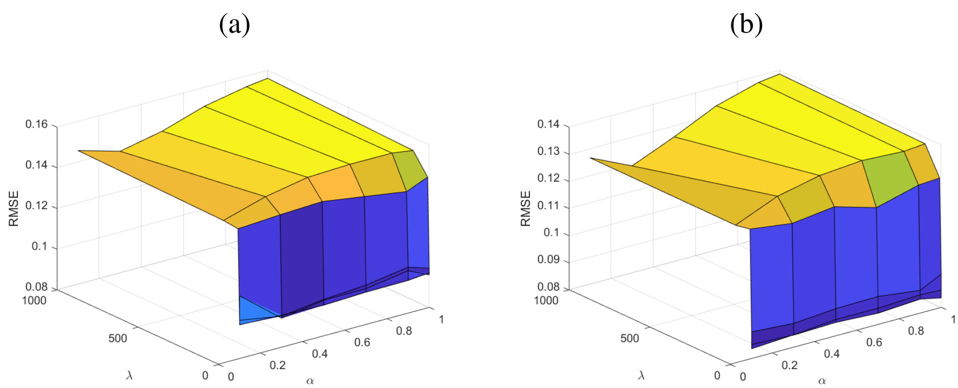

Figure 9. FT-rELM (70-30 Split) - RMSE vs (λ, α): (a) 5-fold Cross-Validation, (b) 10-fold Cross-Validation.

4.1.4. Dataset from [Kau+21; Kau+15]

We also conducted experiments on the dataset from [Kau+21; Kau+15] using 10-fold cross-validation. This dataset contains samples from 2008-2018, totaling 1460 samples, split into 70% training (1022 samples) and 30% validation (438 samples). The data includes information on nutrients, yield, seed nutrient content, and carbon content, normalized using min-max normalization. The results are shown in Table 8. ANFIS-T once again achieved the best accuracy, while FT-rELM was the fastest in training.

Approach	RMSE	Time (s)
ANFIS-T (2 terms, $\lambda = 0.01$, $\alpha = 0.1$)	0.0027 ± 0.0002	2.36
FT-rELM ($\lambda = 0.001$, $\alpha = 0.5$)	0.0512 ± 0.0056	0.035

Table 8. Results for Dataset from [Kau+21; Kau+15]

4.1.5. Discussion

The results obtained from this study provide valuable insights into the performance of machine learning models, particularly ANFIS-T (Adaptive Neuro-Fuzzy Inference System with fractional Tikhonov regularization) and FT-rELM (Fractional Tikhonov regularized Extreme Learning Machine), in predicting Soil Organic Carbon (SOC). Both models were evaluated across different experimental setups, revealing important implications for their application in environmental modeling and soil science. Fractional Tikhonov regularization helps to prevent over-fitting in extreme learning machines and helps to tackle ill-posed problems in ANFIS-T. Since the optimization part of the ANFIS-T includes calculations that involve inverse matrix, and that inverse matrix can lead to an ill-posed problem. So fractional Tikhonov regularization helps to prevent that and the fractional part of the regularization helps to more smooth the learning process.

Performance of ANFIS-T and FT-rELM: The comparative analysis between ANFIS-T and FT-rELM demonstrated that both models have distinct strengths in SOC prediction. ANFIS-T, with its inherent capability to capture non-linear relationships and handle uncertain information, exhibited robust performance across various datasets. The integration of fractional Tikhonov regularization further enhanced its predictive accuracy by mitigating overfitting, a common issue in machine learning models dealing with complex environmental data. On the other hand, FT-rELM showcased its advantages in terms of computational efficiency and ease of implementation. Its ability to produce accurate predictions with reduced computational cost makes it a viable option for large-scale SOC prediction tasks, particularly when rapid assessments are required.

Implications for SOC Prediction: The results underline the importance of model selection in SOC prediction. While ANFIS-T may be preferable in scenarios requiring high accuracy and the ability to model complex interactions, FT-rELM provides a suitable alternative where computational resources are limited or where rapid predictions are necessary. This dual approach offers flexibility in addressing various challenges in SOC modeling, depending on the specific needs of the study or project.

Moreover, the success of fractional Tikhonov regularization in both models highlights the significance of regularization techniques in improving model performance. By controlling the complexity of the models, this method effectively balances the trade-off between bias and variance, leading to more reliable predictions. This finding suggests that incorporating similar regularization techniques into other machine learning models could potentially enhance their performance in environmental prediction tasks.

Limitations and Future Work: Despite the promising results, there are limitations to this study that warrant further investigation. The datasets used in this study were limited to Estonian soil conditions, which may affect the generalizability of the findings to other regions with different soil properties and climatic conditions. Future research should aim to apply these models to a more diverse set of geographical locations to validate their effectiveness on a global scale.

Additionally, while the study focused on comparing ANFIS-T and FT-rELM, there are numerous other machine learning algorithms that could potentially offer better performance in SOC prediction. Future studies should explore a broader range of models, including deep learning approaches, to identify the most optimal techniques for different environmental conditions.

Finally, this study primarily focused on model accuracy and computational efficiency. Other factors, such as model interpretability and ease of use, are also crucial in practical applications. Future work should consider these aspects to ensure that the selected models are not only accurate but also accessible to a wide range of users, including policymakers and land managers.

Conclusion: This study contributes to the growing body of knowledge on the application of machine learning in environmental science, particularly in the prediction of Soil Organic Carbon. The findings demonstrate that both ANFIS-T and FT-rELM are effective tools for SOC prediction, with each model offering distinct advantages. The successful application of fractional Tikhonov regularization underscores the importance of regularization techniques in improving model performance. However, further research is needed to generalize these findings to other regions and explore additional machine learning approaches. By addressing these challenges, future studies can help refine SOC prediction models, ultimately contributing to more effective environmental management and climate change mitigation strategies.

4.2. Case study II: Evapotranspiration prediction

The numerical experiments were conducted in two groups, each using a different dataset. For all experiments, the Root Mean Squared Error (RMSE) was used as the error measure, and 5-fold cross-validation was employed. For ANFIS-T, various values of the regularization parameters were used: $\alpha \in \{0.1, 0.2, \ldots, 0.9, 1\}$ and $\lambda \in \{10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3}\}$. The experiments were implemented in Scilab for ANFIS-T and SVR, and in Matlab for DT.

4.2.1. Datasets

First Dataset: The first dataset was obtained from *data.world*, specifically *data for the bushland, Texas*, provided by the US Department of Agriculture, covering the period from 1996 to 1998. This dataset was formed by merging multiple files containing daily data for a single year. After removing null values, the dataset comprised 1061 samples.

The data is detailed in Table 9.

Name	Unit	Type	Range
Precipitation	mm	Continuous	[0, 114.675]
Irrigation	mm	Continuous	[0, 44.370]
Relative Humidity	%	Continuous	[14.800, 100]
Air Temperature	Celsius	Continuous	[-18.800, 27.600]
Wind Speed	m/s	Continuous	[0, 11.790]
Air Pressure	kPa	Continuous	[87.100, 90.430]

Table 9. Attributes of the First Dataset [MTL23]

Second Dataset: The second dataset was the same used in the first group of experiments in [Gar+22]. The attributes are listed in Table 10. This dataset comprises 216 samples. It is worth noting that the attributes used in our experiments are more comprehensive than those used in [Gar+22]. The best results for the second dataset are shown in Table 13.

Name	Unit	Type	Range
Soil Temperature	Celsius	Continuous	[13.61, 27.80]
Soil Permeability	dS/s	Continuous	[7.04, 13.25]
Soil Conductivity	dS/s	Continuous	[0, 0.158]
Air Temperature	Celsius	Continuous	[12.2, 40.6]
Wind Velocity	m/s	Continuous	[0.57, 3.66]
Vapour Pressure	kPa	Continuous	[0.63, 2.59]
Relative Humidity	%	Continuous	[23.03, 88.28]
Air Pressure	kPa	Continuous	[96.23, 98.46]
Solar Radiance	μmol/m^2/s	Continuous	[10.94, 949.17]
Precipitation	mm	Continuous	[0, 8.2]

Table 10. Attributes of the Second Dataset [MTL23]

4.2.2. Results

First Dataset: The best results obtained for the first dataset are shown in Table 6. Table 4.2.2 lists the parameters used for DT and nu-SVR. These parameters were also used for the experiments with the second dataset. The settings for ANFIS-T are listed in Table 17 as they vary depending on the dataset.

The DT has the highest RMSE among the methods, though not significantly different from the others. The most notable difference is in the training time. DT takes the longest time to train the data, while nu-SVR and ANFIS-T require significantly less training time. ANFIS-T and nu-SVR have almost identical RMSE, but nu-SVR outperforms in terms of training time. However, ANFIS-T offers the best interpretability due to its fewer rules.

Approach	**Rules**	**RMSE**	**Training Time (s)**
DT	221	0.1536	1.87
ANFIS-T (2 terms, $\lambda = 0.1$, $\alpha = 0.9$)	2	0.1294	0.648
nu-SVR	-	0.1269	0.199

Table 11. Cross-Validated Results for the First Dataset [MTL23]

Approach	**Parameter Name/Description**	**Parameter Value**
DT	Minimum Number of Branch Node Observations	10
	Maximal Number of Decision Splits	1
	Minimum Number of Leaf Node Observations	1
nu-SVR	ν	0.5
	Cost Penalty Parameter	1
	Kernel Type	Linear

Table 12. Model Parameters [MTL23]

Second Dataset: The second dataset has about five times fewer samples than the first dataset. The best results are shown in Table 13.

For all methods, the training time is notably shorter compared to the first dataset. Nu-SVR has the worst performance with an RMSE of 2.8267, but it has the shortest training time. Although ANFIS-T requires more training time than nu-SVR, its accuracy is better. DT performs the best in terms of accuracy, but has the longest training time. Once again, ANFIS-T proves to be the most interpretable approach due to the limited number of rules.

Approach	Rules	RMSE	Training Time (s)
DT	45	1.0966	0.665
ANFIS-T (5 terms, $\lambda = 0.001$, $\alpha = 0.1$)	5	1.7643	0.249
nu-SVR	-	2.8267	0.022

Table 13. Cross-Validated Results for the Second Dataset [MTL23]

4.2.3. Discussion

The study compared the performance of three approaches—Decision Trees (DTs), Adaptive Network-based Fuzzy Inference System with fractional Tikhonov regularization (ANFIS-T), and Support Vector Regression (SVR)—for evapotranspiration prediction, focusing on interpretability, accuracy, and training time. The results of this study offer significant insights into the trade-offs between these metrics, particularly in the context of precision agriculture.

Interpretability and Model Complexity: Interpretable models are essential in domains such as agriculture, where decision-makers need to understand and trust the predictions generated by machine learning models. ANFIS-T demonstrated the highest level of interpretability, particularly due to the limited number of rules required to make predictions. For the first dataset, ANFIS-T needed only two rules, while for the second dataset, it required five rules. This simplicity in rule formulation ensures that the model's decisions can be easily understood and validated by domain experts, making it a preferable choice in scenarios where interpretability is a critical requirement.

In contrast, Decision Trees, while generally considered interpretable, showed higher complexity with a larger number of rules (221 for the first dataset and 45 for the second dataset). The increased complexity, particularly in the first dataset, potentially diminishes the clarity of the model's decision-making process, making it harder for stakeholders to interpret the results. Despite its complexity, the Decision Tree approach is still valuable for its relatively straightforward rule-based predictions, but it requires careful consideration of the trade-off between interpretability and accuracy.

Accuracy of Predictions: Accuracy is a crucial metric for evaluating the performance of predictive models. The study found that for the first dataset, the ANFIS-T and SVR approaches provided comparable accuracy (with RMSE values of 0.1294 and 0.1268, respectively), slightly outperforming the Decision Tree model (RMSE of 0.1536). This suggests that while ANFIS-T and SVR are both capable of delivering precise predictions, ANFIS-T does so with the added benefit of enhanced interpretability.

For the second dataset, the Decision Tree model slightly outperformed ANFIS-T and SVR in terms of accuracy (RMSE of 1.0966 for DT compared to 1.7643 for ANFIS-T and 2.8267 for SVR). The relatively higher accuracy of Decision Trees in this scenario suggests that, under certain conditions, more complex models may capture nuances in the data that simpler, more interpretable models like ANFIS-T might miss.

Training Time Considerations: The training time is another critical factor, especially in applications requiring real-time or near-real-time predictions. In this study, SVR consistently showed the shortest training times across both datasets, making it a suitable choice for scenarios where computational efficiency is a priority. ANFIS-T, while not as fast as SVR, still demonstrated reasonable training times, particularly when considering its interpretability advantage. On the other hand, Decision Trees required significantly longer training times, particularly for the first dataset, where the tree complexity was high.

The longer training times associated with Decision Trees might be a drawback in scenarios where rapid model deployment and updates are necessary. However, in settings where the model's interpretability and accuracy take precedence over training time, Decision Trees might still be a viable option.

Conclusion: The findings from this study highlight the importance of considering the specific requirements of the application when selecting a predictive model for evapotranspiration. ANFIS-T's balance between accuracy, interpretability, and training time makes it particularly appealing for precision agriculture applications where transparency and trust in model predictions are paramount. However, in scenarios where model accuracy and efficiency are more critical, and interpretability can be compromised, SVR might be the preferred choice.

Future research could explore the application of these models to larger and more diverse datasets, as well as their performance in operational settings. Furthermore, further investigation of hybrid models that combine the strengths of these approaches could provide even more robust solutions for the prediction of evapotranspiration.

4.3. Case study III: Ex-ante Life Cycle Assessment for Wheat Production

Original data were generated with specified averages and standard deviations for the two case studies and then normalized using Min-Max normalization within the range [0, 1].

In all experiments, 2-fold cross-validation was employed to avoid biased results due to test data selection, despite no cross-validation being used in the main references.

For CANFIS-T, various values of the regularization parameters were used: $\alpha \in \{0.1, 0.2, \ldots, 0.9, 1\}$ and $\lambda \in \{10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3}\}$. For CANFIS-T, e-ANFIS, and RBFN, the Gaussian function was adopted. The numerical experiments were performed with Scilab for CANFIS-T, and Matlab for all other approaches.

RMSE was used as an accuracy measure and training time was used to gauge computational effort. The energy consumption (EC) was measured using Intel Power Gadget 3.6, considering both CPU and DRAM energy consumption [Fer+21; Per+17; Hen+22].

4.3.1. Preliminary experiments

In preliminary experiments, synthetic data was generated randomly within the range [0, 1]. Two datasets were created, each consisting of 2000 samples. The first dataset has 11 attributes and 3 targets, while the second dataset has 19 attributes and 9 targets. These preliminary experiments aimed to observe how accuracy and complexity of the approaches change with an increasing number of attributes and targets.

