TITLE OF THESIS

Development of a causal machine learning model for the diagnosis of African swine fever

by

Steven Lububu

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy: Informatics

Discipline: Information Technology
in the Faculty of Informatics and Design
at the Cape Peninsula University of Technology

District Six, Cape Town.

Date submitted: 25/08/ 2025

CPUT copyright information

The thesis may not be published in parts (in scientific or technical journals) or (as a monograph) unless permission has been granted by the university.

DECLARATION

I, Steven Lububu, declare that the content of this thesis represents my own independent work, and that the thesis has not previously been submitted for academic examination for any qualification. Furthermore, it reflects my own views and not necessarily those of Cape Peninsula University of Technology.

Signed Date: 25/08/ 2025

ABSTRACT

ABSTRACT

This study investigates the causal relationship between African swine fever (ASF) viral load and disease severity in domestic and wild pigs using machine learning models. A causality model with linear regression and random forest regressor was developed to analyse ASF transmission dynamics and symptom severity. The linear regression model achieved an R^2 value of 83.68% with an MAE of 1.27, while the random forest model achieved an R^2 value of 58.10% with an MAE of 1.52, confirming strong predictive performance. The results highlight the effectiveness of biosecurity, surveillance and culling measures in containing ASF and emphasize evidence-based policy making for disease control. This study provides actionable insights for veterinarians, farmers and policy makers, contributing to ASF risk management and prevention strategies. Future research should integrate AI-driven real-time surveillance and genetic analysis to improve ASF outbreak prediction and global containment measures.

Keywords: African swine fever (ASF), ASF diagnosis, causal inference, machine learning and causality model.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to **Professor Tembisa Ngqondi, Professor Masilonyane Mokhele, Dr Tabisa Ncubukezi and Veda Naidoo** for their unfailing support, guidance and encouragement during the development of this work. Their contributions were invaluable to the successful completion of this thesis.

DEDICATION

I would like to dedicate this work to my fiancée, Monique Green. Thanks for your support.

PUBLICATIONS FROM THIS RESEARCH

1 *Lububu, S. and Kabaso, B. 2023. A Systematic Literature Review on Machine Learning and Laboratory Techniques for the Diagnosis of African swine fever (ASF). Paper presented at the 2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD) (pp. 1-8). IEEE.*

2 *Lububu, S. and Kabaso, B. 2023. November. Reflections on Feature Engineering and Design Using Causal Machine Learning (CML) for African Swine Fever (ASF) Diagnosis. Paper presented at the 2023 International Conference on Artificial Intelligence and its Applications (icARTi 2023) (pp. 21-30).*

3 Lububu, S. and Twum-Darko, M., 2024. Comparative analysis of machine learning algorithms to improve the diagnosis of African swine fever. *International Journal of Business Ecosystem & Strategy (2687-2293)*, *6*(5), pp.121-137.

4 Lububu, S. and Twum-Darko, M., 2024. Identification of diagnostic methods for African swine fever: A systematic literature review. *International Journal of Business Ecosystem & Strategy (2687-2293)*, *6*(6), pp.187-202.

5 Lububu, S., 2025. Causality with machine learning using the Lububu method for the diagnosis of African swine fever (ASF). *International Journal of Business Ecosystem & Strategy (2687-2293)*, *7*(2), pp.184-206.

TABLE OF CONTENTS

LIST OF FIGURES and TABLES

ABBREVIATIONS AND ACRONYMS

AI:	Artificial Intelligence
ANN:	Artificial Neural Network
ASF:	African Swine Fever
AV:	Actual Value
BN:	Bayesian Network
BR:	Bayesian Regression
CI:	Causal Inference
CML:	Causal Machine Learning
CPCR:	Convection Polymerase Chain Reaction
CR:	Cross Validation
DP:	Deep Learning
DT:	Decision Tree
EHRS:	Electronic Health Records
ELISA:	Enzyme-Linked Immunosorbent Assay
ES:	Estimand
EURL:	European Union Reference Laboratory
IBE:	Inference to the Best Explanation
IFA:	Indirect Fluorescent Antibody
LR:	Linear Regression
MAE:	Mean Absolute Error
ML:	Machine Learning
MSE:	Mean Squared Error
PCR:	Polymerase Chain Reaction
PV:	Predicted Values
QA:	Quality Assessment
R^2:	R-Squared score
RMSE:	Root Mean Squared Errors
RQS:	Research Questions
SCMs:	Structural Causal Models
SLR:	Systematic Literature Review
UDMLECI:	Understanding the Disease, ML Expertise and Causal Inference

CLARIFICATION OF TERMS

The root mean squared error (RMSE) is a metric often used in regression analysis and predictive modelling to assess the performance of a model's predictions by measuring the difference between the predicted values and the actual observed values. The RMSE is often used for regression tasks due to its simplicity and effectiveness in evaluating the overall performance of the model. In this study, the RMSE is applied to assess the overall performance of the model. $R^{\wedge 2}$ provides a quick indication of how well the model is performing. R-squared ($R^{\wedge 2}$) measures the proportion of variance in the dependent variable (target) that is predictable by the independent variables (characteristics) in the model. It represents the goodness of fit of the model. Higher R^2 values indicate that more variance in the dependent variable can be explained by the independent variables.

- **$R^{\wedge 2}$: = 1.0** means that the model predicts the dependent variable perfectly.
- **$R^{\wedge 2}$: = 0.0** indicates that the model does not explain the variability of the dependent variable.

CHAPTER 1: INTRODUCTION

1.1 Organisation of the chapter

The first chapter (Chapter 1) provides an overview of the entire work. It discusses the background of the research in section 1.2, African Swine Fever in section 1.3, ASF impacts on the pig industry in section 1.4, the importance of developing an effective diagnosis in section 1.5, the research problem in section 1.6, the problem statement in section 1.7, the aim in section 1.8, the objectives in section 1.9 and the research questions in section 1.10. It also discusses the scope and limitations of the thesis in section 1.11, significance of the study in section 1.12, the chapter summary in section 1.3 and the thesis structure in section 1.14.

1.2 Background

African swine fever is a highly contagious and fatal viral disease that affects domestic and wild pigs. It is harmless to humans but can destroy pig herds. It has a significant economic and social impact on the pig industry (Gallardo et al., 2019; Dixon et al., 2020). The following section explains ASF, its impact on the pig industry and the importance of developing effective diagnostic methods.

1.3 African swine fever

According to Badgeley et al. (2019), ASF is caused by a virus belonging to the Asfarviridae family. The disease is characterised by a wide range of clinical signs in affected pigs, including high fever, loss of appetite, internal bleeding and a high mortality rate (up to 100% in some cases). Sánchez et al. (2019) also explain that ASF can be transmitted through direct contact between infected and healthy pigs, contaminated equipment, feed, ticks and contact with wild pigs. Furthermore, Sánchez et al. (2022) outline that, unlike some other pig diseases, there is no effective vaccine and no specific treatment for ASF. Control measures usually include culling infected animals and implementing strict biosecurity measures to prevent further spread.

1.4 ASF Impacts on the pig industry

Richens et al. (2020) and Sánchez et al. (2022) argue that ASF has a significant impact on the pig industry, both economically and socially.

- **Economic impact:** The disease can cause massive losses in pig herds and affect pork production. During outbreaks, whole herds are often culled to prevent further spread. This can lead to a shortage of pork products and thus to price increases.
- **Trade restrictions:** Many countries impose trade restrictions on pork and pork products from ASF-affected regions. This can lead to the loss of international markets and reduced export opportunities for the affected countries.
- **Livelihoods and rural communities:** The pig industry is a source of income for many farmers and rural communities. ASF outbreaks can destroy livelihoods and affect local economies.

1.5 Importance of developing effective diagnostic methods

According to Gallardo et al. (2019), the development of effective diagnostic methods is crucial for early detection, disease surveillance, trade facilitation and research efforts to control and eventually eradicate ASF. Furthermore, forth et al. (2023) stated that these methods play a critical role in mitigating the impact of ASF on the swine industry and ensuring the livelihood of those involved in pork production. In the study, effective diagnostic methods helped to identify infected animals at an early stage so that immediate action can be taken to prevent further spread. In addition, diagnostic methods play a critical role in epidemiologic studies, which enabled the researcher to understand ASF transmission patterns, risk factors and the effectiveness of control measures.

1.6 Research problem

According to Badgeley et al. (2019), lack of accuracy and misdiagnosis have proven to be major challenges for ASF diagnosis. Furthermore, Richens et al. (2020) mentioned that errors in ASF diagnosis are a major challenge for the pig industry. Sánchez et al. (2022) also pointed out that laboratories are not able to separate correlation from causality, leading to ASF misdiagnosis. However, many other researchers propose the use of machine learning for ASF diagnosis, but the problem of inaccurate diagnosis remains (Wang et al., 2021; Pikalo et al., 2021).Causal machine learning was introduced to overcome these challenges by providing an understanding of the causal relationships between variables, an understanding which is crucial for tackling complex problems such as disease spread and control (Achenbach et al., 2015; Pearl et al., 2016).

1.7 Problem statement

This study identifies critical gaps in current diagnostic methods for African swine fever (ASF) and highlights inaccuracies that often lead to misdiagnosis. It also notes that existing machine learning models have problems with causal inference, limiting their effectiveness in establishing cause-and-effect relationships between ASF symptoms and the virus. To address these issues, the study aims to develop a more reliable diagnostic model by combining epidemiologic data with advanced machine learning techniques. The goal is to overcome the biases and limitations of current methods and develop a more robust and accurate approach to ASF diagnosis.