The e-ANFIS scheme for the first data set is illustrated in Figure 10, which consists of 3 ANFIS models. For the second dataset, there are two e-ANFIS schemes, each with different input nodes. The results for the first and second datasets are presented in Tables 14 and 15, respectively. CANFIS-T and DT exhibit similar accuracy, although CANFIS-T has fewer rules. RBFN has the shortest training time but lower accuracy, while e-ANFIS performs the worst in all aspects.

Approach	Rules	RMSE	Training Time (s)	Energy (J)
CANFIS-T ($\lambda = 1$, $\alpha = 0.9$)	3	0.2914 ± 0.00141	0.48	6.384
e-ANFIS	10	0.5785 ± 0.00989	168.23	2115.221
DT	6	0.2934 ± 0.00211	0.63	8.410
RBFN	20	0.5918 ± 0.11623	0.24	3.112

Table 14. Preliminary study, first dataset. Cross-validated test results [Tom+23]

Approach	Rules	RMSE	Training Time (s)	Energy (J)
CANFIS-T ($\lambda = 10$, $\alpha = 0.9$)	3	0.2908 ± 0.00043	1.14	11.742
e-ANFIS	85	0.6998 ± 0.00835	1995.26	21947.863
DT	18	0.2909 ± 0.00165	1.59	19.668
RBFN	45	0.6021 ± 0.02784	0.85	11.22

Table 15. Preliminary study, second dataset. Cross-validated test results [Tom+23]

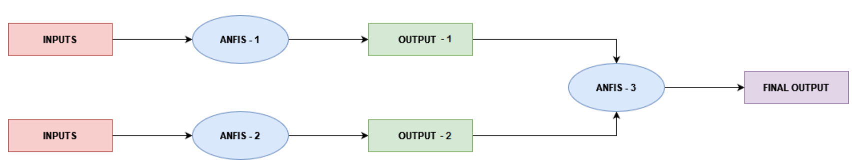

Figure 10. An e-ANFIS scheme [Tom+23]

4.3.2. The case of wheat production

This case study is based on [Gha+22], where e-ANFIS was used to predict the GWP indicator for ex-ante LCA of wheat production. This methodology leverages

Approach	Rules	RMSE	Training Time (s)
CANFIS-T ($\lambda = 0.01$, $\alpha = 0.9$)	2	0.1373 ± 0.00122	0.98
e-ANFIS	2+2+1	0.2546 ± 0.00734	173.22
DT	6	0.1372 ± 0.00212	1.41
RBFN	9	0.5116 ± 0.04231	0.89

Table 17. First case study. Cross-validated test results [Tom+23]

prior knowledge to estimate LCA indicators early in process development, using machine learning approaches [KPK21]. The dataset consists of 5000 samples with attributes derived from [Gha+22], listed in Table 16.

The e-ANFIS scheme used in our experiments is depicted in Figure 10. The inputs and outputs for each ANFIS model are described, and the results are summarized in Table 17 and Figure 11. The average energy consumption is shown in Figure 13a. CANFIS-T and DT have similar accuracy, but CANFIS-T has slightly better training time and energy consumption. RBFN has the shortest training time and lower energy consumption but worse accuracy.

Variable	Description	Average	Standard Deviation
x_1	Field operations	1373.75	215.62
x_2	Transport	5726.73	750.37
x_3	Nitrogen	5851.63	2528.76
x_4	Phosphate	1001.86	433.82
x_5	Manure	624.00	653.85
x_6	Biocides	285.12	80.10
x_7	Medium Voltage	23534.59	6646.94
y_1	Wheat Grain	77723.80	14376.67
y_2	Wheat Straw	39683.33	9401.63
y_3	GWP	624.29	129.93

Table 16. Input and output variables (x_i and y_j respectively, with $i = 1,\ldots,7$ and $j = 1,2,3$). The unit for GWP is kg CO_2. The unit for the rest of the attributes is MJ ha^{-1}. [Tom+23]

4.3.3. Additional case study: SCOR-based supply chain performance

Another case study deals with an adaptation of the study on SCs' performance reported in [LC20], based on ML and SCOR, since this could be adapted to the agri-food context. In this model, seven ANFIS schemes were used to estimate level-1 metrics based on level-2 metrics. We extended this model to include CO2 emissions, which were assumed proportional to energy consumption [He+19]. The dataset consists of 1000 instances with input and output details provided in Table 20.

We partially followed the scheme in [LC20], considering additional outputs like CO2 emissions. The input-output details of our e-ANFIS-based scheme are reported in Table 18.

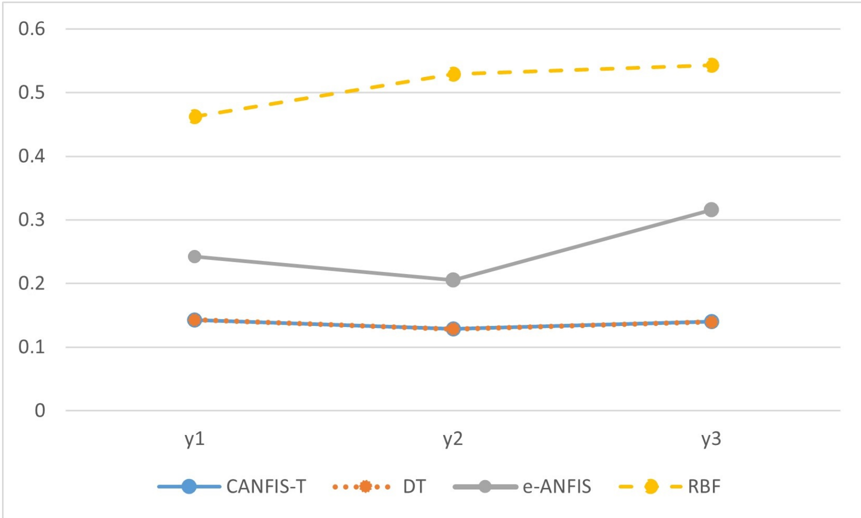

Figure 11. First case study. Cross-validated test results per output variable [Tom+23]

Model	**Input**	**Output**
ANFIS	x_1, x_2, x_3, x_4	y_1
ANFIS	x_5, x_6, x_7, x_8, x_9	y_2
e-ANFIS I/ANFIS-1	$x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}$	$y_4 = x_{25}$
e-ANFIS I/ANFIS-2	x_{22}, x_{23}, x_{24}	$y_5 = x_{26}$
e-ANFIS I/ANFIS-3	x_{25}, x_{26}, x_{27}	y_6
e-ANFIS II/ANFIS-1	$x_{10}, x_{11}, x_{12}, x_{13}$	$y_3 = x_{30}$
e-ANFIS II/ANFIS-2	x_{28}, x_{29}, x_{30}	$y_7 = x_{31}$
e-ANFIS II/ANFIS-3	x_{30}, x_{31}	y_8

Table 18. Second case study: inputs and outputs for the e-ANFIS based scheme [Tom+23]

The results are shown in Table 19 and Figure 12, with energy consumption during training shown in Figure 13b. CANFIS-T and DT have similar accuracy, training time, and energy consumption. RBFN has worse accuracy but the shortest training time and lowest energy consumption. e-ANFIS performed the worst overall.

Approach	Rules	RMSE	Training Time (s)
CANFIS-T ($\lambda = 1000$, $\alpha = 0.9$)	4	0.3133 ± 0.00343	1.91
e-ANFIS	38+122+162+542	0.8528 ± 0.25348	2295.37
DT	28	0.3129 ± 0.00577	1.8
RBFN	35	0.5805 ± 0.12556	1.05

Table 19. Second case study. Cross-validated test results [Tom+23]

Variable	Description	Universe of discourse	Unit
x_1	orders delivered in full	[0; 1]	Dimensionless
x_2	delivery performance to customer commit date	[0; 1]	Dimensionless
x_3	documentation accuracy	[0; 1]	Dimensionless
x_4	perfect condition	[0; 1]	Dimensionless
x_5	value at risk (plan)	[10,000; 100,000]	$
x_6	value at risk (source)	[50,000; 200,000]	$
x_7	value at risk (make)	[50,000; 300,000]	$
x_8	value at risk (deliver)	[20,000; 200,000]	$
x_9	value at risk (return)	[20,000; 200,000]	$
x_{10}	source cycle time	[1; 6]	Days
x_{11}	make cycle time	[1; 7]	Days
x_{12}	delivery cycle time	[1; 7]	Days
x_{13}	delivery retail cycle time	[7; 20]	Days
x_{14}	sourcing cost	[140,000; 300,000]	$
x_{15}	planning cost	[25,000; 50,000]	$
x_{16}	material landed cost	[70,000; 150,000]	$
x_{17}	production cost	[150,000; 380,000]	$
x_{18}	order management cost	[220,000; 480,000]	$
x_{19}	fulfilment cost	[45,000; 70,000]	$
x_{20}	returns cost	[50,000; 200,000]	$
x_{21}	cost of goods sold	[1,300,000; 1,900,000]	$
x_{22}	inventory	[100,000; 2,000,000]	$
x_{23}	accounts receivable	[500,000; 2,000,000]	$
x_{24}	accounts payable	[500,000; 2,000,000]	$
x_{25}	total cost to serve	[2,000,000; 3,530,000]	$
x_{26}	it is the output variable (y) of the ANFIS model 5.	[-1,400,000; 3,500,000]	$
x_{27}	supply chain revenue	[3,500,000; 10,000,000]	$
x_{28}	days sales outstanding	[25; 70]	Days
x_{29}	inventory days of supply	[27; 80]	Days
x_{30}	days payable outstanding	[30; 72]	Days
y_1	perfect order fulfilment	[0; 4]	Dimensionless
y_2	overall value at risk	[150,000; 1,000,000]	$
y_3	order fulfilment cycle time	[10; 40]	Days
y_4	total cost to serve	[2,000,000; 3,530,000]	$
y_5	denominator of the return on working capital	[-1,400,000; 3,500,000]	$
y_6	return on working capital	[-15; 100]	%
y_7	cash-to-cash cycle time	[22; 120]	Days
y_8	CO_2 emissions	[78.5, 314]	Dimensionless

Table 20. Input and output variables (x_i and y_j respectively, with $i = 1,\dots,30$ and $j = 1,\dots,8$) [Tom+23]

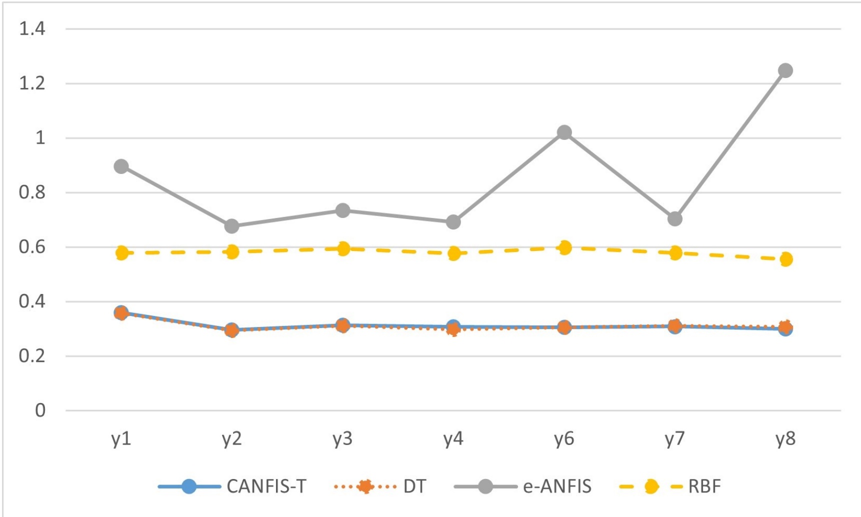

Figure 12. Second case study. Cross-validated test results per output variable [Tom+23]

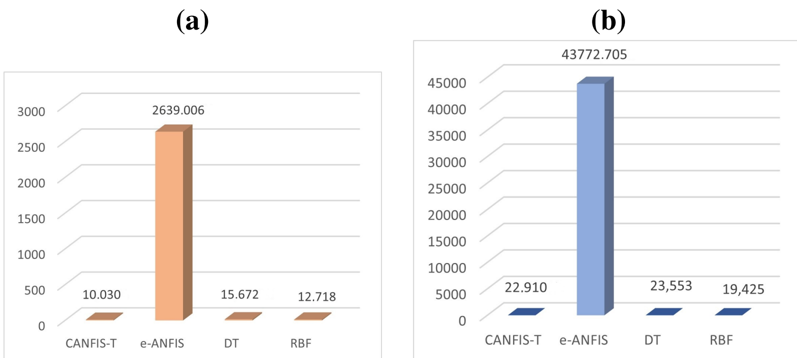

Figure 13. Energy consumption (J): (a) first case study, (b) second case study [Tom+23]

4.3.4. Discussion

The integration of machine learning techniques in the agri-food sector, particularly through the application of the Co-Active Neuro-Fuzzy Inference System with fractional regularization (CANFIS-T), presents a significant advancement in addressing the dual challenges of sustainability and interpretability. This study provides a comparative analysis of CANFIS-T against other state-of-the-art methods such as ensemble ANFIS, Radial Basis Function Networks (RBFN), and Decision Trees (DTs), using two distinct case studies rooted in ex-ante Life Cycle Assessment (LCA) and the Supply Chain Operations Reference (SCOR) model.

Key Findings and Implications: The findings from this research underscore the superiority of CANFIS-T in maintaining accuracy and interpretability, even as the complexity of input and output variables increases. This contrasts sharply with the ensemble ANFIS model, where error propagation significantly hampers performance. The ability of CANFIS-T to effectively manage multiple outputs without a loss in interpretability is particularly notable. This is crucial in the context of agri-food systems, where the implications of machine learning decisions can have profound socio-economic and environmental impacts.

Moreover, the study illustrates that while traditional methods like RBFN and DTs offer computational efficiency, they often fall short in scenarios requiring nuanced interpretation of complex data sets, a necessity in the agri-food sector. The interpretability of CANFIS-T, enabled through its fractional regularization approach, provides a meaningful way to understand and trust the outputs, aligning with the growing need for transparency in AI-driven decisions, especially in areas with significant socio-economic implications.

Contribution to the Field: This work contributes to the field by offering a robust alternative to existing models, addressing the trade-offs between accuracy, sustainability, and interpretability. The introduction of fractional regularization in the CANFIS-T model not only improves the system's performance but also ensures that the model remains accessible and interpretable to a broad range of stakeholders, from farmers to policy-makers.

The research also highlights the importance of considering the environmental costs of AI systems, an area that remains underexplored. As AI continues to be integrated into sustainability practices, the environmental footprint of these technologies must be a consideration. CANFIS-T's design inherently supports a more sustainable AI, which is both energy-efficient and interpretable, making it a viable model for future applications in sustainable agri-food systems.

Limitations and Future Research: While this study successfully demonstrates the advantages of CANFIS-T, there are limitations that warrant further exploration. The case studies, although representative, are limited in scope. Future research could extend this work by applying CANFIS-T across a broader range of agri-food systems and sustainability models. Additionally, the scalability of CANFIS-T in real-time applications remains an area for further investigation. Exploring the integration of CANFIS-T with other AI technologies and its application in large-scale agri-food supply chains could provide valuable insights.

Finally, while this study emphasizes interpretability, further work is needed to enhance the user-friendliness of these models, ensuring that stakeholders can easily interact with and understand the outputs without requiring extensive technical knowledge. Developing more intuitive interfaces and visualizations for CANFIS-T could greatly enhance its usability in practical settings.

Conclusion: In conclusion, the CANFIS-T model represents a significant step forward in the application of machine learning to sustainable agri-food systems. By balancing accuracy, interpretability, and sustainability, this approach not only advances current methodologies but also aligns with broader goals of environmental stewardship and socio-economic responsibility. Future work should build on these findings to further refine and expand the applicability of CANFIS-T in diverse agricultural contexts.

4.4. Case study IV: Rust stripes wheat disease detection

4.4.1. Datasets

Two publicly available datasets were used: RustNet [Tan+22] and Yellow-Rust-19 [Hay+23]. RustNet includes 508 images divided into two categories: disease and no disease. Specifically, there are 281 images depicting disease and 227 images without any disease. RustNet's data were collected from two experimental wheat fields in Pullman, WA, USA, during 2021. Field 1, located at Palouse Conservation Field Station, featured two winter wheat trials: one evaluating fungicides on the 'PS 279' variety and another testing stripe rust resistance in 23 winter wheat cultivars. Both trials were arranged in randomized designs with four replications, planted on November 1, 2020. Urediniospores of P. striiformis were inoculated twice to induce disease. Field 2, at Spillman Agronomy Farm, consisted of spring wheat nurseries with regular irrigation. The Lemhi 66 cultivar, highly susceptible to stripe rust, was planted in three inoculated borders and one non-infected border. Images from Field 2 were exclusively collected from the borders.

The Yellow-Rust-19 dataset comprises 15,000 images, categorized into six classes, each containing 2,500 images. Detailed class-wise information is presented in Table 21. The dataset acquisition process began with the collection of wheat leaves from fields managed by the Republic of Turkey Ministry of Agriculture and Forestry Directorate of Field Crops Central Research Institute in Ankara, Turkey. Each leaf was meticulously examined by two specialists to determine the severity level of yellow rust. Following this assessment, the leaves were photographed to form the raw dataset. To ensure quality and consistency, each image in the raw dataset underwent rigorous preprocessing to enhance clarity and remove noise. Subsequently, the Yellow-Rust-19 dataset was meticulously labeled, assigning each image to the corresponding severity level of yellow rust observed on the wheat leaves. Additionally, data augmentation techniques were applied to diversify the dataset and improve model robustness.

4.4.2. Data Preprocessing

Our methodological approach includes a detailed pre-processing phase, which consists of several key stages: initial image acquisition, grayscale image conversion, resizing, and feature extraction. During the feature extraction stage, we compute critical statistical measures such as mean, standard deviation, variance, correlation, energy, entropy, contrast, skewness, kurtosis, and homogeneity.

Class	Description
0	No visible infection on plant.
MR	Small uredia are present and surrounded by either chlorotic or necrotic areas.
MRMS	Variable-sized uredia are present; some with chlorosis, necrosis, or both.
MS	Medium-sized uredia are present and possibly surrounded by chlorotic areas.
R	Visible chlorosis or necrosis, no uredia are present.
S	Large uredia are present, generally with little or no chlorosis and no necrosis.

Table 21. Description of classes in Yellow-Rust-19 dataset [MET]

4.4.3. Context-Aware Automated Feature Engineering

Feature engineering is an important aspect of machine learning, involving the transformation of raw input data into suitable features to enhance predictive performance [WEG87]. Our approach employs Context-Aware Automated Feature Engineering (CAAFE), specifically designed for tabular datasets. CAAFE leverages a Large Language Model (LLM) to iteratively generate additional semantically meaningful features based on the dataset description [HMH23]. This method not only generates Python code for creating new features but also provides explanations for their utility.

CAAFE operates iteratively on training and validation datasets, D_{train} and D_{valid}, along with a description of the dataset context and features. It constructs a prompt with details of the dataset and the feature engineering task, which is then provided to the LLM. Based on this prompt, the LLM generates code for feature alterations. The generated code is executed on the current datasets (D_{train} and D_{valid}), transforming them into D'_{train} and D'_{valid}. An ML classifier is then trained on D'_{train} and evaluated on D'_{valid} (P'). If the performance on D'_{train} surpasses the performance on the original datasets (D_{train} and D_{valid}), the new feature is retained, and the datasets are updated. Otherwise, the feature is discarded.

The prompt given to the LLM includes semantic and descriptive information about the dataset, such as a user-generated dataset description, feature names, data types, percentage of missing values, and a sample of random rows. Additionally, the prompt includes a template for the expected form of the generated code and explanations to enhance the quality of responses. Chain-of-thought instructions guide the LLM through intermediate reasoning steps for effective code generation. By utilizing CAAFE, we integrate domain knowledge into the feature engineering process while maintaining interpretability and performance. This approach represents a significant advancement in machine learning research, providing efficient and effective means for feature generation in complex datasets.

4.4.4. Experimental Setup

Training and Test Splits. We utilized the same train and test splits as described in the referenced papers: 70% for training and 30% for testing for the RustNet dataset [Tan+22], and 90% for training and 10% for testing for the Yellow-Rust-19 dataset [Hay+23]. The specific splits are detailed in Tables 22 and 23.

Class	Train	Test	Total
disease	208	73	281
no_disease	172	55	227
Total	**380**	**128**	**508**

Table 22. Number of images in train and test split for the RustNet dataset [MET]

Class	Train	Test	Total
0	2250	250	2500
MR	2250	250	2500
MRMS	2250	250	2500
MS	2250	250	2500
R	2250	250	2500
S	2250	250	2500
Total	**13500**	**1500**	**15000**

Table 23. Number of images in train and test split for the Yellow-Rust-19 dataset [MET]

Baselines. Given the randomized nature of the experiments in [Tan+22], we conducted new experiments using the same computational setup described in the original study. Specifically, we utilized ResNet-18, maintaining the original architecture and hyperparameters. The model was initialized with pre-trained weights. ResNet-18, being a state-of-the-art technique, was also used for experiments with the Yellow-Rust-19 dataset to provide a consistent basis for comparison.

TPOT Setting. To ensure a fair comparison, an equivalent time budget was allocated for both TPOT and ResNet methodologies. For the RustNet dataset, a time budget of 20 minutes was set, while for the Yellow-Rust-19 dataset, 100 minutes were allotted. The input data for TPOT consisted of a data matrix with features extracted from the same images used for training and testing ResNet-18. TPOT's hyperparameters were configured with 10 generations and a population size of 100. The resulting pipeline was a multi-layer perceptron classifier with a learning rate of 0.01 and a regularization parameter of 0.0001, aiming to reduce overfitting and enhance generalization.

CAAFE Setting. We utilized OpenAI's language models, including GPT-3.5, within the CAAFE framework [Ope21; Ope23]. This integration enabled the generation of semantically meaningful features iteratively, enhancing feature engineering effectiveness. We conducted ten feature engineering iterations using the CAAFE framework. For evaluating the effectiveness of generated features, we used TabPFN (Tabular Predictive Functional Network) as proposed by Hollmann et al. [Hol+23].

Performance Metrics. For classification evaluation, we used Accuracy, Precision, Recall, and F1-score. These metrics provide a comprehensive assessment of model performance.

Accuracy measures the overall correctness:

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}. \tag{4.1}$$

Precision quantifies true positives relative to total positive predictions:

$$Precision = \frac{\text{True Positives}}{\text{True Positives + False Positives}}. \tag{4.2}$$

Recall calculates true positives relative to total actual positives:

$$Recall = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}. \tag{4.3}$$

F1-score, the harmonic mean of precision and recall, balances both metrics:

$$F1 - score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision + Recall}}. \tag{4.4}$$

Together, these metrics ensure a robust evaluation of classification models.

4.4.5. Results

Feature Engineering A demonstration of CAAFE using the RustNet and YellowRust-19 datasets is illustrated in Figures 14 and 15, respectively. User inputs are highlighted in blue, ML-classifier-generated data in red, and LLM-generated code is presented with syntax highlighting. The code includes comments for each generated feature, adhering to a predefined template in CAAFE's prompt. This template comprises the feature name, its utility description, the features utilized in the generated code, and sample values for these features. In the RustNet dataset, the retained features generated by CAAFE after 10 iterations include 'mean_variance_ratio', calculated as the mean divided by the variance, and 'contrast_energy_ratio', computed as the contrast divided by the energy. In the YellowRust dataset, the preserved features generated by CAAFE after 10 iterations consist of 'contrast_entropy_ratio', obtained by dividing the contrast by the entropy, 'std_skewness_product', representing the product of standard deviation and skewness, 'kurtosis_entropy_ratio', calculated as kurtosis divided by entropy, 'contrast_energy_ratio', derived from contrast divided by energy, and 'homogeneity_skewness_product', obtained by multiplying homogeneity by skewness. For the RustNet dataset, incorporating the features generated by CAAFE into TPOT improved the accuracy from 93.02% achieved using TPOT alone on the validation dataset to 95.42%, as shown in Table 24. For the Yellow-Rust dataset, integrating the generated features from CAAFE with TPOT achieved performance very comparable to that achieved using TPOT alone (see Table 24).

```
Dataset Description:
This dataset contains handcrafted statistical features from images. The images are
wheat images and there are two types of images, i.e with disease and without
disease. This dataset predicts whether an image is infected with disease or not.

Attribute Information:
- mean: mean of the image (numerical)
- std: standard deviation of the image (numerical)
- var: variance of the image (numerical)
- skewness: skewness of the image (numerical)
- entropy: entropy of the image (numerical)
- kurtosis: kurtosis of the image (numerical)
- contrast: contrast of the image (numerical)
- correlation: correlation of the image (numerical)
- energy: energy of the image (numerical)
- homogeneity: homogeneity of the image (numerical)
- class: class variable 1: disease
                        0: no disease

Some samples from the dataset:
<Sample 1, class=1>:
mean: 0.769010384, std: 0.761649904,var: 0.580110568, skewness: 0.466414969,
entropy: 0.963101,kurtosis: 0.369918743, contrast: 0.489227072,
correlation: 0.756602121, energy: 0.688121882,homogeneity: 0.954844229
.
.
.
<Sample 5, class=1>:
mean:0.636383104,std: 0.87729013,var: 0.769637974,skewness: 0.586431784,
entropy: 0.973940969,kurtosis: 0.37709383,contrast: 0.488186784,
correlation: 0.526767611,energy: 0.799576354,homogeneity: 0.947847399
<Sample 6, class=0>:
mean: 0.677572277,std: 0.854699258,var: 0.730510817,skewness: 0.549493916,
entropy: 0.980216291,kurtosis: 0.395917515,contrast: 0.474908547,
correlation: 0.69608474,energy: 0.762089244,homogeneity: 0.983787588
.
.
.
<Sample 10, class=0>:
mean: 0.568418388,std: 0.879868278,var: 0.774168186,skewness: 0.7706753,
entropy: 0.964696511,kurtosis: 0.47484991,contrast: 0.483253361,
correlation: 0.717102329,energy: 0.859660247,homogeneity: 0.926171716
```

```
# Feature name and description: "mean_variance_ratio", ratio of mean to variance.
# Usefulness: This feature could highlight the relationship between the mean and
    variance of the image, which could be useful in identifying patterns related
    to the presence of a disease.
# Input samples: 'mean': [0.4967, 0.7235, 0.5506],
                'var': [0.879374249, 0.774168186, 0.691100425]


df['mean_variance_ratio'] = df['mean'] / df['var']
```

```
Iteration 1
Performance before adding features ROC 0.912, ACC 0.946.
Performance after adding features ROC 0.911, ACC 0.949.
Improvement ROC -0.001, ACC 0.003. The code was executed and changes to ´df´
were kept.
```

```python
# Feature name and description: "mean_variance_ratio", ratio of mean to variance.
# Usefulness: This feature could highlight the relationship between the mean and
    variance of the image, which could be useful in identifying patterns related
    to the presence of a disease.
# Input samples: 'mean': [0.4967, 0.7235, 0.5506],
                 'var': [0.879374249, 0.774168186, 0.691100425]


df['mean_variance_ratio'] = df['mean'] / df['var']
```

```
Iteration 1
Performance before adding features ROC 0.912, ACC 0.946.
Performance after adding features ROC 0.911, ACC 0.949.
Improvement ROC -0.001, ACC 0.003. The code was executed and changes to ´df´
were kept.
```

```python
# Feature name and description: "contrast_energy_ratio", ratio of contrast to
    energy.
# Usefulness: This feature could highlight the relationship between the contrast
    and energy of the image, which could be useful in identifying patterns
    related to the presence of a disease.
# Input samples: 'contrast': [0.5, 0.51, 0.48], 'energy': [1.0, 0.78, 0.92]


df['contrast_energy_ratio'] = df['contrast'] / df['energy']
```

```
Iteration 2
Performance before adding features ROC 0.911, ACC 0.949.
Performance after adding features ROC 0.910, ACC 0.950.
Improvement ROC -0.001, ACC 0.001. The code was executed and changes to ´df´
were kept.
```

```python
# Explanation why the column `correlation` is dropped
# The correlation of an image might not provide much useful information in terms
    of disease detection as it measures the degree to which the two variables in
    the image move in relation to each other, which might not be directly related
    to the presence of a disease.


df.drop(columns=['correlation'], inplace=True)
```

```
Iteration 5
Performance before adding features ROC 0.911, ACC 0.949.
Performance after adding features ROC 0.911, ACC 0.950.
Improvement ROC 0.000, ACC 0.001. The code was executed and changes to ´df´
were kept.
```

```python
# Explanation why the column `contrast` is dropped
# The contrast of an image might not provide much useful information in terms
    of disease detection as it measures the degree of brightness and darkness
    of an image and it might not be very useful in disease detection.


df.drop(columns=['contrast'], inplace=True)
```

```
Iteration 7
Performance before adding features ROC 0.911, ACC 0.950.
Performance after adding features ROC 0.913, ACC 0.952.
Improvement ROC 0.002, ACC 0.002. The code was executed and changes to ´df´
were kept.
```

```
# Explanation why the column ‘energy‘ is dropped
# The contrast of an image might not provide much useful information in terms
    of disease detection as it measures local change in an image and it
    might not be very useful in disease detection.


df.drop(columns=['energy'], inplace=True)
```

```
Iteration 8
Performance before adding features ROC 0.913, ACC 0.952.
Performance after adding features ROC 0.913, ACC 0.953.
Improvement ROC 0.000, ACC 0.001. The code was executed and changes to ´df´
were kept.
```