1.8 Research aim

The aim of this study is to develop a causal machine learning (CML) model that can create links between the ASF virus and the symptoms of the disease.

1.9 Research objectives

The four main objectives that correlate with the research questions of this study are:

1. Identification and review of existing solutions for ASF diagnosis.
2. Identification of ASF trends analysis that create links between virus and clinical symptoms.
3. Develop a CML model that can establish relationships between the ASF virus and the symptoms of the disease.
4. Evaluation of the CML model in terms of its performance

1.10 Research questions

This section outlines the specific research questions that this study aims to address.

Main research question

How can a causal machine learning model be developed for ASF diagnosis?

Sub-questions

1. What solutions are available on Machine Learning and laboratory techniques for ASF diagnosis?
2. What are the ASF trends analysis that can facilitate cause-effect relationships?
3. How can a CML model be developed to create the links between the ASF virus and the symptoms of the disease?
4. How can the CML model be evaluated in terms of its performance?

1.11 Scope and limitations

The scope of this thesis is limited to the development and application of the CML model for ASF diagnosis. It specifically addresses the integration of causal inference and ML computing and focuses on the development of a CML model capable of establishing a relationship between ASF virus and symptoms. The study investigates the use of causal inference using machine learning techniques for diagnosis. The study also identifies possible causal relationships between ASF viruses and symptoms. The study has several limitations that may affect the scope and applicability of the results. Firstly, the literature review may not include all causal machine learning and laboratory solutions, especially those that are proprietary, new, or insufficiently documented. Secondly, the challenges identified, such as inadequate ASF diagnosis in some laboratories in some countries, may not reflect conditions in other laboratories in other countries and therefore may not be universally relevant. Thirdly, the use of online data or publicly available data harbours the risk of distortions or inaccuracies. These data may not reflect all facets of the ASF dataset and may omit significant information about the complexity of the model. These limitations must be taken into consideration when interpreting the findings of the study and considering their contribution to the diagnosis of ASF. Finally, the practical effectiveness of the proposed model for ASF diagnosis has been experimentally tested, which provides some confidence in its performance in practice. However, the results may be subject to regional biases influenced by the laboratory, opinion, experience and knowledge of ASF diagnostic studies, which may limit the generalisability of the results to other laboratories with different characteristics or operational requirements.

1.12 Significance of the study

The significance of this study lies in the fact that it enables the causal relationship in ASF diagnosis to be applied in terms of time and methodology using ML. Improved diagnosis of African swine fever (ASF) can bring significant benefits to the pig industry, animal health and disease control. This work has significant implications for the practical application of the CML model for ASF diagnosis. The development of a CML model using causality theory methodology is a practical solution to the major challenges in ASF diagnosis as it improves accuracy, precision, enables early detection and reduces economic losses. Theoretically, this study applies causality theory to a machine learning predictive model to improve its prediction, accuracy and precision in diagnosis. Methodologically, the study contributes to a new, experimentally tested CML model for ASF diagnosis, which represents an advance in the study of causal inference and ML algorithms. In addition, the performance of the developed CML model was compared with another algorithm to guide the future application of the CML model for ASF diagnosis and set a precedent for research in this area. Beyond practical applications, this research improves our scientific understanding of ASF. It also provides a multidimensional approach to understanding, preventing and managing the disease.

Based on the significance provided, the study is deductive for the following reasons:

1. **Starting from theory:** the study explicitly applies causality theory to develop a predictive model through machine learning. This means that it starts from an established theoretical framework about causal relationships in ASF diagnosis.

2. **Hypothesis testing:** By using theory as the basis for developing and testing a Causal Machine Learning (CML) model, the study moves from general theoretical principles to specific predictions. The experimental tests and the comparison with another algorithm indicate that the study is designed to test the hypotheses derived from the theoretical framework.

3. **Applying a top-down approach:** Deductive research usually follows a top-down approach — from general theory to specific observations or experiments. In this case, the study uses

causal theory to guide the development of the model, which is then experimentally tested for its accuracy, precision, and practical application in diagnosing ASF.

According to Gallaire, Minker and Nicolas (1984), a deductive approach begins with a broad theory or hypothesis. From this theory, specific predictions are derived that determine the design of experiments or studies. In addition, Pandey (2019) has discussed that the collected data is then analysed to determine whether the results are consistent with these predictions, allowing researchers to confirm, refine, or disprove the original theory. Thus, the design and methodology of the study are consistent with a deductive rather than an inductive approach.

1.13 Chapter Summary

Chapter 1 provides an overview of the research, focusing on ASF, a highly contagious and fatal swine disease that has a significant economic and social impact on the pig industry. The chapter emphasises the importance of developing effective ASF diagnosis due to the current diagnostic inaccuracies which often lead to misdiagnosis. It outlines the research problem, problem statement, aim, objectives and research question and emphasises the need for a more reliable causal machine learning (ML) model. Existing ML models struggle to make accurate causal inferences between ASF symptoms and the virus, limiting their effectiveness. The chapter also discusses the complexity of applying causal ML to ASF diagnosis, which requires the integration of epidemiologic data and machine learning techniques. In addition, it addresses the limitations of current approaches, which may be biased or limited by specific laboratory and ML data and attempts to develop a more general and robust model for improved ASF diagnosis.

1.14 Thesis Structure

Chapter 1 provides an overview of the research and sets the framework for the study. Chapter 2 covers the literature review, discusses the background of the research, and includes Research Objective 1, which is to review and identify existing machine learning (ML) solutions and laboratory techniques for ASF diagnosis and evaluate their methods, strengths, and weaknesses. Chapter 3 explains the research philosophy and methodology used in the study. Chapter 4 presents the structuring and evolution of the causal machine learning (CML) model. Chapter 5 discusses the findings and results of the study. Chapter 6 discusses each of the research objectives in detail. Finally, Chapter 7 concludes the thesis and makes suggestions for future applications of the model. The next chapter (Chapter 2) presents the literature reviewed for this study.

CHAPTER 2: LITERATURE REVIEW

2.1. Organisation of the chapter

This chapter begins with an organisation of the chapter in section 2.1, the introduction to the review in section 2.2, an overview of ASF in section 2.3, the impact of ASF on pigs in section 2.4, the epidemiology in section 2.5, the existing methods and techniques for diagnosing the virus in section 2.6, the limitations and challenges associated with current ASF diagnostic methods in section 2.7, the significance of causal machine learning for ASF in section 2.8, background information on the application of ML in disease diagnosis in section 2.9, A brief historical overview in section 2.10, the current status in section 2.11, theoretical foundations in section 2.12, Machine learning in section 2.13, Causality Model for diagnosis in section 2.14, the reflection on the SLR in section 2.15, gaps and limitations in existing research in section 2.16 and chapter summary in section 2.17.

2.2. Introduction

African swine fever (ASF) is a highly contagious and devastating viral disease that affects domestic and wild pigs. ASF was detected in Africa particularly in Kenya in 1921. This disease can create various challenges to the farm industry especially those who are dealing with pigs. The ASF disease can cause economic losses and destabilise the food industry. On one hand, researchers are trying to use CML to diagnose ASF and implement preventive techniques. The role of a CML is to establish the cause-effects relationship between ASF virus and the symptoms. The CML is about inferring causal relationships from observation of ASF datasets. However, the existing statistic methods relies on correlation and association with ASF datasets. Consequently, the association and correlation process can provide inaccurate diagnostic of ASF virus.

This literature review clarifies the use of CML methods to generate predictive methods cable to establish relationships between factors. This review is a knowledge-based that shapes the research objectives and questions of the study. Also, the review offers a holistic overview of the combination of ASF context and causal ML techniques. To this end, a CML shows how causal ML techniques enable our understanding of ASF and how to implement preventive measures. In this review the current state of ASF is discussed as well as the gaps and challenges.

2.3. Overview of African Swine Fever

African swine fever (ASF) is a fatal disease that is spread between domestic and wild pigs. The disease is generated by ASF virus family called Asfarviridae. The ASF virus lives in DNA and blood cells of the animals. The ASF virus does not affect human, but it affects pigs only (Sánchez-Vizcaíno et al., 2019; Olesen et al., 2024).

2.3.1 History

ASF was detected in early 20th century. In 1950, the ASF virus was spread to Europe.
In 1980, the ASF virus reached Asia, America and the URSS (Sánchez-Vizcaíno et al., 2019). Researchers and scientists tried to stop and manage the spread of ASF virus by imposing strict control measures of isolation and limitations of movement from country to country, but it was unsuccessful, and the virus reached the rest of the world (Franzoni et al., 2023; Kwon et al., 2024; Song et al.,2024).

2.3.2 ASF Factors

a. **Infection:** the use of infected equipment and lack of strict security control can enable the spread of ASF virus among pigs.
b. **Infections aspects:** direct contact with infected pigs, environment etc.
c. **Vaccination:** ASF does not have a vaccination. The methods used include isolation of pigs, quarantine and management control.

2.4. Impact of ASF on the Pig Industry

ASF spreads through domestic and wild pigs but does not affect humans. For instance, Haines et al. (2013) discussed that ASF has a severe impact on the pig's population worldwide. Gallardo et al.(2015b) also said that ASF can cause economic losses and disrupts the international trade on the long-term.