Figure 14. Exemplary run of CAAFE on the RustNet image dataset. User generated input is shown in blue, ML-classifier generated data shown in red, and LLM generated code is shown with syntax highlighting. The generated code contains a comment for each generated/deleted feature that follows a template provided in our prompt (feature name, description of usefulness, features used in the generated code, and sample values of these features). In this run, CAAFE improves the ACC on the validation dataset from 0.946 to 0.953. There were a total of 10 iterations, but only those that improved the ACC are included in the figure. [MET]

```
This dataset contains handcrafted statistical features from images.
The images are wheat images and they are categorized in six classes
with six different levels of infections.


Attribute Information:

- mean: mean of the image (numerical)
- std: standard deviation of the image (numerical)
- var: variance of the image (numerical)
- skewness: skewness of the image (numerical)
- entropy: entropy of the image (numerical)
- kurtosis: kurtosis of the image (numerical)
- contrast: contrast of the image (numerical)
- correlation: correlation of the image (numerical)
- energy: energy of the image (numerical)
- homogeneity: homogeneity of the image (numerical)
- class: class variable
        0: no infection
        1: Small uredia are present and surrounded by either chlorotic or necrotic
            areas
        2: Variable-sized uredia are present; some withchlorosis, necrosis,
            or both
        3: Medium-sized uredia are present and possibly surrounded by chlorotic
            areas
        4: Visible chlorosis or necrosis, no uredia are present
        5: Large uredia are present, generally with little or no chlorosis and
            no necrosis.

Some samples from the dataset:
<Sample 1, class=0>:
mean:0.412029,std: 0.598975,var: 0.436435,skewness: 0.525828,
entropy: 0.779089,kurtosis: 0.085871,contrast: 0.186343,
correlation: 0.430613,energy: 0.364397,homogeneity:0.649717
.
.
.
<Sample 6, class=5>:
mean:0.543930,std: 0.629128,var: 0.471242,skewness: 0.553025,
entropy:0.839109,kurtosis: 0.085644 ,contrast: 0.496534,
correlation: 0.357381,energy: 0.224802,homogeneity: 0.333801
```

```
# Feature name and description: 'contrast_entropy_ratio'
# Usefulness: This feature is useful because it gives us a measure of how
    the contrast is related to the entropy.
# Higher contrast and lower entropy could indicate a more uniform image,
    which might be linked to the absence of a disease.
# Input samples: 'contrast': [0.186343, 0.118865,0.556068],
    'entropy': [0.779089, 0.764887, 0.908173]


df['contrast_entropy_ratio'] = df['contrast'] / df['entropy']
```

```
Iteration 2
Performance before adding features ROC 0.638, ACC 0.894.
Performance after adding features ROC 0.638, ACC 0.896.
Improvement ROC 0.001, ACC 0.002. The code was executed and changes to ´df´
were kept.
```

```
# Feature name and description: 'std_skewness_product'
# Usefulness: This feature is useful because it gives us a measure
    of how the standard deviation is related to the skewness.
# A high product could indicate a high variability and asymmetry
    in the image, which might be linked to the presence of a disease.
# Input samples: 'std': [0.598975, 0.591202, 0.620972],
    'skewness': [0.622890,0.808570, 0.467779]


df['std_skewness_product'] = df['std'] * df['skewness']
```

```
Iteration 3
Performance before adding features ROC 0.638, ACC 0.896.
Performance after adding features ROC 0.640, ACC 0.896.
Improvement ROC 0.002, ACC 0.000. The code was executed and changes to ´df´
were kept.
```

```
# Feature name and description: 'kurtosis_entropy_ratio'
# Usefulness: This feature is useful because it gives us a measure of
    how the kurtosis is related to the entropy.
# A higher ratio could indicate a more peaked distribution with a higher
    uncertainty, which might be linked to the presence of a disease.
# Input samples: 'kurtosis': [0.085871 , 0.225140, 0.144952],
    'entropy': [0.651275, 0.922197, 0.839109]


df['kurtosis_entropy_ratio'] = df['kurtosis'] / df['entropy']
```

```
Iteration 5
Performance before adding features ROC 0.640, ACC 0.896.
Performance after adding features ROC 0.641, ACC 0.895.
Improvement ROC 0.001, ACC -0.001. The code was executed and changes to ´df´
were kept.
```

```
# Feature name and description: 'contrast_energy_ratio'
# Usefulness: This feature is useful because it gives us a measure of how the
    contrast is related to the energy.
# A higher ratio could indicate a more contrasty image with less energy, which
    might be linked to the presence of a disease.
# Input samples: 'contrast': [0.186343, 0.118865, 0.556068],
    'energy': [0.340916 , 0.436955, 0.224802]


df['contrast_energy_ratio'] = df['contrast'] / df['energy']
```

```
Iteration 6
Performance before adding features ROC 0.641, ACC 0.895.
Performance after adding features ROC 0.648, ACC 0.901.
Improvement ROC 0.006, ACC 0.005. The code was executed and changes to ´df´
were kept.
```

```
# Feature name and description: 'homogeneity_skewness_product'
# Usefulness: This feature is useful because it gives us a measure of how the
    homogeneity is related to the skewness.
# A high product could indicate a high homogeneity and asymmetry in the image,
    which might be linked to the presence of a disease.
# Input samples: 'homogeneity': [0.649717, 0.293929 , 0.333801 ],
    'skewness': [0.622890, 0.553025, 0.603786]


df['homogeneity_skewness_product'] = df['homogeneity'] * df['skewness']
```

```
Iteration 9
Performance before adding features ROC 0.648, ACC 0.901.
Performance after adding features ROC 0.648, ACC 0.901.
Improvement ROC 0.000, ACC 0.000. The code was executed and changes to ´df´
were kept.
```

Figure 15. Exemplary run of CAAFE on the Yellow-Rust-19 image dataset. User generated input is shown in blue, ML-classifier generated data shown in red, and LLM generated code is shown with syntax highlighting. The generated code contains a comment for each generated/deleted feature that follows a template provided in our prompt (feature name, description of usefulness, features used in the generated code, and sample values of these features). In this run, CAAFE improves the ACC on the validation dataset from 0.896 to 0.901. There were a total of 10 iterations, but only those that improved the ACC are included in the figure. [MET]

AutoML The results for both datasets using TPOT and ResNet-18 are presented in Tables 24 and 25, respectively. For the RustNet dataset, TPOT achieved an accuracy of 93.02%, while ResNet-18 reached a test accuracy of 85.20%. The training and test accuracy of ResNet-18 for the RustNet dataset are shown in Figure 16(a). The number of epochs is limited due to the fixed time budget.

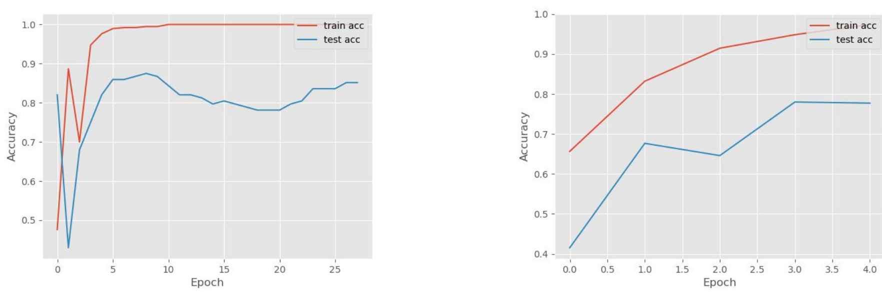

Figure 16. ResNet-18 Accuracy for (a) RustNet dataset, (b) Yellow-Rust-19 dataset [MET]

For the Yellow-Rust-19 dataset, the time budget was extended to 100 minutes because of the large number of images. TPOT achieved a test accuracy of 96.4%, whereas ResNet-18 achieved only 77.7%. The training and test accuracy of ResNet-18 for the Yellow-Rust-19 dataset are presented in Figure 16(b). The limited number of epochs within the allotted time indicates the high computational effort required. The confusion matrices for both approaches and datasets are shown in Figures 17 and 18. As observed, for both datasets, TPOT produced a higher sum of true positive and true negative cases compared to ResNet-18. Similarly, the sum of false positive and false negative cases was lower for TPOT than for ResNet-18.

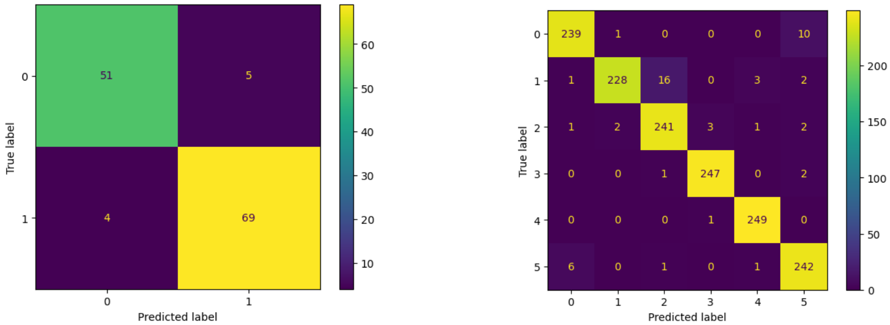

Figure 17. TPOT Confusion Matrix for (a) RustNet dataset, (b) Yellow-Rust-19 dataset [MET]

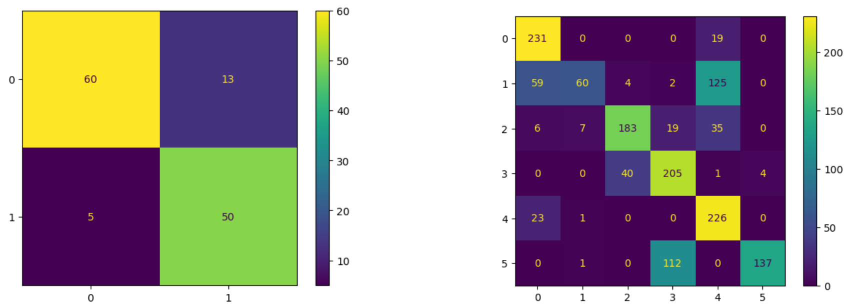

Figure 18. ResNet-18 Confusion Matrix for (a) RustNet dataset, (b) Yellow-Rust-19 dataset [MET]

Model	**Dataset**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
TPOT	RustNet	93.02%	92.99	92.90	92.80
CAAFE with TPOT	RustNet	95.35%	95.79	94.85	95.22
TPOT	Yellow-Rust-19	96.40%	96.46	96.40	96.39
CAAFE with TPOT	Yellow-Rust-19	95.42%	95.26	94.85	95.22

Table 24. TPOT performance on both datasets [MET]

Dataset	Accuracy	Precision	Recall	F1-Score
RustNet	85.2%	86.13	86.54	86.15
Yellow-Rust-19	77.70%	79.36	77.77	77.70

Table 25. ResNet-18 performance on both datasets [MET]

4.4.6. Discussion

The study presents a significant advancement in the application of Automated Machine Learning (AutoML) for agricultural disease detection, specifically targeting wheat stripe rust. By integrating AutoML with context-aware feature engineering, this research demonstrates a novel approach that not only enhances the accuracy of disease detection but also addresses the growing need for scalable, efficient, and interpretable machine learning models in agriculture.

Some general remarks on the complexity of TPOT: The complexity of TPOT is not formally defined but we can understand its complexity by breaking down different components of it.

- The computational complexity of TPOT is influenced by several factors, including the number of input variables, the size of the search space (which includes all possible pipelines), and the number of generations and population size used in the genetic programming algorithm.
- As the number of input variables increases, the search space expands exponentially, leading to a corresponding increase in the computational cost. This is because more complex pipelines involving different combinations of pre-processing steps, models, and hyperparameters need to be evaluated.
- Additionally, the number of copies (i.e., the population size in the genetic algorithm) also directly impacts the computational complexity. A larger population size results in more candidate pipelines being evaluated in each generation, increasing the overall computational burden.
- TPOT's complexity can be expressed as $O(N \times G \times E \times C)$, where N is the number of input variables, G is the number of generations, E is the number of evaluations per generation, and C is the cost of evaluating each pipeline (which itself depends on the complexity of the models and the dataset size).
- Therefore, while TPOT provides a powerful automated approach to model optimization, it is computationally intensive, especially for large datasets or when the search space is broad. Careful tuning of the genetic programming parameters (such as limiting the number of generations or reducing the population size) can help manage the computational demands.

Performance of AutoML in Disease Detection: The proposed methodology, which combines AutoML with rigorous feature engineering, has shown remarkable performance in detecting stripe rust in wheat crops. The results clearly indicate that this approach outperforms traditional deep learning models like ResNet-

18, achieving significantly higher accuracy on both the RustNet and Yellow-Rust-19 datasets. Specifically, the accuracy improvements from 85.2% to 95.35% for the RustNet dataset and from 77.7% to 95.42% for the Yellow-Rust-19 dataset underscore the potential of this method in practical agricultural applications. One of the key factors contributing to this success is the use of context-aware automated feature engineering. By using advanced language models to generate semantically meaningful features, the approach captures subtle patterns and nuances that are often missed by conventional image processing techniques. This feature engineering process, which iteratively refines the dataset, has proven to be highly effective in enhancing the discriminative power of the extracted features, leading to superior model performance.

Comparison with Traditional Deep Learning Models: The comparison with ResNet-18, a state-of-the-art deep learning model, highlights the advantages of the proposed AutoML framework. While deep learning models have traditionally been favored for their high accuracy in image classification tasks, they are often criticized for their computational intensity and the substantial resources required for training. In contrast, the AutoML approach not only achieves higher accuracy but also does so with significantly lower computational costs and shorter training times. This is particularly important in agricultural settings, where resources may be limited, and the timely deployment of models is critical for effective disease management.

Furthermore, the interpretability of the AutoML-generated models presents another significant advantage over deep learning models. In agriculture, where the stakes of incorrect decisions are high, being able to understand and trust the model's predictions is crucial. The use of transparent and interpretable models, as demonstrated in this study, ensures that the decision-making process remains accessible to stakeholders, from farmers to policymakers.

Implications for Sustainable Agriculture: The implications of this research extend beyond the immediate benefits of improved disease detection. The study also addresses the broader issue of sustainability in AI, particularly in the context of agricultural applications. The proposed method exemplifies a green AI approach, optimizing for both accuracy and resource efficiency. By reducing the computational footprint of machine learning models, this approach aligns with the growing emphasis on sustainable AI practices, which is becoming increasingly important as AI systems are integrated into critical sectors like agriculture.

Moreover, the ability to automate the feature engineering and model selection process through AutoML reduces the need for expert intervention, making advanced machine learning techniques more accessible to a broader audience. This democratization of technology can significantly enhance the scalability and impact of AI in agriculture, contributing to improved food security and sustainable farming practices worldwide.

Limitations and Future Research: Despite the promising results, there are several limitations to this study that warrant further exploration. The datasets used, while publicly available and representative, may not capture the full variability encountered in real-world agricultural settings. Future research should focus on validating the proposed approach across a wider range of conditions, including different crops, environmental factors, and disease types.

Additionally, while the context-aware feature engineering approach has proven effective, its reliance on large language models introduces a dependency on high-quality textual data and expert knowledge. Exploring ways to further automate this process or integrate it with other data modalities could enhance the robustness and applicability of the method.

Finally, the long-term deployment of such models in real-world agricultural systems will require ongoing monitoring and adaptation to ensure continued accuracy and relevance. Future studies should investigate methods for the continual learning and adaptation of these models to new data and changing environmental conditions.

Conclusion: In conclusion, this study introduces a powerful and efficient approach to wheat disease detection, exploring the strengths of AutoML and advanced feature engineering. The results demonstrate that this method not only outperforms traditional deep learning models but also offers a more sustainable and interpretable solution for agricultural applications. As the agricultural sector continues to face the challenges of disease management and food security, such innovative approaches will be critical in ensuring the resilience and sustainability of global food systems.

5. CONCLUSION

This thesis has addressed several critical issues in the agricultural and environmental sciences by focusing on the application of advanced machine learning techniques to improve sustainability and productivity in agriculture. These challenges have been approached through innovative methods that leverage automation and data-driven strategies, with the aim of providing solutions that are both efficient and interpretable, with practical applications in real-world agricultural contexts.

The first major contribution of this thesis, as detailed in [MET], involves the detection and classification of diseased wheat crops using automated machine learning (AutoML) techniques. Traditional methods for detecting crop diseases often depend on manual processes or non-adaptive algorithms, which are labor-intensive, time-consuming, and prone to errors due to environmental variability. In contrast, this research introduces a context-aware AutoML framework that dynamically adapts to different environmental and crop conditions. The proposed system significantly improves the robustness and reliability of the disease detection process, outperforming traditional methods in both precision and adaptability. This advancement is crucial for precision agriculture, as it provides a scalable and efficient solution that can be deployed across various agricultural settings. The improvements in accuracy and adaptability can lead to better crop management, reduced resource wastage, and more informed decision-making for farmers and agricultural stakeholders.

In [Tom+23], the focus shifted to ensuring sustainability within agri-food systems by integrating machine learning techniques into a cohesive framework for sustainable agriculture. This research specifically examined the use of sustainable and interpretable computing schemes in early-stage decision-making through data-driven ex-ante Life Cycle Assessment (LCA) for crop production. By using a combination of models, the research provided a comprehensive toolkit that allows stakeholders to make informed decisions, optimize resource utilization, and enhance crop yields sustainably. The framework emphasizes the importance of early interventions, allowing stakeholders to anticipate and mitigate potential environmental impacts before they occur.

The third significant contribution, as presented in [MTL23], addresses the critical issue of prediction of evapotranspiration (ET), which is vital for effective water management in agriculture. Accurate prediction of ET is essential to determine the water requirements of crops, thereby facilitating efficient irrigation practices. This study compared various interpretable machine learning models, including Decision Trees (DTs), Adaptive Neuro Fuzzy Inference System with Fractional Tikhonov Regularization (ANFIS-T), and Support Vector Regression (SVR). The findings concluded that ANFIS-T, due to its balance of accuracy and interpretability, is particularly suitable for applications where transparent decision-making is crucial. The model's ability to provide high interpretability with a minimal number of rules is especially beneficial for understanding and managing water use in agriculture, making it a preferred choice for practitioners and researchers.

Finally, [Muh+22] tackled the challenge of predicting soil organic carbon (SOC), which is essential for maintaining soil health and facilitating carbon sequestration. SOC plays a significant role in enhancing soil fertility and mitigating climate change by acting as a carbon sink. The study compared the performance of ANFIS-T and FT-rELM (Fractional Tikhonov regularized Extreme Learning Machine) in predicting SOC levels across various datasets. The research found that while ANFIS-T generally achieved higher accuracy, FT-rELM offered faster training times, highlighting the importance of balancing accuracy with computational efficiency. This is particularly relevant in scenarios where real-time data processing is critical, such as in large-scale agricultural operations. The findings contribute to a deeper understanding of soil resources, promoting sustainable agricultural practices that improve soil health and productivity.

5.1. Limitations and Future Research Directions

While the proposed models demonstrate substantial improvements in accuracy, interpretability, and computational efficiency, several limitations were identified that require further exploration. One significant limitation is the reliance on high-quality labeled data for the AutoML framework, which may limit its applicability in regions where such data are scarce or difficult to obtain. Additionally, the models were tested primarily on specific datasets, raising questions about their generalizability to other crops, environmental conditions, or geographical regions. Future research should aim to validate these models across a broader range of conditions to ensure their robustness and adaptability in diverse agricultural settings.

Another area for future research involves the scalability of these models. As agricultural datasets continue to grow in size and complexity, the ability of models like ANFIS-T to maintain high levels of accuracy and interpretability without becoming overly complex or computationally intensive will be critical. Research into hybrid models that combine the strengths of different machine learning techniques could offer a solution, potentially improving both scalability and performance.

Furthermore, while this thesis focused on the technical aspects of model development and evaluation, there is a need for more research on the practical implementation of these models in real-world agricultural systems. This includes investigating how these models can be integrated with real-time monitoring systems, automated decision-support tools, and other technologies that are increasingly being used in modern agriculture. Such integration would enhance the practical utility of the models, making them more accessible and useful to a wider range of stakeholders, from smallholder farmers to large agricultural enterprises.

Finally, future work should explore the development of more user-friendly interfaces and visualization tools for these models. Making the outputs of these models more accessible to non-experts is crucial for their widespread adoption. By improving the user interface and experience, the models can be made more intuitive, allowing stakeholders to easily interpret the results and make informed decisions without needing extensive technical expertise.

5.2. Practical Implications and Integration of Models

The practical implications of this research are significant and wide-ranging, offering tangible benefits to practitioners in the agri-food sector. The context-aware AutoML framework, for example, has the potential to revolutionize crop monitoring systems by enabling early and accurate detection of diseases. This can lead to timely interventions, reducing crop losses and improving overall yield. Moreover, the integration of ANFIS-T in water management practices can optimize irrigation strategies, leading to more efficient water use and contributing to sustainability goals.

One of the key insights from this research is the potential for these models to work together in a cohesive framework rather than as isolated solutions. For example, integrating the SOC prediction model with the evapotranspiration prediction system could provide a more holistic understanding of soil health and water requirements, enabling more precise and effective agricultural management. Similarly, combining the context-aware AutoML system with the sustainable computing framework could enhance the decision-making process by providing real-time, data-driven insights that consider both short-term crop health and long-term sustainability.

Together, these models can provide a comprehensive approach to managing the various challenges facing agriculture today. This integrated approach ensures that decisions made in an area, such as irrigation management, are informed by data from other areas, such as soil health or disease detection, leading to more coordinated and effective agricultural practices. This holistic perspective is essential for the advancement of precision agriculture, where the goal is to optimize every aspect of farming to achieve maximum productivity with minimal environmental impact.

5.3. Conclusion

In conclusion, this thesis has demonstrated the potential of advanced machine learning techniques to address some of the most pressing challenges in agricultural sustainability and productivity. By focusing on the key areas of interpretability, scalability, and efficiency, the proposed models not only improve current practices but also pave the way for future innovations in the agri-food sector. The integration of these models into a cohesive framework for sustainable agriculture represents a significant step forward, enabling more informed decision-making, better resource management, and ultimately, more sustainable and productive agricultural practices.

The findings of this research have important implications for the future of agriculture, particularly in the context of increasing global food demand and the need to minimize environmental impacts. As the agricultural sector continues to evolve, the contributions of this thesis will play a crucial role in supporting the transition to more sustainable and resilient agricultural systems. By addressing the identified limitations and exploring new areas of research, future work will further enhance the applicability and impact of these models, ensuring that they can be widely adopted and used effectively in different agricultural contexts.

The adoption of these innovative approaches by stakeholders in the agricultural sector will be essential in meeting the growing demand for food while minimizing environmental impacts. As these models are further refined and integrated into real-world systems, they have the potential to transform agriculture, making it more efficient, sustainable, and resilient in the face of global challenges.

BIBLIOGRAPHY

[Aby+16] Hamid Abyaneh et al. "Soil temperature estimation using an artificial neural network and co-active neuro-fuzzy inference system in two different climates". In: *Arabian Journal of Geosciences* 9 (May 2016). DOI: 10.1007/s12517-016-2388-8.

[Ace+18] Maricelis Acevedo et al. "The role of wheat in global food security". In: May 2018, pp. 81–110. ISBN: 9781138300590. DOI: 10.4324/9780203733301-4.

[Alc+10] Jesus Alcala-Fdez et al. "KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework". In: *Journal of Multiple-Valued Logic and Soft Computing* 17 (Jan. 2010), pp. 255–287.

[APS98] Richard Allan, L. Pereira, and Martin Smith. *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56*. Vol. 56. Jan. 1998.

[Ayt09] Ali Aytek. "Co-active neurofuzzy inference system for evapotranspiration modeling". In: *Soft Comput.* 13 (May 2009), pp. 691–700. DOI: 10.1007/s00500-008-0342-8.

[Ban+98] Wolfgang Banzhaf et al. "Genetic Programming: An Introduction on the Automatic Evolution of computer programs and its Applications". In: Jan. 1998.

[Bar+12] Rodrigo Barros et al. "A Survey of Evolutionary Algorithms for Decision-Tree Induction". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 42 (Jan. 2012), pp. 291–312. DOI: 10.1109/TSMCC.2011.2157494.

[Bed21] Jatin Bedi. "Transfer learning augmented enhanced memory network models for reference evapotranspiration estimation". In: *Knowledge-Based Systems* 237 (Nov. 2021), p. 107717. DOI: 10.1016/j.knosys.2021.107717.

[BPC23] Rina Bora, Deepa Parasar, and Shrikant Charhate. "A detection of tomato plant diseases using deep learning MNDLNN classifier". In: *Signal, Image and Video Processing* 17 (Apr. 2023), pp. 1–9. DOI: 10.1007/s11760-023-02498-y.

[BPM14] Thomas Brooks, D. Pope, and Michael Marcolini. *Airfoil Self-Noise*. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5VW2C. 2014.

[Bre+84] L. Breiman et al. *Classification and Regression Trees*. Taylor & Francis, 1984. ISBN: 9780412048418. URL: https://books.google.com.pk/books?id=JwQx-WOmSyQC.

[CG14] Edwin Carlinet and Thierry Géraud. "A Comparative Review of Component Tree Computation Algorithms". In: *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 23 (July 2014). DOI: 10.1109/TIP.2014.2336551.

[CG16] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[Che20] Xianming Chen. "Pathogens which threaten food security: Puccinia striiformis, the wheat stripe rust pathogen". In: *Food Security* 12 (Mar. 2020). DOI: 10.1007/s12571-020-01016-z.

[Col] Trinity College. *Fuzzy Sets and Rule Bases*. URL: https://www.maths.tcd.ie/~ormondca/notes/Fuzzy%20Sets%20and%20Rule%20Bases.pdf.

[Deb+02] Kalyan Deb et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II". In: *Evolutionary Computation, IEEE Transactions on* 6 (May 2002), pp. 182–197. DOI: 10.1109/4235.996017.

[DK17] Finale Doshi-Velez and Been Kim. "Towards a rigorous science of interpretable machine learning". In: *arXiv preprint arXiv:1702.08608* (2017).

[Dua+22] Efraín Duarte et al. "Digital mapping of soil organic carbon stocks in the forest lands of Dominican Republic". In: *European Journal of Remote Sensing* 55 (Dec. 2022), pp. 213–231. DOI: 10.1080/22797254.2022.2045226.

[Eld+24] Hassan Eldeeb et al. "AutoMLBench: A comprehensive experimental evaluation of automated machine learning frameworks". In: *Expert Systems with Applications* 243 (2024), p. 122877. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2023.122877. URL: https://www.sciencedirect.com/science/article/pii/S0957417423033791.

[Er-+13] Salah Er-Raki et al. "Micrometeorology Tools for Measuring Evapotranspiration from the Leaf to the Region." In: Feb. 2013, p. 22. ISBN: 978-1-62417-138-3.

[Eur16] European Union. *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. 2016.

[Fer+21] Mariza Ferro et al. "Towards a Sustainable Artificial Intelligence: A Case Study of Energy Efficiency in Decision Tree Algorithms". In: *Concurrency and Computation Practice and Experience* 35 (Dec. 2021). DOI: 10.1002/cpe.6815.

[For+12] Félix-Antoine Fortin et al. "DEAP: Evolutionary algorithms made easy". In: *Journal of Machine Learning Research, Machine Learning Open Source Software* 13 (July 2012), pp. 2171–2175.

[Gar+22] M. Carmen Garrido et al. "Evaporation Forecasting through Interpretable Data Analysis Techniques". In: *Electronics* 11.4 (2022). ISSN: 2079-9292. DOI: 10.3390/electronics11040536. URL: https://www.mdpi.com/2079-9292/11/4/536.

[Gha+22] Hassan Ghasemi Mobtaker et al. "A comparative of modeling techniques and life cycle assessment for prediction of output energy, economic profit, and global warming potential for wheat farms". In: *Energy Reports* 8 (Mar. 2022), pp. 4922–4934. DOI: 10.1016/j.egyr.2022.03.184.

[Ghe+21] Maryam Ghebleh et al. "Using environmental variables and Fourier Transform Infrared Spectroscopy to predict soil organic carbon". In: *Catena* 202 (Feb. 2021). DOI: 10.1016/j.catena.2021.105280.

[Gho+22] Vahid Gholami et al. "Evaluating the effects of vegetation and land management on runoff control using field plots and machine learning models". In: *Environmental Science and Pollution Research* 30 (Nov. 2022). DOI: 10.1007/s11356-022-24347-0.

[GL07] Peter Glavič and Rebeka Lukman. "Review of sustainability terms and their definitions". In: *Journal of Cleaner Production* 15 (Dec. 2007), pp. 1875–1885. DOI: 10.1016/j.jclepro.2006.12.006.