Saka et al. (2010) outlined in his research some preventive techniques and control measures of ASF such as biosecurity and isolation. These techniques can save and protect pig's population as there is no appropriate treatment. Furthermore, Balestreri et al. (2024) discussed that ASF has a high mortality rate in contaminated pigs up to 100%. A trade restriction can be imposed in countries or regions where ASF is identified. Export bans can severely affect a country's ability to sell its pork products abroad, leading to a surplus in the domestic market and falling prices (Gallardo et al., 2015b; Mehinagic et al.,2024).

According to Saka et al. (2010), the pig industry is very sensitive to fluctuations in supply and demand. ASF outbreaks can lead to price fluctuations in the pig market, affecting both producers and consumers. ASF outbreaks may discourage investment in pig farming because of the disease risk and possible financial losses. Farmers may downsize their pig herds or abandon the industry altogether. Leivers et al. (2023) outline that reduced supply due to culling and trade restrictions may lead to an increase in pork prices in affected regions. This can impact consumers and make pork less affordable. ASF affects not only pig farmers, but also industries associated with pig farming, such as feed production, transport and processing. Job losses and financial difficulties may extend beyond pig farming. Prolonged or repeated ASF outbreaks may lead to structural changes in the pig industry.

2.5 The Epidemiology, Transmission and Clinical Symptoms

ASF is a devastating disease that has an impact on pigs either domestic or wild (Buschet et al., 2021). The ASF virus does not have a vaccine for treatment. The primary control measures involve pigs' separation or isolation, application of strict biosecurity measure and an educational training or awareness to stop the spread of ASF (Leivers et al., 2023; Wang et al., 2023). Below are the explanations about the epidemiology, transmission and clinical signs of ASF.

2.5.1 Epidemiology

ASF virus belongs to the Asfarviridae family. The ASF virus is resistant and cable to survive in the area for decades. It is resilient disease that can live in pork and blood (Gomez-Vazquez & Martinez-Lopez, 2023; Penrith et al., 2023).

1. **In domestic pigs:** the ASF virus can be transmitted through close contact with infected pigs. The spread of ASF virus can happen through the transportations, feed and the insect's bit.
2. **In wild pigs:** the male pigs or boars are the carriers and can spread the ASF virus to the female pigs and piglets.
3. **Outbreak:** ASF is spread in different parts of Africa, Asia, Europe and America.

2.5.2 Transmission

Below are the different ways of transmissions:

1. **Close interactions:** close interactions between contaminated pigs may cause ASF virus among pigs.
2. **Secondary exposure:** secondary exposure includes the infected environments
3. **Bloodsucking insects:** this includes the biting bugs, biting pests and stinging pests.

4. **Garbage food:** this includes the wasted feeding (Penrith et al., 2023).

2.5.3 Clinical signs

ASF can manifest itself in various clinical forms ranging from acute and highly fatal to subacute or chronic (Gomez-Vazquez et al., 2023). Clinical signs include:

1. **High fever:** Sudden high fever is common in infected pigs.
2. **Depression and lassitude:** The pigs become lethargic and may have difficulty standing or walking.
3. **Lack of appetite:** contaminated pigs do not have appetite.
4. **Respiratory Effects:** Coughing, nasal discharge and breathing problem.
5. **Digestive effects:** Diarrhoea, vomiting and stomach.
6. **Skin lesions:** Skin redness, skin ulcers and bleeding occur in some pigs.
7. **Miscarriages:** abortions can happen to pregnant sows.
8. **Mortality:** the mortality rate is very high to infected pigs, often to 100%.

2.6. Existing methods and techniques for ASF diagnosis

Railey et al. (2023) and Yang et al. (2023) discussed some existing methods used to diagnose ASF such as clinical observations and laboratory tests. In practice, there is a combination of clinical observations and laboratory tests used to diagnose and control ASF. Also, Morelle et al. (2023) and Zuo et al. (2023) outlined few existing methods such as clinical signs, observation, assessment, separation, tests, biosecurity, and surveillance.

Often, clinical observation can offer evidence of ASF occurrences. In most of the cases, infected pigs show the following signs including fever, lack of appetite, stomach run and respiratory problems (Railey et al., 2023). According to Morelle et al. (2023), most of the laboratory tests use polymerase chain reaction (PCR) to diagnose ASF. This existing method is frequently used to detect ASF virus that is hiding in blood cells. The PCR is rapid and provide accurate results, they said.

LAMP (Loop-mediated isothermal amplification) is an ASF diagnostic technique that is easy and quick to detect the virus (Zuo et al., 2023). According to Zuo et al. (2023), there is another technique named ELISA (enzyme-linked immunosorbent assays). The ELISA tests for ASF diagnostics evolve a combination of PCR diagnostics for accuracy. The process requires more skilled people to use it.

2.7. Limitations and Challenges of Current ASF Diagnostic Methods

Accuracy (True Positive Rate), precision (True Negative Rate), interference (False Positives), sample bias (Collection errors), quality control, efficacy and safety are the key-limitations and challenges affiliated with ASF diagnostics methods (Bru et al., 2023; Milton et al., 2023).
In addition, same laboratories lack resources to diagnose ASF with accuracy. For instance, Milton et al. (2023) outlined that the diagnosis tests such as blood tests are not accurate in the initial stage of ASF virus contamination because the virus is still in the development stage. Sometimes, accuracy or precision results can happen, but it is incorrect. Interference with other pigs' bacterial microbes can cause a range of other ASF viruses that can lead to misdiagnosis or a confusion in diagnosis (Bru et al., 2023).

In most of the cases, the diagnosis of ASF virus requires niche laboratories and well-educated employees. Lack of appropriate laboratories and well-trained personal can delay the diagnosis process (Founta et al., 2023). Another limitation is lack of finance to purchase equipment and train employees. Resource limitations is a big challenge particularly for underdevelopment countries or regions. Another limiting factor is based on transportation of collected samples from infected regions to diagnostic laboratory.

2.8. The Significance of CML for the Diagnosis of ASF

A CML uses algorithms to establish cause-effect relationships between factors (Balke & Pearl, 1994; Pearl, 2019; Richens et al., 2020). The idea is that to facilitate early and easy detection by outlining new preventing techniques of ASF virus (Ahrendt et al., 2011; Singh et al., 2014; Pearl et al., 2016; Grampurohit & Sagarnal, 2020; Richens et al., 2020; Shamji, et al., 2023). Below are the few advantages that CML offers such as easy detection, enhance pig's health, accuracy and precision in initiating causal relationships between ASF virus and symptoms.

2.9 Background on the Application of Machine Learning in Disease Diagnosis

The application of Machine Learning in medical sector has transformed the healthcare system. Systematically, the healthcare system has improved in terms of disease diagnosis and treatment (Ahsan et al., 2022; Su et al., 2022). The application of machine learning allowed early detection and treatment of various diseases such as cancer, tuberculous, ASF etc. However, the use of ML in healthcare system the privacy and ethical concerns should be considered (Beeram et al.,2023).

2.9.1 Radiology

Radiology is a medical technique that involves the use of images technology. In most cases algorithms are developed to analyse medical samples and images such as X-rays, CT scans and MRIs. These algorithms can detect anomalies and diseases. Often, the algorithm used to analyse medical images is called as Convolutional Neural Networks (CNNs). This algorithm plays an important role in improving the accuracy of image diagnosis. The CNN can automatically learn and identify patterns in healthcare images with accuracy (Kolukisa & Bakir-Gungor, 2023; Shukur & Mijwil, 2023).

2.9.2 Electronic Health Records (EHR)

EHR is a medical records tool cable to digitise patients' information and makes it available for diagnosis. Huang et al. (2023) and Kufel et al. (2023) discussed that predictive modelling with EHR data can identify a patient who is at risk of certain diseases. The predictive modelling using HER can enable early intervention and individual treatment.

2.9.3 Genomic Data Analysis

ML is cable to diagnose and process genetic data and detect genetic patterns linked with the disease. According to Huang et al. (2023) and Shukur & Mijwil (2023), ML based genetic diagnosis can provide a personalised medicine and treatment to a patient.

2.9.4 Predictive Disease Modelling

ML models can predict disease outbreaks. The models can trend and analyse different data samples such as social media, weather forecasts and historical disease datasets (Khan et al., 2022; Kolukisa & Bakir-Gungor, 2023).

2.9.5 Drug Discovery and Development

ML can also be used to diagnose many datasets of chemicals containing and their combination of with biological aspects. According to Khan et al. (2022) and Beeram et al. (2023), artificial intelligence modelling can enable the prediction of drug patients and by determining potential side effects.

2.9.6 Natural Language Processing (NLP)

The NLP modelling can be used to clinical records, research and healthcare review for extracting needed information or treatment. Zhang et al. (2022) outlined that NLP deals with decision making for the prediction and treatment process.

2.9.7 Challenges and Ethical Considerations

Huang et al. (2023) and Shukur & Mijwil (2023) pointed out few challenges related to the use of ML to diagnosis including privacy concerns, bias and the seek for validation of policies and procedures.

2.10 A brief historical overview of the field of research

This study is an experimental design using CML for ASF diagnosis. The CML modelling combines algorithms with causal inference to improve the ASF diagnosis. Bellow section provides an overview of milestones and development in this study.