[Gu+23] Chunyan Gu et al. "Assessing narrow brown leaf spot severity and fungicide efficacy in rice using low altitude UAV imaging". In: *Ecological Informatics* 77 (July 2023), p. 102208. DOI: 10.1016/j.ecoinf.2023.102208.

[Hay+23] Tolga Hayıt et al. "The classification of wheat yellow rust disease based on a combination of textural and deep features". In: *Multimedia Tools and Applications* 82 (May 2023), pp. 1–19. DOI: 10.1007/s11042-023-15199-y.

[He+19] Bin He et al. "Product carbon footprint across sustainable supply chain". In: *Journal of Cleaner Production* 241 (Sept. 2019), p. 118320. DOI: 10.1016/j.jclepro.2019.118320.

[Hen+22] Peter Henderson et al. *Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning*. 2022. arXiv: 2002.05651 [cs.CY].

[HLS15] Frank Hutter, Jörg Lücke, and Lars Schmidt-Thieme. "Beyond Manual Tuning of Hyperparameters". In: *KI - Künstliche Intelligenz* 29 (July 2015). DOI: 10.1007/s13218-015-0381-0.

[HMH23] Noah Hollmann, Samuel Müller, and Frank Hutter. *LLMs for Semi-Automated Data Science: Introducing CAAFE for Context-*

Aware Automated Feature Engineering. 2023. arXiv: 2305.03403 [cs.AI].

[Hol+23] Noah Hollmann et al. "TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second". In: (2023). arXiv: 2207.01848 [cs.LG].

[Hom+20] Wladyslaw Homenda et al. "Multicriteria Decision Making: Scale, Polarity, Symmetry, Interpretability". In: July 2020, pp. 1–8. DOI: 10.1109/FUZZ48607.2020.9177705.

[Hos13] Parisa Hosseinzadehtalaei. "Daily soil temperature modeling using neuro-fuzzy approach". In: *Theoretical and Applied Climatology* 118 (Dec. 2013). DOI: 10.1007/s00704-013-1084-9.

[HSW04] Samuel Huan, Sunil Sheoran, and Ge Wang. "A Review and Analysis of Supply Chain Operations Reference (SCOR) Model". In: *Supply Chain Management: An International Journal* 9 (Feb. 2004), pp. 23–29. DOI: 10.1108/13598540410517557.

[IN00] Hisao Ishibuchi and Manabu Nii. "Neural networks for soft decision making". In: *Fuzzy Sets and Systems* 115.1 (2000), pp. 121–140. ISSN: 0165-0114. DOI: https://doi.org/10.1016/S0165-0114(99)00022-6. URL: https://www.sciencedirect.com/science/article/pii/S0165011499000226.

[Jan93] J. S. R. Jang. "ANFIS: Adaptive-Network-Based Fuzzy Inference System". In: *IEEE Transactions on Systems Man & Cybernetics* 23 (3 1993), pp. 665–685. DOI: https://doi.org/10.1109/21.256541.

[Kan+22] E. Kanari et al. "A robust initialization method for accurate soil organic carbon simulations". In: *Biogeosciences* 19.2 (2022), pp. 375–387. DOI: 10.5194/bg-19-375-2022. URL: https://bg.copernicus.org/articles/19/375/2022/.

[Kau+15] Karin Kauer et al. "Soil carbon dynamics estimation and dependence on farming system in a temperate climate". In: *Soil and Tillage Research* 154 (2015), pp. 53–63. ISSN: 0167-1987. DOI: https://doi.org/10.1016/j.still.2015.06.010. URL: https://www.sciencedirect.com/science/article/pii/S0167198715001270.

[Kau+21] Karin Kauer et al. "Soil Particulate and Mineral-Associated Organic Matter Increases in Organic Farming under Cover Cropping and Manure Addition". In: *Agriculture* 11 (Sept. 2021), p. 903. DOI: 10.3390/agriculture11090903.

[Kmo+21] Alexander Kmoch et al. "EstSoil-EH: A high-resolution eco-hydrological modelling parameters dataset for Estonia". In: *Earth System Science Data* 13 (Jan. 2021), pp. 83–97. DOI: 10.5194/essd-13-83-2021.

[KPK21a] Paraskevi Karka, Stavros Papadokonstantakis, and Antonis Kokossis. “Digitizing sustainable process development: From ex-post to ex-ante LCA using machine-learning to evaluate bio-based process technologies ahead of detailed design”. In: *Chemical Engineering Science* 250 (Dec. 2021), p. 117339. DOI: 10.1016/j.ces.2021.117339.

[KPK21b] Paraskevi Karka, Stavros Papadokonstantakis, and Antonis Kokossis. “Digitizing sustainable process development: From ex-post to ex-ante LCA using machine-learning to evaluate bio-based process technologies ahead of detailed design”. In: *Chemical Engineering Science* 250 (Dec. 2021), p. 117339. DOI: 10.1016/j.ces.2021.117339.

[LC20] Francisco Rodrigues Lima Junior and Luiz Carpinetti. “An adaptive network-based fuzzy inference system to supply chain performance evaluation based on SCOR® metrics”. In: *Computers & Industrial Engineering* 139 (Jan. 2020). DOI: 10.1016/j.cie.2019.106191.

[Leo+22] Ai Leon et al. “An ex ante life cycle assessment of wheat with high biological nitrification inhibition capacity”. In: *Environmental Science and Pollution Research* 29 (Jan. 2022). DOI: 10.1007/s11356-021-16132-2.

[Li13] Boyang Li. “Soft Decision-Making SVM Using in Fault Detection and Diagnosis”. In: *Information Technology Journal* 12 (2013), pp. 5017–5021. DOI: 10.3923/itj.2013.5017.5021. URL: https://scialert.net/abstract/?doi=itj.2013.5017.5021.

[Lip18] Zachary C Lipton. “The mythos of model interpretability”. In: *Queue* 16.3 (2018), pp. 31–57.

[Liu+23] Yang Liu et al. “Leaf area index estimation under wheat powdery mildew stress by integrating UAV-based spectral, textural and structural features”. In: *Computers and Electronics in Agriculture* 213 (2023), p. 108169. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2023.108169. URL: https://www.sciencedirect.com/science/article/pii/S0168169923005574.

[LL17] Scott M Lundberg and Su-In Lee. “A unified approach to interpreting model predictions”. In: *Advances in neural information processing systems* 30 (2017).

[LOL09] Gorka Landeras, Amaia Ortiz Barredo, and J.J. Lopez. “Forecasting Weekly Evapotranspiration with ARIMA and Artificial Neural Network Models”. In: *Journal of Irrigation and Drainage Engineering* 135 (June 2009), pp. 323–334. DOI: 10.1061/(ASCE)IR.1943-4774.0000008.

[MA19] Mohamed Mattar and Prof Alazba. "GEP and MLR approaches for the prediction of reference evapotranspiration". In: *Neural Computing and Applications* 31 (Oct. 2019). DOI: 10.1007/s00521-018-3410-8.

[Men13] Corrado Mencar. "Interpretability of Fuzzy Systems". In: Nov. 2013. ISBN: 978-3-319-03199-6. DOI: 10.1007/978-3-319-03200-9_3.

[MET] **Muhammad Uzair**, Radwa Elshawi, and Stefania Tomasiello. "Context-Aware AutoML Approach to detect a wheat disease". In: *Scientific Reports* ().

[MH00] Sushmita Mitra and Yoichi Hayashi. "Neuro-fuzzy rule generation: Survey in soft computing framework". In: *IEEE Transactions on Neural Networks* 11.3 (2000), pp. 748–768.

[MJ95] E. Mizutani and J.-S.R. Jang. "Coactive neural fuzzy modeling". In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. Vol. 2. 1995, 760–765 vol.2. DOI: 10.1109/ICNN.1995.487513.

[MK15] Anurag Malik and Anil Kumar. "Pan Evaporation Simulation Based on Daily Meteorological Data Using Soft Computing Techniques and Multiple Linear Regression". In: *Water Resources Management* 29 (Apr. 2015). DOI: 10.1007/s11269-015-0915-0.

[MKP17] Anurag Malik, Anil Kumar, and Jamshid Piri. "Daily suspended sediment concentration simulation using hydrological data of Pranhita River Basin, India". In: *Computers and Electronics in Agriculture* 138 (2017), pp. 20–28. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2017.04.005. URL: https://www.sciencedirect.com/science/article/pii/S0168169916304331.

[Mol20] C. Molnar. *Interpretable Machine Learning*. Leanpub, 2020. ISBN: 9780244768522. URL: https://books.google.com.pk/books?id=jBm3DwAAQBAJ.

[MRT11] Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. "A randomized algorithm for the decomposition of matrices". In: *Applied and Computational Harmonic Analysis* 30 (Jan. 2011), pp. 47–68. DOI: 10.1016/j.acha.2010.02.003.

[MTL23] **Muhammad Uzair**, Stefani Tomasiello, and Evelin Loit. "Interpretable Approaches to Predict Evapotranspiration". In: *Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022)*. Ed. by Ajith Abraham et al. Cham: Springer Nature Switzerland, 2023, pp. 275–284. ISBN: 978-3-031-27524-1.

[Muh+22] **Muhamnmad Uzair** et al. "Predicting the soil organic carbon by recent machine learning algorithms". In: *2022 IEEE Intl Conf on*

Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech). 2022, pp. 1–7. DOI: 10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9928005.

[Nas+95] Warwick Nash et al. *Abalone*. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C55C7W. 1995.

[Ols+16a] Randal Olson et al. "Automating Biomedical Data Science Through Tree-Based Pipeline Optimization". In: Mar. 2016, pp. 123–137. ISBN: 978-3-319-31203-3. DOI: 10.1007/978-3-319-31204-0_9.

[Ols+16b] Randal S. Olson et al. "Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science". In: *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. GECCO '16. Denver, Colorado, USA: ACM, 2016, pp. 485–492. ISBN: 978-1-4503-4206-3. DOI: 10.1145/2908812.2908918. URL: http://doi.acm.org/10.1145/2908812.2908918.

[Ols+16c] Randal S. Olson et al. *Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science*. 2016. arXiv: 1603.06212 [cs.NE].

[OM19] Randal Olson and Jason Moore. "TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning". In: May 2019, pp. 151–160. ISBN: 978-3-030-05317-8. DOI: 10.1007/978-3-030-05318-5_8.

[Ope21] OpenAI Community. *GPT-3 can't count syllables - or doesn't "get" haiku*. https://community.openai.com/t/gpt-3-cant-count-syllables-or-doesnt-get-haiku/18733. Accessed on: 2024-03-1. 2021.

[Ope23] OpenAI. *openai/openai-cookbook: Examples and guides for using the openai api*. https://github.com/openai/openai-cookbook. (Accessed on 03/1/2023). 2023.

[Ott+22] Denise Ott et al. "LCA as decision support tool in the food and feed sector: evidence from R&D case studies". In: *Environment Systems and Decisions* 43 (Aug. 2022). DOI: 10.1007/s10669-022-09874-y.

[Ped+12] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (Jan. 2012).

[Per+17] Rui Pereira et al. "Energy efficiency across programming languages: how do energy, time, and memory relate?" In: Oct. 2017, pp. 256–267. DOI: 10.1145/3136014.3136031.

[PM23] D. Pansy and Murali Malaisamy. "UAV hyperspectral remote sensor images for mango plant disease and pest identification using MD-

FCM and XCS-RBFNN". In: *Environmental Monitoring and Assessment* 195 (Aug. 2023). DOI: 10.1007/s10661-023-11678-9.

[Pri+22] Nancy Prioux et al. "Environmental assessment coupled with machine learning for circular economy". In: *Clean Technologies and Environmental Policy* 25 (Feb. 2022), pp. 1–14. DOI: 10.1007/s10098-022-02275-4.

[PV16] Santosh Patil and Shriniwas Valunjkar. "Utility of Coactive Neuro-Fuzzy Inference System for Runoff Prediction in Comparison with Multilayer Perception". In: *International Journal of Engineering Research* 5 (Jan. 2016), pp. 156–160. DOI: 10.17950/ijer/v5i1/036.

[Rea+08] John Reap et al. "A survey of unresolved problems in life cycle assessment. Part 1: Goal and scope and inventory analysis". In: *The International Journal of Life Cycle Assessment* 13 (June 2008), pp. 290–300. DOI: 10.1007/s11367-008-0008-x.

[RSG16] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ""Why should I trust you?": Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, pp. 1135–1144.

[Rud19] Cynthia Rudin. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". In: *Nature Machine Intelligence* 1.5 (2019), pp. 206–215.

[SBS00] B. Scholkopf, Peter Bartlett, and A. Smola. "Support Vector Regression with Automatic Accuracy Control". In: *Proceedings of the 8th International Conference on Artificial Neural Networks* (Jan. 2000). DOI: 10.1007/978-1-4471-1599-1_12.

[Sim13] P. Simon. *Too Big to Ignore: The Business Case for Big Data*. Wiley and SAS Business Series. Wiley, 2013. ISBN: 9781118642108. URL: https://books.google.gr/books?id=Dn-Gdoh66sgC.

[Sin88] V.P. Singh. *Hydrologic Systems: Rainfall-runoff modeling*. Hydrologic Systems. Prentice Hall, 1988. ISBN: 9780134480510. URL: https://books.google.com.pk/books?id=CbUPAQAAIAAJ.

[SJS22] Gitika Sharma, Sushma Jain, and Ashima Singh. "DeepEvap: Deep reinforcement learning based ensemble approach for estimating reference evapotranspiration". In: *Applied Soft Computing* 125 (June 2022). DOI: 10.1016/j.asoc.2022.109113.

[SR20] Veronica Saiz-Rubio and Francisco Rovira-Más. "From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management". In: *Agronomy* 10 (Feb. 2020). DOI: 10.3390/agronomy10020207.

[SSJ22] Gitika Sharma, Ashima Singh, and Sushma Jain. "Hybrid deep learning techniques for estimation of daily crop evapotranspiration using limited climate data". In: *Computers and Electronics in Agriculture*

202 (Nov. 2022), p. 107338. DOI: 10.1016/j.compag.2022.107338.

[Suo+20] Mingliang Suo et al. "Soft decision-making based on decision-theoretic rough set and Takagi-Sugeno fuzzy model with application to the autonomous fault diagnosis of satellite power system". In: *Aerospace Science and Technology* 106 (July 2020). DOI: 10.1016/j.ast.2020.106108.

[Tan+22] Zhou Tang et al. "Affordable High Throughput Field Detection of Wheat Stripe Rust Using Deep Learning with Semi-Automated Image Labeling". In: (Apr. 2022). DOI: 10.20944/preprints202204.0177.v1.

[Tho+13] Chris Thornton et al. *Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms*. 2013. arXiv: 1208.3719 [cs.LG].

[TK22] Georgios Archimidis Tsalidis and Gijsbert Korevaar. "Environmental assessments of scales: The effect of ex-ante and ex-post data on life cycle assessment of wood torrefaction". In: *Resources, Conservation and Recycling* 176 (2022), p. 105906. ISSN: 0921-3449. DOI: https://doi.org/10.1016/j.resconrec.2021.105906. URL: https://www.sciencedirect.com/science/article/pii/S0921344921005152.

[Tom+23] Stefania Tomasiello et al. "Data-driven approaches for sustainable agri-food: coping with sustainability and interpretability". In: *Journal of Ambient Intelligence and Humanized Computing* 14 (Oct. 2023), pp. 1–12. DOI: 10.1007/s12652-023-04702-w.

[TPL22a] Stefania Tomasiello, Witold Pedrycz, and Vincenzo Loia. *Contemporary Fuzzy Logic, A Perspective of Fuzzy Logic with Scilab*. May 2022. DOI: 10.1007/978-3-030-98974-3.

[TPL22b] Stefania Tomasiello, Witold Pedrycz, and Vincenzo Loia. "On Fractional Tikhonov Regularization: Application to the Adaptive Network-Based Fuzzy Inference System for Regression Problems". In: *Trans. Fuz Sys.* 30.11 (Nov. 2022), pp. 4717–4727. ISSN: 1063-6706. DOI: 10.1109/TFUZZ.2022.3157947. URL: https://doi.org/10.1109/TFUZZ.2022.3157947.

[UN 15] UN General Assembly. *Transforming our world: the 2030 Agenda for Sustainable Development*. 21 October 2015, A/RES/70/1, General Assembly Seventieth Session, New York (last access: 19 January 2022). 2015.

[Van+20] Coen Van der Giesen et al. "A critical view on the current application of LCA for new technologies and recommendations for improved practice". In: *Journal of Cleaner Production* 259 (Mar. 2020), p. 120904. DOI: 10.1016/j.jclepro.2020.120904.

[Vin+19] Ricardo Vinuesa et al. *The role of artificial intelligence in achieving the Sustainable Development Goals*. Apr. 2019.

[WEG87] Svante Wold, Kim Esbensen, and Paul Geladi. "Principal Component Analysis". In: *Chemometrics and Intelligent Laboratory Systems* 2 (Aug. 1987), pp. 37–52. DOI: 10.1016/0169-7439(87)80084-9.

[Xie+22] Boqiang Xie et al. "Estimation of Soil Organic Carbon Content in the Ebinur Lake Wetland, Xinjiang, China, Based on Multisource Remote Sensing Data and Ensemble Learning Algorithms". In: *Sensors* 22.7 (2022). ISSN: 1424-8220. DOI: 10.3390/s22072685. URL: https://www.mdpi.com/1424-8220/22/7/2685.

[Zha+20] Zhaoyu Zhai et al. "Decision support systems for agriculture 4.0: Survey and challenges". In: *Computers and Electronics in Agriculture* 170 (2020), p. 105256. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2020.105256. URL: https://www.sciencedirect.com/science/article/pii/S0168169919316497.

ACKNOWLEDGEMENTS

I would like to begin by expressing my heartfelt gratitude to the almighty Allah for guiding me through this journey and bestowing upon me the strength and perseverance to complete my Ph.D. thesis.

I am highly thankful to my supervisors, Stefania Tomasiello, Radwa El Shawi, and Evelin Loit, for their continuous support, guidance, and encouragement. They really helped me shaping my work and also helped me to grow professionally and personally.

I would also like to extend my heartfelt gratitude to my parents, wife, and extended family, who always stood by my side during this journey and always supported me.

I acknowledge the use of Large Language Models (LLMs) for text rephrasing during the final writing phase of this thesis.

I acknowledge the support through the funding of SusAn, FACCE ERA-GAS, ICT-AGRI-FOOD and SusCrop ERA-NET. I would extend my gratitude to the staff at the University of Tartu's Institute of Computer Science for their assistance and support throughout my Ph.D. studies. The institute provided a supportive and inspiring research environment that greatly contributed to my work.

SISUKOKKUVÕTE

Pehme otsustamine agri-toidu 4.0 jaoks

Väitekiri esitab ulatusliku uurimuse uuenduslikest, andmepõhistest lähenemistest, mis käsitlevad jätkusuutliku põllumajanduse kriitilisi väljakutseid, eriti Agri-Food 4.0 kujunemisjärgus paradigmas. Uurimus keskendub masinõppe (ML) mudelite täpsuse, efektiivsuse ja tõlgendatavuse parandamisele, et ennustada olulisi põllumajanduslikke parameetreid, nagu evapotranspiratsioon, mulla orgaaniline süsinik (SOC), Ex-ante elutsükli hindamine nisu tootmisel ja põllukultuuride haiguste tuvastamine (roostetriibu nisuhaiguse tuvastamine) kasutades pilte, eesmärgiga panustada jätkusuutlike ja vastupidavate agri-toidusüsteemide arengusse.

Põllumajandus seisab silmitsi kasvava survega, mis tuleneb kliimamuutustest, ressursside piirangutest ja toiduga kindlustatuse vajaduse suurenemisest. Need väljakutsed nõuavad arenenud otsustustööriistu, mis optimeerivad ressursikasutust ja parandavad saagikust, säilitades samal ajal keskkonna jätkusuutlikkuse. Agri-Food 4.0 kontseptsioon hõlmab selliste tehnoloogiate integreerimist nagu asjade internet (IoT), pilvearvutus ja tehisintellekt (AI), et revolutsiooniliselt muuta põllumajandust. See väitekiri on kooskõlas nende eesmärkidega, pakkudes välja tõlgendatavad ja jätkusuutlikud otsustamistööriistad, mis käsitlevad nii tõhusust kui ka läbipaistvust, mis on vajalikud põllumajanduse otsustustugisüsteemides (DSS).

Uurimistöö eesmärgid on:

- Arendada tõlgendatavaid masinõppemudeleid, mis tasakaalustavad täpsust ja tõlgendatavust, mis on olulised põllumajanduse sidusrühmadele.
- Parandada põllumajandustavade jätkusuutlikkust, integreerides masinõppemudelid valdkonnapõhiste teadmistega.
- Rakendada automatiseeritud masinõppe (AutoML) tehnikaid põllukultuuride haiguste tuvastamise optimeerimiseks, keskendudes konkreetselt nisuhaiguste tuvastamisele.

Mulla orgaanilise süsiniku (SOC) ennustamine: Mulla orgaaniline süsinik (SOC) on mulla tervise oluline komponent, mis mõjutab selle struktuuri, veemahutavust ja toitainete kättesaadavust, mis kõik on olulised jätkusuutliku põllukultuuride tootmise jaoks. SOC on ka oluline tegur ülemaailmses süsinikuringes, kuna mullad on suurimad maismaa süsiniku reservuaarid. Väikesed muutused SOC tasemetes võivad märkimisväärselt mõjutada atmosfääri CO2 kontsentratsioone, mõjutades seega kliimamuutusi.

Väitekiri uurib erinevaid masinõppe tehnikaid SOC tasemete täpsemaks ja tõhusamaks ennustamiseks. Traditsioonilised SOC hindamise meetodid, mis sageli hõlmavad keerulist ja töömahukat proovide võtmist ja analüüsi, on täiustatud andmepõhiste mudelitega nagu ANFIS-T (Adaptive Neuro-Fuzzy

Inference System koos murdosatikhhonovi regulatsiooniga). Uurimistöö näitab, et ANFIS-T mitte ainult ei paranda ennustuste täpsust, vaid säilitab ka kõrge tõlgendatavuse taseme, mis on praktilises põllumajanduses oluline. See on eriti oluline, kuna see on kooskõlas ülemaailmsete jõupingutustega, nagu EL-i mullastrateegia aastaks 2030, mis rõhutab SOC suurendamist kasvuhoonegaaside vähendamise eesmärkide täitmiseks.

Evapotranspiratsiooni ennustamine: Evapotranspiratsioon (ET) on vee aurustumise ja taimede transpiratsiooni summa. See on kriitiline parameeter veeressursside juhtimisel, eriti põllumajanduses, kus tõhusad niisutusstrateegiad on vajalikud vee kasutamise optimeerimiseks. Põllumajandus moodustab märkimisväärse osa ülemaailmsest veetarbimisest, kus ET esindab suurimat osa veekaost põllumajandussüsteemides.

Traditsioonilised ET mõõtmise ja hindamise meetodid, nagu lüsimeetrid ja evaporimeetrid, on sageli piiratud nende kõrge maksumuse ja raskustega nende ulatuslikul kasutuselevõtul. Väitekiri hindab masinõppemudelite, eriti ANFIS-T, kasutamist ET ennustamisel. Need mudelid pakuvad täpseid ennustusi, olles samas piisavalt tõlgendatavad, et neid saaksid kasutada põllumehed ja põllumajandusjuhid teadlike otsuste tegemiseks niisutuspraktikate kohta. Parandades ET ennustuste täpsust, aitab uurimistöö kaasa põllumajanduse jätkusuutlikumale veekasutusele, mis on kliimamuutuste ja veepuuduse kontekstis üha olulisem.

Ex-ante elutsükli hindamine (LCA) nisu tootmisel: Elutsükli hindamine (LCA) on kõikehõlmav metoodika toodete keskkonnamõjude hindamiseks alates tooraine kaevandamisest kuni tootmise, kasutamise ja kõrvaldamiseni. Traditsioonilised LCA-d viiakse tavaliselt läbi pärast toote elutsükli lõppu (ex-post), piirates nende kasulikkust proaktiivsete ja jätkusuutlike otsuste tegemisel.

Väitekiri edendab ex-ante LCA kasutamist, mis hindab võimalikke keskkonnamõjusid teadus- ja arendustegevuse etapis, võimaldades varasemaid sekkumisi ja jätkusuutlikumaid tootedisainilahendusi. Nisu tootmise kontekstis integreerib väitekiri masinõppe LCA-ga, et ennustada ja leevendada keskkonnamõjusid, nagu kasvuhoonegaaside heitkogused ja ressursikasutus. See lähenemisviis on eriti kooskõlas Euroopa Liidu rohelise kokkuleppega, mille eesmärk on vähendada süsinikuheidet kõigis sektorites, sealhulgas põllumajanduses. ML-i integreerimine LCA-sse mitte ainult ei paranda LCA ennustamisvõimekust, vaid muudab selle ka tõlgendatavamaks ja sidusrühmadele rakendatavaks.

Põllukultuuride haiguste tuvastamine: Põllukultuuride haigused, eriti need, mis mõjutavad põhikultuure nagu nisu, kujutavad endast märkimisväärseid ohte ülemaailmsele toiduga kindlustatusele. Nisu roostetriibuhaigus, mille põhjustajaks on seenpatogeen Puccinia striiformis, on üks kõige laastavamaid nisukultuuri mõjutavaid haigusi. Traditsioonilised haiguste tuvastamise meetodid, mis tuginevad suures osas käsitsi inspekteerimisele, ei ole ulatuslikud ja on vigadele altid.

Väitekiri pakub välja uue AutoML-i raamistiku, mis hõlmab kontekstitundlikku tunnusjoonte inseneeriat, et parandada nisu roostetriibuhaiguse tuvastamise täpsust ja efektiivsust. Kasutades mehitamata õhusõidukitega kogutud pilte ja arenenud pilditöötlusvõtteid, ületab raamistik täpsuse ja arvutusliku efektiivsuse osas traditsioonilised süvaõppemudelid nagu ResNet-18. See lähenemisviis mitte ainult ei lihtsusta haiguste tuvastamise protsessi, vaid suurendab ka jälgimissüsteemide mastaapsust, muutes nende rakendamise suurtele põllumajanduslikele aladele teostatavaks.

Kokkuvõte: Väitekiri järeldab, et arenenud masinõppetehnikate integreerimine valdkonnapõhiste teadmistega parandab oluliselt võimet käsitleda põllumajanduse võtmeprobleeme. Parandades nende mudelite tõlgendatavust, täpsust ja efektiivsust, panustab uurimistöö jätkusuutlikumate ja vastupidavamate agritoidusüsteemide arengusse. Tulevased uurimused võiksid uurida nende tehnikate laiemat rakendamist erinevatel põllukultuuridel ja keskkondades, täiendavalt täiustades mudeleid ja raamistikke, et käsitleda ülemaailmse põllumajanduse arenguvajadusi. Lisaks on tehisintellekti mudelite energiatarbimise vähendamisele suunatud jätkuvad jõupingutused, nagu rõhutatakse väitekirjas, kriitilise tähtsusega, et viia põllumajanduslikud uuendused vastavusse keskkonna jätkusuutlikkuse eesmärkidega.