2.10.1 Important milestones

1. **Information acquisition:** the data acquisition was the starting point for this study. The data acquisition involves medical history, pig's individual genetic, external factors that affects health or ecosystem functions and past records or trends.
2. **Selection of key data:** this involves identifying important variables
3. **Regression analysis Techniques:** this step involves modelling relationships between variables. The process is about using ML techniques to detect causal effects.
4. **Categorising ML Techniques:** this involves the selection of appropriate ML models. In this study the researcher used linear regression, logistic regression and random forest.
5. **Cross-validation:** the cross-validation assess the performance of a ML model. In this study, the research used recall, MSE and F1-score.

2.10.2 Crucial Milestone:

i. **Data platforms:**

The researcher combined different data sources to gain a deeper understanding of ASF information such laboratory tests, google libraries and medical research.

ii. **Transparency and clarity:**

the transparency and clarity are so important in the process of cause-effects relationships (Oviedo et al. (2022). The process led to understanding and transparency in decision-making.

iii. **Monitoring the process:**

This step involves the surveillance and analysis of ASF datasets. In terms of ASF virus, the monitoring process allows early detection of the virus to prevent the spread of infection (Javaid et al., 2023; Kucharski et al. (2023).

2.11. Status

The laboratory test is the principal ASF diagnostic method (Richens et al., 2020). However, the laboratory test is slow and encounters multiple diagnostic errors (Sánchez et al., 2022). Therefore, a CML method has been suggested as an alternative possibility to allow diagnostic accuracy and precision (Pearl et al., 2019).

2.12 Theoretical Foundations

The theoretical foundation of this study includes the creation and analysis of a CML model capable of estimating cause-effects between ASF virus and symptoms. The research questions and objectives of this study form the basis theoretical framework.

2.13 Machine Learning (ML)

Machine learning has been described as subset of artificial intelligence (AI). It enables the learning based on the knowledge (Mahesh, 2020). ML algorithms can learn the patterns, analyse the data and decide, he said. Three types of ML exist such as supervised learning, unsupervised learning and reinforcement learning (Thakur & Sharma, 2016). The supervised learning can train the data based on labelled results. Unsupervised learning analyses hidden data with no labelled outcomes. While a reinforcement learning can learn upon errors and analyses the environmental factors.

In this research, ML has been used to process ASF datasets, then estimates the cause-effect relationship between ASF virus and symptoms. Also, the role of ML in this study is to lower the potential of misdiagnosis and uncertainty (Halev et al., 2023; Dinhobl et al., 2023). The researcher's objective is to establish cause-effect relationships with accuracy and precision (Zhang, Su & Chen, 2021).

2.14 Causality Model for Diagnosis

Causality model estimates the difference between causality and correlation (Richens et al., 2020; Pedro Sanchez et al., 2022). Causality deals with identification of cause-effect between factors (Schölkopf, 2022; Cui & Athey, 2022). Causality in this study enables the identification of ASF virus and establishes its relationship with the symptoms of the disease. The development of the model, its understanding and interpretation are guided by CML framework.

2.15 Reflection on the Systematic Literature Review

The reflection on the Systematic literature review is based on analysis and assessment of the current research on the use of machine learning algorithms to diagnose ASF virus. The review reveals that ML algorithms can detect ASF virus at starting point with accuracy. The review also reveals that laboratory test is the most used for ASF diagnosis, but the results are always inaccurate. Dispute the promising future of the ML algorithms for ASF diagnosis, more work is needed to improve the performance of the models. Finally, the review suggested that both techniques are important for the diagnosis and the CML technique can be used as an alternative possibility to laboratory test.

2.16. Addressing Gaps and Limitations in Existing Research.

The ongoing application of laboratory test for ASF virus diagnostics is a major limitation (Richens et al., 2020; Pedro Sanchez et al., 2022). Lack of trust into current technology is also considered as another limitation. People do not trust technology and it is difficult to shift their mindsets (Pikalo et al., 2021). The key challenge with laboratory test is that it cannot establish cause-effect relationships between ASF virus and symptoms (Sanchez et al., 2022). To address this challenge, the study proposes the application of a CML model that enables cause-effect relationships between factors.

2.17. Chapter Summary

This study presented ML and laboratory test as the two techniques used to diagnose ASF disease. ML techniques were presented in this study as an alternative and improvement of the laboratory test. It can be used in conjunction with laboratory experiments to strengthen the ASF diagnostics. The ideal of using these techniques is to identify ASF and implement strict control measures to prevent the spread of the virus. The next chapter (Chapter 3) will discuss the research philosophy and methodology.

CHAPTER 3: METHODOLOGY

3.1 organisation of the chapter

This chapter sets out the research methodology and philosophy. Following this organisation and the introduction in section 3.2; the methodology follows in section 3.3, with the research philosophy covered in section 3.4 and the research design in section 3.5. Proposed new methodology in 3.6 section. Data collection is described in 3.7 section, then data analysis in section 3.8. Section 3.9 focuses on research ethics, section 3.10 on data validity, and section 3.11 on the limitations and potential challenges of the research. Reflexibility in section 3.12 and the chapter summary in section 3.12.

3.2 Introduction

Research questions and objectives require the thorough formulation of a research plan (Blaikie, 2001; Mohamed Shaffril et al., 2021). Data collection, processing and analysis are guided by a strategy. This chapter provides an overview and explanation of the research techniques and procedures utilised to conduct this study. The research philosophies and paradigms employed are also discussed. Once the paradigms have been established and discussed, the experimental design and strategy used to develop a CML model and how it was employed in this study is discussed. This chapter also sets out the ethical considerations underlying the study, the methods of data collection and the strategies for selecting participants. In addition, issues of reliability and validity are addressed to ensure the credibility and consistency of the research findings.

3.3 Methodology

Methodology is the theoretical study of the procedures used in a particular field of research. It provides a guide for understanding the ‘how’ and ‘why’ of research and is used to plan and structure research initiatives (Mullany & Stockwell, 2021). To determine the methodology for this study, the assumptions of the research philosophy, ontology, epistemology and axiology are identified, described and justified in relation to the needs of this research.

3.4 Philosophical Framework

The term 'research philosophy' refers to a collection of beliefs about the nature of the reality being studied; and the type of research philosophy used in a particular field of study is determined by the nature of the knowledge being researched (Scotland, 2012). Research philosophy has been the subject of numerous papers and dissertations in which four basic themes have been identified and described: positivism, interpretivism, pragmatism and realism (Žukauskas et al., 2018). The philosophy of this work lies within the assumptions of positivism. Positivism is a method in which one first makes assertions and then tests, refines or abandons some of them in favour of other, better substantiated assertions. Thus, Alharahsheh and Pius (2020) state that Positivism is based on the philosophical stance of the natural scientist who works with observable reality in society and derives generalisations from it. Positivism refers to the meaning of what is universally given and focuses strictly on the observation of pure data and facts, without being influenced by the interpretation of human bias (Lancaster, 2007; Scotland, 2012; Saunders et al., 2012; Creswell, 2014:8). In this sense, positivism in machine learning involves a systematic and experimental approach to developing, evaluating and improving a model based on data-driven insights, scientific principles and rigorous testing to create reliable and effective machine learning systems for ASF diagnosis.

Wyly (2014) discussed that positivist machine learning begins with the collection of data relevant to the problem at hand. This data can come from various sources such as databases, sensors or user interactions. The data is then pre-processed to ensure its quality, i.e., it is cleaned, missing values are handled, and it is formatted for analysis. Positivist machine learning involves the selection or creation of a model based on experimental evidence and scientific principles. Different algorithms and architectures are selected or designed based on their ability to adapt to the data, generalise well and show experimental performance in solving the problem. In this sense, this study used the methods best suited to the development and application of the CML for ASF diagnosis, in line with the philosophy that emphasises the practical application of concepts and theories and their usefulness in solving real-life situations.

3.4.1 Ontology

Regarding the question, 'What is reality?', when it comes to reality, ontological philosophy focuses on the differences between reality and our experience of reality and how this affects everything in our environment (Abidi, 2011). Ontology is concerned with the categories, properties and interactions between things and the nature of being, existence or reality (Crotty, 2003). An ontological position is that there is never an observable single reality of truth.

However, we can try to uncover the truth by exploring what is experimentally available from more than one perspective.

The ontological aspect of machine learning (ML) in the context of African Swine Fever (ASF) involves understanding the fundamental nature of the disease, its characteristics and its impact on pig populations. It also involves defining and structuring the knowledge. In this sense, ontology in the context of ASF involves understanding the biological aspects of the disease, including its causes, transmission mechanisms, symptoms and impact on pig populations.

Ontological aspect enables the use of knowledge-based conclusions and inferences in an ML model. Kulmanov, Smaili, Gao and Hoehndorf (2021), for example, have discussed that domain-specific knowledge can be used to narrow down the search and find optimal or near-optimal solutions faster or find better solutions. In the life sciences, domain-specific knowledge is often encoded in ontologies and in databases and knowledge bases that use ontologies for annotation. Asim, Wasim, Khan, Mahmood, and Abbasi (2018) also describe an ontology as a formal and structural way of representing the concepts and relationships of a shared conceptualization. More specifically, it can be defined as concepts, relationships, attributes and hierarchies present in the domain (Asim et al., 2018).

By encoding expert knowledge about ASF in the ontology, ML algorithms can draw conclusions, infer relationships and make informed decisions based on the structured knowledge representation. Ontology-based machine learning for ASF is an iterative process. The ontological aspect of machine learning to solve the ASF problem essentially consists of structuring and organizing knowledge about the disease in such a way that effective relationships can be established between the ASF virus and the symptoms of the disease. The main goal of this research was to develop a CML model that can establish relationships between ASF virus and symptoms. The investigated truth was about the development and application of the CML model for the ASF diagnosis, which always remains the same regardless of the number of diagnoses.