PUBLICATIONS

CURRICULUM VITAE

Personal data

Name:	Muhammad Uzair
Date of birth:	March 12, 1991
Citizenship:	Pakistan
Contact:	uzair.dev@gmail.com

Education

2021–2024	PhD in Computer Science, University of Tartu, Estonia
2016–2021	Masters in Computer Science, University of Tartu, Estonia
2010–2014	BSc(Honors) in Computer Science, University of Gujrat, Pakistan

Employment

2021–Present	Junior Research Fellow of Artificial Intelligence, Institute of computer science, University of Tartu, Tartu, Estonia
2023–Present	Senior Data Engineer, BMBaltics OÜ

Teaching

Spring 2024	Teaching Assistant for Fuzzy Logic and Soft Computing Course at University of Tartu, Estonia
Spring 2023	Teaching Assistant for Fuzzy Logic and Soft Computing Course at University of Tartu, Estonia
Fall 2022	Teaching Assistant for Algorithmics Course at University of Tartu, Estonia
Spring 2022	Teaching Assistant for Fuzzy Logic and Soft Computing Course at University of Tartu, Estonia

Thesis Supervision

Ayushmat Bhardwaj Soni (MSc): Neuro-fuzzy systems for binary classification

Scientific work

Main fields of interest:

- Fuzzy Logic
- AutoML

ELULOOKIRJELDUS

Isikuandmed

Name:	Muhammad Uzair
Date of birth:	March 12, 1991
Citizenship:	Pakistan
Contact:	uzair.dev@gmail.com

Haridus

2021–2024	PhD informaatikas, Tartu Ülikool, Eesti
2016–2021	Magistrikraad informaatikas, Tartu Ülikool, Eesti
2010–2014	BSc (kiituskirjaga) informaatikas, Gujrati ülikool, Pakistan

Teenistuskäik

2021–Present	Tehisintellekti nooremteadur, Arvutiteaduse instituut, Tartu Ülikool
2023–Present	Vanem andmeinsener, BMBaltics OÜ

Õpetamine

Kevad 2024	Aine "Hägusloogika ja pehme andmetöötlus"õppeassistent, Arvutiteaduse instituut, Tartu Ülikool
Kevad 2023	Aine "Hägusloogika ja pehme andmetöötlus"õppeassistent, Arvutiteaduse instituut, Tartu Ülikool
Sügis 2022	Aine Älgoritmika"õppeassistent, Arvutiteaduse instituut, Tartu Ülikool
Kevad 2022	Aine "Hägusloogika ja pehme andmetöötlus"õppeassistent, Arvutiteaduse instituut, Tartu Ülikool

Juhendamine

Ayushmat Bhardwaj Soni (MSc): Neuro-hägused süsteemid binaarseks klassifitseerimiseks

Teadustegevus

Peamised uurimisvaldkonnad:

- fuzzy logic
- AutoML

DISSERTATIONES INFORMATICAE PREVIOUSLY PUBLISHED IN DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS

19. **Helger Lipmaa.** Secure and efficient time-stamping systems. Tartu, 1999, 56 p.
22. **Kaili Müürisep.** Eesti keele arvutigrammatika: süntaks. Tartu, 2000, 107 lk.
23. **Varmo Vene.** Categorical programming with inductive and coinductive types. Tartu, 2000, 116 p.
24. **Olga Sokratova.** Ω-rings, their flat and projective acts with some applications. Tartu, 2000, 120 p.
27. **Tiina Puolakainen.** Eesti keele arvutigrammatika: morfoloogiline ühestamine. Tartu, 2001, 138 lk.
29. **Jan Villemson.** Size-efficient interval time stamps. Tartu, 2002, 82 p.
45. **Kristo Heero.** Path planning and learning strategies for mobile robots in dynamic partially unknown environments. Tartu 2006, 123 p.
49. **Härmel Nestra.** Iteratively defined transfinite trace semantics and program slicing with respect to them. Tartu 2006, 116 p.
53. **Marina Issakova.** Solving of linear equations, linear inequalities and systems of linear equations in interactive learning environment. Tartu 2007, 170 p.
55. **Kaarel Kaljurand.** Attempto controlled English as a Semantic Web language. Tartu 2007, 162 p.
56. **Mart Anton.** Mechanical modeling of IPMC actuators at large deformations. Tartu 2008, 123 p.
59. **Reimo Palm.** Numerical Comparison of Regularization Algorithms for Solving Ill-Posed Problems. Tartu 2010, 105 p.
61. **Jüri Reimand.** Functional analysis of gene lists, networks and regulatory systems. Tartu 2010, 153 p.
62. **Ahti Peder.** Superpositional Graphs and Finding the Description of Structure by Counting Method. Tartu 2010, 87 p.
64. **Vesal Vojdani.** Static Data Race Analysis of Heap-Manipulating C Programs. Tartu 2010, 137 p.
66. **Mark Fišel.** Optimizing Statistical Machine Translation via Input Modification. Tartu 2011, 104 p.
67. **Margus Niitsoo**. Black-box Oracle Separation Techniques with Applications in Time-stamping. Tartu 2011, 174 p.
71. **Siim Karus.** Maintainability of XML Transformations. Tartu 2011, 142 p.
72. **Margus Treumuth.** A Framework for Asynchronous Dialogue Systems: Concepts, Issues and Design Aspects. Tartu 2011, 95 p.
73. **Dmitri Lepp.** Solving simplification problems in the domain of exponents, monomials and polynomials in interactive learning environment T-algebra. Tartu 2011, 202 p.

74. **Meelis Kull.** Statistical enrichment analysis in algorithms for studying gene regulation. Tartu 2011, 151 p.
77. **Bingsheng Zhang.** Efficient cryptographic protocols for secure and private remote databases. Tartu 2011, 206 p.
78. **Reina Uba.** Merging business process models. Tartu 2011, 166 p.
79. **Uuno Puus.** Structural performance as a success factor in software development projects – Estonian experience. Tartu 2012, 106 p.
81. **Georg Singer.** Web search engines and complex information needs. Tartu 2012, 218 p.
83. **Dan Bogdanov.** Sharemind: programmable secure computations with practical applications. Tartu 2013, 191 p.
84. **Jevgeni Kabanov.** Towards a more productive Java EE ecosystem. Tartu 2013, 151 p.
87. **Margus Freudenthal.** Simpl: A toolkit for Domain-Specific Language development in enterprise information systems. Tartu, 2013, 151 p.
90. **Raivo Kolde.** Methods for re-using public gene expression data. Tartu, 2014, 121 p.
91. **Vladimir Šor.** Statistical Approach for Memory Leak Detection in Java Applications. Tartu, 2014, 155 p.
92. **Naved Ahmed.** Deriving Security Requirements from Business Process Models. Tartu, 2014, 171 p.
94. **Liina Kamm.** Privacy-preserving statistical analysis using secure multi-party computation. Tartu, 2015, 201 p.
100. **Abel Armas Cervantes.** Diagnosing Behavioral Differences between Business Process Models. Tartu, 2015, 193 p.
101. **Fredrik Milani.** On Sub-Processes, Process Variation and their Interplay: An Integrated Divide-and-Conquer Method for Modeling Business Processes with Variation. Tartu, 2015, 164 p.
102. **Huber Raul Flores Macario.** Service-Oriented and Evidence-aware Mobile Cloud Computing. Tartu, 2015, 163 p.
103. **Tauno Metsalu.** Statistical analysis of multivariate data in bioinformatics. Tartu, 2016, 197 p.
104. **Riivo Talviste.** Applying Secure Multi-party Computation in Practice. Tartu, 2016, 144 p.
108. **Siim Orasmaa.** Explorations of the Problem of Broad-coverage and General Domain Event Analysis: The Estonian Experience. Tartu, 2016, 186 p.
109. **Prastudy Mungkas Fauzi.** Efficient Non-interactive Zero-knowledge Protocols in the CRS Model. Tartu, 2017, 193 p.
110. **Pelle Jakovits.** Adapting Scientific Computing Algorithms to Distributed Computing Frameworks. Tartu, 2017, 168 p.
111. **Anna Leontjeva.** Using Generative Models to Combine Static and Sequential Features for Classification. Tartu, 2017, 167 p.
112. **Mozhgan Pourmoradnasseri.** Some Problems Related to Extensions of Polytopes. Tartu, 2017, 168 p.

113. **Jaak Randmets.** Programming Languages for Secure Multi-party Computation Application Development. Tartu, 2017, 172 p.
114. **Alisa Pankova.** Efficient Multiparty Computation Secure against Covert and Active Adversaries. Tartu, 2017, 316 p.
116. **Toomas Saarsen.** On the Structure and Use of Process Models and Their Interplay. Tartu, 2017, 123 p.
121. **Kristjan Korjus.** Analyzing EEG Data and Improving Data Partitioning for Machine Learning Algorithms. Tartu, 2017, 106 p.
122. **Eno Tõnisson.** Differences between Expected Answers and the Answers Offered by Computer Algebra Systems to School Mathematics Equations. Tartu, 2017, 195 p.

DISSERTATIONES INFORMATICAE UNIVERSITATIS TARTUENSIS

1. **Abdullah Makkeh**. Applications of Optimization in Some Complex Systems. Tartu 2018, 179 p.
2. **Riivo Kikas**. Analysis of Issue and Dependency Management in Open-Source Software Projects. Tartu 2018, 115 p.
3. **Ehsan Ebrahimi**. Post-Quantum Security in the Presence of Superposition Queries. Tartu 2018, 200 p.
4. **Ilya Verenich**. Explainable Predictive Monitoring of Temporal Measures of Business Processes. Tartu 2019, 151 p.
5. **Yauhen Yakimenka**. Failure Structures of Message-Passing Algorithms in Erasure Decoding and Compressed Sensing. Tartu 2019, 134 p.
6. **Irene Teinemaa**. Predictive and Prescriptive Monitoring of Business Process Outcomes. Tartu 2019, 196 p.
7. **Mohan Liyanage.** A Framework for Mobile Web of Things. Tartu 2019, 131 p.
8. **Toomas Krips.** Improving performance of secure real-number operations. Tartu 2019, 146 p.
9. **Vijayachitra Modhukur.** Profiling of DNA methylation patterns as biomarkers of human disease. Tartu 2019, 134 p.
10. **Elena Sügis.** Integration Methods for Heterogeneous Biological Data. Tartu 2019, 250 p.
11. **Tõnis Tasa.** Bioinformatics Approaches in Personalised Pharmacotherapy. Tartu 2019, 150 p.
12. **Sulev Reisberg.** Developing Computational Solutions for Personalized Medicine. Tartu 2019, 126 p.
13. **Huishi Yin.** Using a Kano-like Model to Facilitate Open Innovation in Requirements Engineering. Tartu 2019, 129 p.
14. **Faiz Ali Shah.** Extracting Information from App Reviews to Facilitate Software Development Activities. Tartu 2020, 149 p.
15. **Adriano Augusto**. Accurate and Efficient Discovery of Process Models from Event Logs. Tartu 2020, 194 p.
16. **Karim Baghery.** Reducing Trust and Improving Security in zk-SNARKs and Commitments. Tartu 2020, 245 p.
17. **Behzad Abdolmaleki.** On Succinct Non-Interactive Zero-Knowledge Protocols Under Weaker Trust Assumptions. Tartu 2020, 209 p.
18. **Janno Siim.** Non-Interactive Shuffle Arguments. Tartu 2020, 154 p.
19. **Ilya Kuzovkin.** Understanding Information Processing in Human Brain by Interpreting Machine Learning Models. Tartu 2020, 149 p.
20. **Orlenys López Pintado.** Collaborative Business Process Execution on the Blockchain: The Caterpillar System. Tartu 2020, 170 p.
21. **Ardi Tampuu.** Neural Networks for Analyzing Biological Data. Tartu 2020, 152 p.

22. **Madis Vasser.** Testing a Computational Theory of Brain Functioning with Virtual Reality. Tartu 2020, 106 p.
23. **Ljubov Jaanuska.** Haar Wavelet Method for Vibration Analysis of Beams and Parameter Quantification. Tartu 2021, 192 p.
24. **Arnis Parsovs.** Estonian Electronic Identity Card and its Security Challenges. Tartu 2021, 214 p.
25. **Kaido Lepik.** Inferring causality between transcriptome and complex traits. Tartu 2021, 224 p.
26. **Tauno Palts.** A Model for Assessing Computational Thinking Skills. Tartu 2021, 134 p.
27. **Liis Kolberg.** Developing and applying bioinformatics tools for gene expression data interpretation. Tartu 2021, 195 p.
28. **Dmytro Fishman.** Developing a data analysis pipeline for automated protein profiling in immunology. Tartu 2021, 155 p.
29. **Ivo Kubjas.** Algebraic Approaches to Problems Arising in Decentralized Systems. Tartu 2021, 120 p.
30. **Hina Anwar.** Towards Greener Software Engineering Using Software Analytics. Tartu 2021, 186 p.
31. **Veronika Plotnikova.** FIN-DM: A Data Mining Process for the Financial Services. Tartu 2021, 197 p.
32. **Manuel Camargo.** Automated Discovery of Business Process Simulation Models From Event Logs: A Hybrid Process Mining and Deep Learning Approach. Tartu 2021, 130 p.
33. **Volodymyr Leno.** Robotic Process Mining: Accelerating the Adoption of Robotic Process Automation. Tartu 2021, 119 p.
34. **Kristjan Krips.** Privacy and Coercion-Resistance in Voting. Tartu 2022, 173 p.
35. **Elizaveta Yankovskaya.** Quality Estimation through Attention. Tartu 2022, 115 p.
36. **Mubashar Iqbal.** Reference Framework for Managing Security Risks Using Blockchain. Tartu 2022, 203 p.
37. **Jakob Mass.** Process Management for Internet of Mobile Things. Tartu 2022, 151 p.
38. **Gamal Elkoumy.** Privacy-Enhancing Technologies for Business Process Mining. Tartu 2022, 135 p.
39. **Lidia Feklistova.** Learners of an Introductory Programming MOOC: Background Variables, Engagement Patterns and Performance. Tartu 2022, 151 p.
40. **Mohamed Ragab.** Bench-Ranking: A Prescriptive Analysis Approach for Large Knowledge Graphs Query Workloads. Tartu 2022, 158 p.
41. **Mohammad Anagreh.** Privacy-Preserving Parallel Computations for Graph Problems. Tartu 2023, 181 p.
42. **Rahul Goel.** Mining Social Well-being Using Mobile Data. Tartu 2023, 104 p.

43. **Anti Ingel.** Algorithms using information theory: classification in brain-computer interfaces and characterising reinforcement-learning agents. Tartu 2023, 142 p.
44. **Shakshi Sharma.** Fighting Misinformation in the Digital Age: A Comprehensive Strategy for Characterizing, Identifying, and Mitigating Misinformation on Online Social Media Platforms. Tartu 2023, 158 p.
45. **Kristiina Rahkema.** Quality Analysis of iOS Applications with Focus on Maintainability and Security Aspects. Tartu 2023, 182 p.
46. **Ivan Slobozhan.** Studying Online Social Media Engagement in CIS Countries during Protests, Mass Demonstrations and War. Tartu 2023, 81 p.
47. **Nurlan Kerimov.** Building a catalogue of molecular quantitative trait loci to interpret complex trait associations. Tartu 2023, 248 p.
48. **Pavlo Tertychnyi.** Machine Learning Methods for Anti-Money Laundering Monitoring. Tartu 2023, 117 p.
49. **Abasi-amefon Obot Affia.** A Framework and Teaching Approach for IoT Security Risk Management. Tartu 2023, 180 p.
50. **Raimond-Hendrik Tunnel.** Video Game Design and Development Bachelor's Curriculum for Estonia. Tartu 2024, 137 p.
51. **Ahto Salumets.** Bioinformatics analysis of various aspects in immunology. Tartu 2024, 198 p.
52. **Mohammed Abdulhameed Shaif Ali.** Deep Learning Methods for Cell Microscopy Image Analysis. Tartu 2024, 143 p.
53. **Pille Pullonen-Raudvere**. Foundations of Efficient and Secure Algorithm Development for Secure Multiparty Computation. Tartu 2024, 265 p.
54. **Marili Rõõm.** Multiple approaches to learners' success and factors affecting it in computer programming MOOCs. Tartu 2024, 170 p.
55. **Shivananda Rangappa Poojara.** Design and Orchestration of Scalable, Event-Driven Serverless Data Pipelines for Internet of Things (IoT) Applications. Tartu 2024, 172 p.
56. **Hassan Abdulgaleel Hassan Salim Eldeeb.** Empowering Machine Learning Pipelines with Automated Feature Engineering. Tartu 2024, 121 p.