3.4.2 Epistemology

From an epistemological point of view, our focus is on How do we know that an animal has ASF? And how do we know that this symptom is the cause?

Epistemologically, our understanding of ASF diagnosis and symptom causation relies on the systematic collection and validation of evidence. To know that an animal has ASF, we rely on a combination of empirical observations, laboratory tests such as PCR and serologic tests, and pathologic examinations. According to Alharahsheh et al. (2020), epistemology deals with the discovery of factors that can be observed and measured. The factors to be observed and measured give the credibility and meaning of the data. In the context of this study, the epistemological of ML is about solving ASF issues and how knowledge about ASF is obtained, evaluated, processed and used in algorithms. The key point is to ensure the reliability and usefulness of the knowledge to create a cause-effect model capable of establishing relationships between ASF virus and symptoms. To this end, the researcher knows that the epidemiological aspect of ASF is generated by various forms of virus.

3.4.3 Axiology

Axiology is the study of values. In this study, axiology plays a role in investigating ASF. The current ASF dataset is based on the key aspects such as the epidemiology, clinical signs, transmission methods, and control strategies to reflect the value of public health, scientific aspect and economic sustainability. The data also outlines the ethical challenges of ASF management, especially the culling of contaminated pigs, where animal welfare should be balanced with socio-economic and epidemiological concerns.

Table 3.1 presents the philosophical assumptions accepted for this study.

Table 3.1: Adapted Research Philosophy

Concept	**Clarification**
Ontology	The ontological aspect of machine learning (ML) in the context of African Swine Fever (ASF) involves understanding the fundamental nature of the disease, its characteristics and its impact on pig populations such as cause of contaminations, ASF symptoms and virus.
Epistemology	The epistemology of ASF is based on the understanding of the disease and its symptoms. The understanding of ASF focuses on the diagnosis of the disease such as the observation, laboratory tests and pathological examinations. The dataset provided reflects these principles by presenting structured epidemiologic and clinical data, including incubation periods, mortality rates, transmission routes, and symptomatic variation in domestic and wild boar. These results improve the understanding of ASF progression and facilitate evidence-based surveillance and control measures.

	By adhering to epistemological principles, ASF research ensures that results are not only descriptive but also explanatory and contribute to predictive modelling and effective disease containment strategies, ultimately facilitating the containment and control of ASF transmissions.
Methodology	From a methodological perspective, the process involves collecting data, preparing it for analysis, identifying relevant characteristics, selecting a suitable model, followed by training the model, evaluating it and finally implementing it.
Axiology	Axiology, the study of values, is fundamental to ASF research and informs its ethical, economic and scientific priorities. The dataset reflects a strong commitment to scientific rigor, public health, and economic stability by focusing on epidemiology, clinical signs, modes of transmission, and control measures. The inclusion of clinical and biological aspects emphasizes the value placed on objectivity and accuracy in disease investigation. ASF research also balances animal welfare concerns with the need for effective containment strategies, such as culling and biosecurity measures. By emphasizing evidence-based decision-making, the dataset contributes to better disease management and prevention measures. The ethical implications of disease control, particularly the trade-off between economic sustainability and humane treatment, underscore the broader values that guide ASF research. Ultimately, this approach ensures that the results of ASF research are scientifically sound, ethically defensible and practically applicable to improve disease control and agricultural resilience.
Positivism	Positivist research is a scientific approach that relies on observable, measurable and objective data to explain phenomena. It is based on the belief that reality is independent of human perception and can be investigated using empirical evidence and logical reasoning (Ali, 2024). The researcher in this paradigm uses a quantitative method, such as experiments, to test hypotheses and establish causal relationships between viruses.

3.5 Research Design

3.5.1 Conceptual Framework

The CML model was estimated to identify causal relationships using a small ASF dataset for diagnosis. Within this conceptual framework, the following steps are performed: The observed input includes ASF data collected from the European Union Reference Laboratory for African Swine Fever (EURL). The data were prepared and pre-processed. After data pre-processing, the pre-processed data were filtered using the data extraction procedure. The selected categorical features were then converted into numbers and used to create a learning model. The model was trained and tested based on the correctness of the data. RMSE was used to assess the model's performance. If the performance of the CML model is satisfactory, it is used and deployed. If not, the CML model needs to be improved.

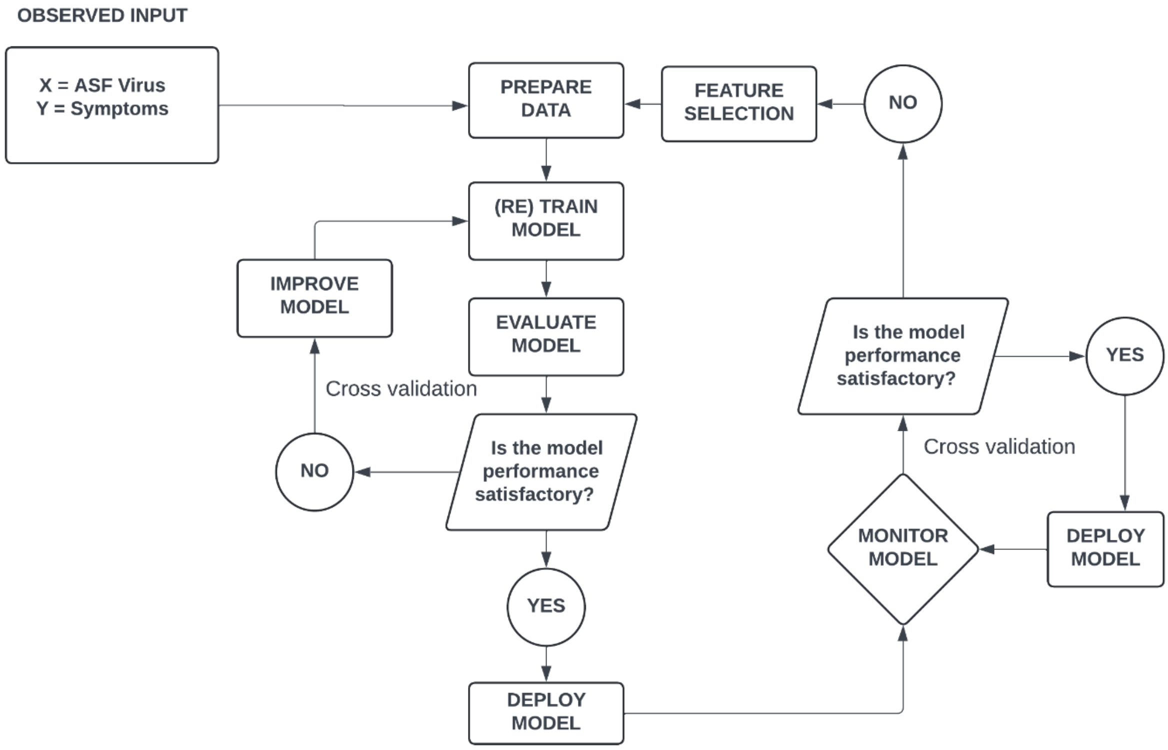

Figure 3.1: Conceptual Framework

3.5.2 Overall Research Design or Approach used in this Study

This research is categorised as applied research as it focuses on solving practical problems in the real world, namely ASF diagnosis; and it involves using various technologies in developing a new method, namely using CML model for ASF diagnosis. Furthermore, this study is a type of correlational research, as it does establish relationships between the ASF virus and the symptoms. The CML model finds the relationships between the ASF virus and the clinical symptoms.

3.5.3 Nature of this Research

This study lies between the two study frameworks of applied research and basic research. Basic research is concerned with deriving a generalised solution rather than solving a real-world problem. The pursuit of a fundamental understanding is the focus, while the application focus is low. Applied research is about solving problems in context by developing innovative solutions that are better than existing solutions. The quest to understand the fundamentals is low and the utility aspect is high (Frazer & Stokes, 1997). In this study, the focus is on the use of technologies rather than the necessary background knowledge from pure study.

3.5.4 Research strategy

A research strategy is a plan or an approach that is chosen to address a specific question or problem. It involves various techniques and methods used for collecting, analysing and interpreting the data (Bell, 2020). This study applied an experimental design as a research strategy. According to Greenhil et al. (2020) and Veli et al. (2021), an experiment design can be used for testing hypotheses and draw conclusions. In the same context, Miller et al. (2020) discussed that an experiment design is a way of solving problems by creating new ideas. This study is about creating new ideas and developing new artifacts to solve problems related to ASF disease. In support of this study, Sallam (2023) outlined that the development of an artifact must seek for answers to the problems. To this end, this study can create an artifact and test it.

3.5.5 Experimental design

The experimental design approach was used as a lens to develop, test, deploy and evaluate the causality model as it ensures a systematic, scientific approach at every stage.

I. **Design and development:** the experimental design defined the cause-effect relationship between symptoms and virus. It also defined the development of robust model features and algorithms for ASF detection.

II. **Test:** the causality model was tested for its capabilities to infer causality but not a simple association of factors. The following techniques were used for testing such as cross-validation and counterfactual analysis.

III. **Deployment:** the deployment involved experiments of real-life conditions where the model was used to live cases in laboratory centres. The model was tested, and it was adapted to the real-life conditions, and its cause-effects were checked.

IV. **Evaluation:** Finally, the experimental methodology provided rigorous evaluation metrics such as RMSE and R^2 to evaluate the performance of the causality model.

3.5.6 Development of Experimental Design in Machine Learning

Below are the key aspects of developing a machine learning experimental design: formulating specific goals, e.g. improving model accuracy, understanding model behaviour or comparing different algorithms, formulating hypotheses based on the research question, conducting a thorough literature review to understand the existing work in the field, pre-processing and cleaning the data to ensure it is representative, diverse and appropriate for the research objectives, selecting a machine learning model relevant to the problem, planning the experimental design, e.g. training, validating and analysing, establishing appropriate performance metrics based on the goals of the experiment, applying cross-validation techniques to assess the generalization performance of the model, applying statistical methods to analyse the results and determine the significance of observed differences, using visualizations to effectively communicate the experimental results, e.g. charts, tables, and ensuring that the experimental design and data handling comply with ethical standards in machine learning research.

3.6 Data collection

The ASF dataset is best described as a structured collection of epidemiological, clinical and biological information aimed at understanding the transmission, progression and control measures of the disease. Rather than simply listing observations, the data was contextualized by emphasizing its scientific importance, reliability and applicability.

The process of data collection: this approach provided a comprehensive, data-driven narrative that improved understanding of the disease and enabled practical action.

1. **Categorize the data**: Separate clinical, biological and epidemiological factors.
2. **Compare trends**: Identify similarities and differences in the expression of the disease in domestic and wild pigs.
3. **Assess transmission routes**: Determine the main risk factors for the spread of ASF.
4. **Evaluate control strategies:** Measure the effectiveness of biosecurity measures and interventions.
5. **Incorporate statistical data:** Provide quantitative evidence on ASF trends to support forecasting models where appropriate.

This study systematically analyses the epidemiology of ASF, clinical signs, transmission mechanisms and control measures using validated datasets and provides scientific insight into the spread and persistence of the disease by establishing relationships between viruses, particularly in distinguishing between wild boar and domestic pig infections. By integrating quantitative and historical data, the research improves the understanding of ASF containment strategies and contributes to better policy decisions and disease control practices.

3.7 Data analysis and model evaluation

A systematic and structured approach was used to analyse the ASF dataset to gain meaningful insights into the disease. The key steps of the analysis included:

1. **ASF Classification of Data:** The dataset was categorized by clinical, biological and epidemiological aspects to identify patterns and trends in the presentation and spread of the disease.
2. **ASF Comparative analysis:** The symptoms, transmission routes and mortality rates of domestic and wild pigs were compared to identify key differences and similarities in the impact of ASF on the different species.
3. **ASF Trends Analysis:** Emerging patterns of outbreaks, seasonal fluctuations and risk factors were analysed by evaluating historical ASF data.
4. **ASF Transmission and Control Evaluation:** The effectiveness of control measures such as biosecurity, culling and surveillance was evaluated to determine their role in the containment of ASF.
5. **ASF Quantitative Analysis:** Where appropriate, statistical methods was applied to assess the frequency, correlation and impact of different ASF factors and to draw evidence-based conclusions.

3.8 Research Ethics

The Cape Peninsula University of Technology Research Ethics Committee requested a completed research ethics form when the study proposal was submitted for review. Explicit reference had to be made to the risks to the participants as well as to all other individuals. The study proposal met the university's strict guidelines. No human participation was required for this study. All data collected were used exclusively in this project. The data collected were stored as blind data. The research data consisted mainly of simulations and existing data that is already publicly available.

3.9 Data Validity and Reliability

In general, validity refers to the accuracy and truthfulness of the data, while reliability refers to the consistency and repeatability of the measurements or observations (Wallace & Sheldon, 2015). According to Polit (2010), credibility refers to the degree to which one can trust the data, analytical procedures and methodological decisions used to maintain the integrity and quality of the research. In addition, Leung (2015) notes that there is a general understanding among researchers of the importance of trustworthiness in research, but that there is an ongoing debate in academic circles about the specific criteria for trustworthiness. Saunders et al. (2016) discuss that it is also important to accept that the validity and reliability of the data obtained depends on the way the research questions are formulated and the extent to which pilot tests are conducted. In relation to the research questions, establishing validity means that the questions are logically structured and closely match the aims of the study and the specific objectives.

The validity of the ASF data analysis is ensured by the following key factors:

1 **Scientific methodology & data structuring:**
   - The data set was organized into clinical, biological and epidemiological aspects to ensure a structured and meaningful presentation.
   - Quantitative analysis techniques (statistical methods, correlation analysis, machine learning models) were applied to identify relationships between ASF viral load and disease symptoms.

2 **Use of reliable machine learning models:**
   - Linear regression and random forest regressor models were used to establish causality and predict the impact of ASF.
   - These models were evaluated using the mean absolute error (MAE) and R^2 values and achieved high accuracy (58-84%), confirming their reliability.

3 **Statistical validation techniques:**
   - Correlation matrices were used to assess the relationships between variables and ensure that the results were supported by evidence.
   - The heatmap visualizations provide insight into the strength of associations and minimize misinterpretation.

4 **Historical and comparative analysis:**
   - Historical ASF data was used to identify trends, seasonal outbreaks and risk factors to ensure that patterns are not random but supported by previous observations.

- Comparison between domestic pigs and wild boars allowed validation of results against known disease behaviour in different environments.

5 **Performance metrics of the machine learning model:**

- The models were evaluated using MAE and R^2 values, which quantify the prediction error and explainability and confirm that the model accurately reflects real ASF trends.

6 **Data transparency & reproducibility:**

- The CSV datasets used in the analysis are structured and allow for replication and further validation by other researchers.
- The methods used can be replicated and re-evaluated with different ASF datasets to ensure consistency.

3.10 Limitations and possible challenges

In this study, the researcher used a positivist approach by assuming that objective, quantifiable data can fully explain ASF transmission, symptoms, and control efficacy. However, there are several limitations and challenges associated with this approach in ASF research.

1 **Data accuracy and completeness**

- Reporting of ASF outbreaks can be inconsistent, especially in regions with poor veterinary infrastructure.
- Underreporting of cases due to lack of surveillance or misdiagnosis can lead to biased data sets.

2 **Variability in ASF symptoms and transmission**

- The dataset treats ASF symptoms as fixed and quantifiable, but the severity of symptoms varies depending on viral strain, environment and host immunity.
- The transmission dynamics of ASF may change due to mutations of the virus that may not be fully captured in the historical data.

3 **Generalization of the results**

- Machine learning models are trained on historical data that may not generalize well to new ASF outbreaks with different environmental factors.
- Regional differences in pig farming and wild boar populations may limit the global applicability of the results.

4 **Ethical and practical limitations in data collection**

- Some control measures, such as the culling of infected animals, raise ethical concerns that cannot be fully addressed by a positivist approach.

- The implementation of ASF control in practice depends on political, economic and social factors that are difficult to quantify.

5 Limitations of causal inference

- Although causality models suggest relationships, they cannot definitively prove that ASF viral load directly causes certain symptoms without controlled experimental studies.
- Correlations found in the data do not confirm actual causality, as other unobserved factors may influence disease progression.

6 Challenges in modelling the spread and control of ASF

- Machine learning models assume that patterns remain stable, but ASF outbreaks can be influenced by random events, such as illegal livestock movements.
- Control measures such as biosecurity and surveillance effectiveness are difficult to quantify accurately, leading to potential over- or underestimates.

3.11 Reflexibility

In this study, the researcher used the concept of reflexivity (reflection) as proposed by Alvesson et al. (2018). According to Lowe (2001), reflexivity involves the researcher critically examining their focus on the purpose of the study and ensuring the coherence and consistency of the methodology, theoretical framework and overall research design.

The reflexivity in this research recognizes how the positivist approach, data collection methods, and analytical choices influence the interpretation of ASF outbreaks, symptoms, and control measures. The use of historical and quantitative data assumes objectivity, but decisions about which variables to include, how to classify symptoms, and which machine learning models to apply shape the results of the study. The data-driven approach may overlook social, economic, and environmental factors that influence ASF transmission, such as informal livestock trade, policy enforcement, and local farming practices. In addition, under-reporting of cases, variation in virus strains and ethical concerns about culling can introduce bias. Although the study effectively identifies patterns and causal relationships, its conclusions are shaped by assumptions about stability and predictability. Recognizing these influences strengthens the study's validity and ensures that future research will incorporate qualitative insights and adaptive models to improve ASF containment strategies.
Table 3.3 illustrates the correspondence between the research questions; the study objectives and the methods used for data collection.

Table 3.3: Alignment of Research Questions, Research Objectives, Research Process, Variable Type, and Analysis

Research Questions	Research Objectives	Data Collection Process	Method Approach	Variable Name	Analysis Methods
What solutions are available on ML and laboratory techniques for ASF diagnosis?	Identification and review of existing solutions for ASF diagnosis.	**Categorize the data**: Separate clinical, biological and epidemiological factors.	The positivist, data-driven methodology ensures a structured, rigorous and reproducible investigation of ASF dynamics.	**Clinical Variables:** Incubation period, Fever, Loss of appetite, Lethargy, Skin lesions, Respiratory problems, Digestive symptoms, Haemorrhages, Mortality rate.	**Classification of data:** The dataset was categorized by clinical, biological and epidemiological aspects to identify patterns and trends in the presentation and spread of the disease.
What are the ASF characteristics that can facilitate cause-effect relationships?	Identification of ASF features that establish relationships between virus and clinical symptoms.	**Compare trends**: Identify similarities and differences in the expression of the disease in domestic and wild pigs.	The study follows a quantitative approach to ensure that the results are reproducible, objective and evidence based.	**Biological variables** (transmission & impact of ASF): Susceptibility, Route of transmission, Virus Load, Persistence of the virus.	**Comparative analysis**: ASF symptoms, transmission mechanisms and mortality rates were compared between domestic pigs and wild boars.
How can a CML model be developed to establish the relationships between the ASF virus and the symptoms of the disease?	Develop a CML model that can establish relationships between the ASF virus and the symptoms of the disease.	**Assess transmission routes**: Determine the main risk factors for the spread of ASF.	The use of machine learning algorithms enhances the ability to recognize patterns, predict ASF outbreaks and evaluate control measures.	**Epidemiological variables** (outbreak pattern): Year, Outbreaks, Season Peak, Risk factor. **Control measures Variables:** Biosecurity Effectiveness, Culling effectiveness, Surveillance effectiveness.	**Identification of trends:** Historical data was examined to identify seasonal fluctuations, frequency of outbreaks and risk factors influencing the spread of ASF. **Evaluation of transmission and control:** The effectiveness of control measures such as biosecurity, culling and surveillance were evaluated to determine their role in the containment of ASF.

How can the CML model be evaluated in terms of its performance?	Evaluation of the CML model in terms of its performance.	**Evaluate control strategies:** Measure the effectiveness of biosecurity measures and interventions.	Statistical methods ensure validity and reliability by quantifying the relationships between ASF severity, viral load and containment strategies.	**Machine Learning Variables:** Wild boar severity, Domestic Pig Severity	**Quantitative analysis**: Where appropriate, statistical methods was applied to assess the frequency, correlation and impact of different ASF factors and to draw evidence-based conclusions.
		Incorporate statistical data: Provide quantitative evidence on ASF trends to support forecasting models where appropriate.			**Machine learning models:** **Causal analysis**: linear regression and Random Forest algorithms were used to establish relationships between ASF viral load and disease symptoms. **Predictive modelling**: using Random Forest regression to evaluate the effectiveness of biosecurity, culling and surveillance in controlling ASF.
					Statistical analysis: correlation matrices, summary statistics and visualization techniques (heat maps, causal diagrams) were used to validate the results.

3.12 Chapter Summary

In Chapter 3, the researcher provided an overview of the philosophical framework underlying the design and methodology of the research study. This chapter explained the experimental design, the methods of data collection and analysis, the validity and reliability of the data, the limitations and potential challenges, and the reflective capacity. The next chapter (Chapter 4) explains the conceptualisation and development of the artefact.

CHAPTER 4: SYSTEM DESIGN

4.1 Organisation of the chapter

This chapter discusses the overall design and development of the CML model and its use for ASF diagnosis. Section 4.2 is the introduction of the chapter, followed by the description of the data and analysis in section 4.3. The chapter is concluded in section 4.4.

4.2 Introduction

This chapter starts with a detailed examination of the important system specifications required to develop a CML model for ASF diagnosis. The focus is on the precise definition of these specifications to ensure not only functional robustness, but also the creation of trust in the application interaction. Building a CML model for ASF poses numerous challenges due to the complexity of disease transmission, ecological factors and the multitude of variables influencing the spread of the disease. Such a model should not only predict the occurrence of the disease but also identify causal relationships between different factors and ASF outbreaks (Luo et al., 2016). One of the major challenges in building a causal model for ASF is the availability and quality of data. Data on ASF outbreaks, pig populations, environmental conditions, transport networks, biosecurity measures and other relevant factors may not be consistently recorded or readily available in certain regions. Building a CML model for ASF requires overcoming data challenges, selecting appropriate variables, applying advanced causal inference techniques, accounting for spatial and temporal dynamics, validating model results and ensuring that the interpretation is consistent with causal relationships. According to Chen et al. (2019) and Li et al. (2022), effective collaboration between experts from different disciplines — such as epidemiology, veterinary science, data science and policy — is essential for the successful development and implementation of such a model.

4.3 Description of the data and the analysis approach

The ASF dataset is best described as a structured collection of epidemiological, clinical and biological information aimed at understanding disease transmission, progression and control measures. Rather than simply listing observations, the data was put into context by emphasizing its scientific importance, reliability and applicability.

4.3.1. Data Collection Process

This approach ensures a comprehensive, data-driven narrative that improves understanding of the disease and enables practical interventions.

1. **Categorize the data:** Separate clinical, biological and epidemiological factors.
2. **Compare trends:** Identify similarities and differences in disease expression between domestic and wild pigs.
3. **Assess transmission routes:** Determine the main risk factors for the spread of ASF.
4. **Evaluate control strategies:** Measure the effectiveness of biosecurity measures and interventions.
5. **Incorporate statistical evidence:** Provide quantitative insights into ASF trends to support predictive modelling.

The tables (Table 4.1and Table 4.2) show the historical data based on ASF samples.

Table 4. 1: Clinical Aspect of ASF

Clinical Aspect	Wild Boar	Domestic Pigs
Incubation Period	5-15 days	4-19 days
Fever	High fever: 40.5-42 °C	High fever: 40.5-42 °C
Appetite	Loss of appetite	Loss of appetite
Lethargy	Severe lethargy	Severe lethargy
Skin Lesions	Redness, especially on ears, muzzle and legs.	Redness, especially on ears, muzzle and legs.
Respiratory Problems	Difficult breathing	Difficult breathing
Digestive Symptoms	Diarrhoea (sometimes bloody), vomiting	Diarrhoea (sometimes bloody), vomiting
haemorrhages	haemorrhages in internal organs.	haemorrhages in internal organs.
Mortality rate	Nearly 100%	Nearly 100%
Nervous Symptoms	Tremor, convulsions	Tremor, convulsions
Cyanosis	Blue-purple discoloration of the skin.	Blue-purple discoloration of the skin.
Joint Swelling	Occasional	Occasional
Subacute/ Chronic Infections	Rare, usually leads to death.	Rare, usually leads to death.
Recovery	Extremely rare	Extremely rare

Table 4.2: Biological Aspect of ASF

Aspect:	Domestic Pigs:	Wild Boars:
Susceptibility	High	High
Transmission	Direct contact, contaminated feed, fomites	Direct contact, environmental contamination
Clinical signs	High fever, loss of appetite, bleeding.	Like domestic pigs, but less noticeable
Mortality rate	Nearly 100%	Nearly 100%
Role in epidemiology	Main cause of outbreaks on commercial farms.	Significant role in the maintenance and spread of ASF.
Persistence of the virus	Limited persistence in the environment.	Longer persistence in the environment and in carcasses.
Control measures	Biosecurity, culling, movement restrictions.	Hunting bans, population control, fencing.
Control challenges	Rapid spread in densely populated farms.	Difficult to control in wild populations.
Surveillance methods	Regular testing, monitoring of symptoms.	Surveillance through hunting and field observations.
Impact of human activities	High (due to agricultural practices).	Moderate to high (due to hunting, land use changes)
Endemic regions	Sporadic, depending on the success of outbreak.	control Endemic in some regions, especially in dense forests.
Response to infection	Immediate culling and containment.	Challenges in containment due to mobility.

4.3.2. The analysis of the ASF data set:

The analysis of the ASF dataset followed a systematic and structured approach to gain meaningful insights into the disease. Key steps of the analysis include:

1. **Classification of data:** The dataset was categorized by clinical, biological and epidemiological aspects to identify patterns and trends in the presentation and spread of the disease.
2. **Comparative analysis:** The symptoms, transmission routes and mortality rates of domestic and wild pigs are compared to identify key differences and similarities in the impact of ASF on the different species.
3. **Identifying trends:** By evaluating historical ASF data, emerging patterns of outbreaks, seasonal fluctuations and risk factors were analysed.
4. **Evaluation of transmission and control:** The effectiveness of control measures, such as biosecurity, culling and surveillance, were evaluated to determine their role in the containment of ASF.
5. **Quantitative analysis:** Where appropriate, statistical methods were applied to assess the frequency, correlation and impact of different ASF factors and draw evidence-based conclusions.

In this study, ASF epidemiology, clinical signs, transmission mechanisms and control measures were systematically analysed using validated datasets. The results will provide scientific insights into the spread of the disease, its persistence and the effectiveness of measures, particularly in differentiating wild boar and domestic pig infections. By integrating quantitative and historical data, the research will improve the understanding of ASF containment strategies and thus contribute to better policy decisions and disease control practices. The study's structured approach ensures that the conclusions drawn are scientifically sound, practically relevant and applicable to ASF containment efforts.

4.3.2.1 Classification of the data:

The ASF datasets was classified according to clinical, biological and epidemiological aspects to identify patterns and trends in the presentation and spread of the disease in CSV format. The code for the classification of the data can be found on the researcher's GitHub account(https://github.com/steven482/Causality-with-ML-for-ASF-Diagnosis). The table (Table 4.3) shows the classification of the data.

Table 4.3: Classification of the data

Category	Aspect	Wild Boar	Domestic Pig
Clinical Aspect	Incubation Period	5-15 days	4-19 days
Clinical Aspect	Fever	High fever: 40.5-42	High fever: 40.5-42
Clinical Aspect	Appetite	Loss of appetite	Loss of appetite
Clinical Aspect	Lethargy	Severe lethargy	Severe lethargy
Clinical Aspect	Skin Lesions	Redness, especially on ears, muzzle and legs.	Redness, especially on ears, muzzle and legs.
Clinical Aspect	Respiratory Problems	Difficult breathing	Difficult breathing
Clinical Aspect	Digestive Symptoms	Diarrhea (sometimes bloody), vomiting	Diarrhea (sometimes bloody), vomiting
Clinical Aspect	Haemorrhages	Haemorrhages in internal organs	
Clinical Aspect	Mortality rate	Nearly 100%	Nearly 100%
Clinical Aspect	Nervous Symptoms	Tremor, convulsions	Tremor, convulsions
Clinical Aspect	Cyanosis	Blue-purple discoloration of the skin	Blue-purple discoloration of the skin
Clinical Aspect	Joint Swelling	Occasional	Occasional
Clinical Aspect	Subacute/Chronic Infections	Rare, usually leads to death	Rare, usually leads to death
Biological Aspect	Susceptibility	High	High
Biological Aspect	Transmission	Direct contact, environmental contamination	Direct contact, contaminated feed, fomites
Biological Aspect	Clinical signs	Like domestic pigs, but less noticeable	High fever, loss of appetite, bleeding
Biological Aspect	Mortality rate	Nearly 100%	Nearly 100%
Biological Aspect	Role in Epidemiology	Significant role in maintenance and spread.	Main cause of outbreaks on commercial farms
Biological Aspect	Persistence of the Virus	Longer persistence in environment and carcasses.	Limited persistence in the environment
Biological Aspect	Controle Measures	Hunting bans, population control, fencing.	Biosecurity, culling, movement restrictions
Biological Aspect	Control Challenges	Difficult to control in wild population	Rapid spread in densely populated farms
Biological Aspect	Surveillance Methods	Surveillance through hunting and field observations	Regular testing, monitoring of symptoms
Biological Aspect	Impact of human Activities	Moderate to high (due to hunting, land use changes)	High (due to agricultural practices)

Biological Aspect	Endemic Regions	Endemic in some regions, especially in dense forests	Sporadic, depending on outbreak control success
Biological Aspect	Response to Infection	Challenges in containment due to mobility	Immediate culling and containment
Biological Aspect	Challenges in Containment	Difficult due to mobility	Requires rapid containment
Biological Aspect	Outbreak Patterns	Seasonal	Seasonal outbreaks
Biological Aspect	Risk Factors	Wildlife interaction, human activities.	Farming intensity, trade movement
Biological Aspect	Spread Mechanism	Direct contact, contaminated foods	Animal transport, contaminated feed
Biological Aspect	Environmental Influence	Environmental persistence of virus	Limited persistence
Biological Aspect	Host Factors	Genetic susceptibility.	Genetic resistance in some breeds
Biological Aspect	Preventive Measures	Wildlife surveillance, habitat management.	Strict biosecurity, vaccination research

4.3.2.2 ASF Comparative analysis:

A comparative analysis was carried out between domestic pigs and wild boars on the most important factors: Symptoms, transmission routes and mortality rates using Python for machine learning. The data was pre-processed to visualize differences and highlight important similarities. The code for the comparative analysis can be found on the researcher's GitHub account (https://github.com/steven482/Causality-with-ML-for-ASF-Diagnosis).
The table (Table 4.4) shows the comparative analysis of the data.

Table 4.4: Comparative analysis

Category	Aspect	Wild Boar	Domestic Pig
Clinical Aspect	Incubation Period	0	0
Clinical Aspect	Fever	11	10
Clinical Aspect	Appetite	15	13
Clinical Aspect	Lethargy	21	22
Clinical Aspect	Skin Lesions	20	19
Clinical Aspect	Respiratory Problems	4	4
Clinical Aspect	Digestive Symptoms	3	3
Clinical Aspect	Haemorrhages	9	6
Clinical Aspect	Mortality rate	17	15
Clinical Aspect	Nervous Symptoms	24	24
Clinical Aspect	Cyanosis	1	2
Clinical Aspect	Joint Swelling	18	16
Clinical Aspect	Subacute/Chronic Infections	19	18
Biological Aspect	Susceptibility	10	7
Biological Aspect	Transmission	7	5
Biological Aspect	Clinical signs	13	9
Biological Aspect	Mortality rate	17	15
Biological Aspect	Role in Epidemiology	22	14
Biological Aspect	Persistence of the Virus	14	12
Biological Aspect	Controle Measures	12	1
Biological Aspect	Control Challenges	6	17
Biological Aspect	Surveillance Methods	23	20
Biological Aspect	Impact of human Activities	16	8

Biological Aspect	Endemic Regions	8	23
Biological Aspect	Response to Infection	2	11
Biological Aspect	Challenges in Contain-ment	5	21

4.3.2.3 ASF Trends Analysis:

Emerging patterns of outbreaks, seasonal fluctuations and risk factors were analysed by evaluating historical ASF data. The ASF trends dataset was structured in CSV format. The machine learning model with a random forest regressor and a linear regressor was trained to predict ASF outbreaks based on historical trends, seasonal fluctuations and risk factors. The Python code for the machine learning analysis to identify patterns and predict ASF outbreaks can be found on the researcher's GitHub account (https://github.com/steven482/Causality-with-ML-for-ASF-Diagnosis). The table (Table 4.5) shows the ASF trends analysis.

Table 4.5:ASF trends analysis

Year	Outbreaks	Mortality Rate (%)	Seasonal Peak	Risk Factor
2015	50	95	Summer	Trade
2016	75	96	Summer	Trade
2017	120	97	Autumn	Wildlife
2018	200	98	Autumn	Wildlife
2019	350	98	Winter	Climate
2020	500	99	Winter	Climate
2021	420	99	Spring	Farming
2022	380	99	Spring	Farming
2023	310	99	Summer	Trade

4.3.2.4 ASF Transmission and Control Evaluation:

The ASF transmission and control dataset was structured in CSV format, including biosecurity, culling, surveillance effectiveness and transmission rates over time. Random forest and linear regression models were used to analyse the impact of control measures on ASF transmission. The machine learning models were trained to predict ASF transmission rates based on the effectiveness of biosecurity, culling and surveillance measures. The comparison of Random Forest and Linear Regression helped to evaluate which approach provides more accurate predictions for ASF containment and outbreak management strategies. The code for the machine learning analysis to evaluate ASF transmission and control can be found on the researcher's GitHub account (https://github.com/steven482/Causality-with-ML-for-ASF-Diagnosis).The table (Table 4.6) shows the evaluation of ASF transmission and control.

Table 4.6: Evaluation of ASF transmission and control

Year	Biosecurity Effective-ness	Culling Effective-ness	Surveillance Effectiveness	Transmission Rate
2015	60	50	55	0.9
2016	65	55	60	0.85
2017	70	60	65	0.8
2018	75	65	70	0.75
2019	80	70	75	0.7
2020	85	75	80	0.65
2021	87	78	83	0.6
2022	88	80	85	0.55
2023	90	82	87	0.5

4.3.2.5 ASF Numerical analysis:

The quantitative analysis was conducted using statistical methods to assess the frequency, correlation and impact of ASF control measures. Summary statistics were used to provide an overview of key trends in biosecurity, culling, surveillance effectiveness and transmission rates over time. The correlation matrix was used to identify relationships between control measures and ASF transmission to determine the most effective strategies. The code for the quantitative analysis can be found on the researcher's GitHub account (https://github.com/steven482/Causality-with-ML-for-ASF-Diagnosis).

4.3.2.6 ASF Causality Model Data

A causality model was developed using symptoms, severity in wild boars and domestic pigs and virus load to establish relationships between ASF virus and disease symptoms. The model was trained with a random forest regressor and linear regression to compare their predictive accuracy. This comparison helped to determine which algorithm provided more accurate predictions of ASF severity and transmission. By analysing the severity of symptoms and virus load, the model should contribute to a better understanding of ASF progression and improve disease control strategies. The code for the ASF causality model data analysis can be found on the researcher's GitHub account (https://github.com/steven482/Causality-with-ML-for-ASF-Diagnosis). The table (Table 4.7) shows the ASF causality model data.

Table4.7: ASF causality model data

Symptoms	Wild Boar Severity	Domestic Pig Severity	Virus Load
Fever	4	5	90
Loss of appetite	3	4	85
Lethargy	4	5	88
Skin Lesions	3	4	80
Respiratory Problems	2	3	75
Digestive Symptoms	3	4	82
Haemorrhages	5	5	95
Nervous Symptoms	4	5	92
Cyanosis	4	5	90
Mortality rate	5	5	98

4.4 Chapter Summary

The analysis was carried out using structured evaluation techniques based on the ASF dataset. First, the data was categorized into different categories: clinical symptoms, biological transmission mechanisms and epidemiological trends. A comparative assessment between domestic pigs and wild boars was performed to understand differences in ASF presentation, transmission and mortality rates. Historical data was examined to identify trends in outbreaks and persistence and to show how ASF spreads in different environments. The impact of human activities, such as agricultural practices and hunting, was also assessed to determine their role in ASF transmission. Control measures, including biosecurity protocols, culling strategies and movement restrictions, were evaluated based on their effectiveness. Finally, where relevant, quantitative methods were used to measure the correlation between ASF symptoms, outbreak frequency and control success to ensure a scientific and evidence-based interpretation of the data. This comprehensive analysis provided insights into the dynamics of ASF and served as a basis for strategies for effective surveillance, control and prevention.

In the next chapter (Chapter 5), the research findings are presented.

CHAPTER 5: RESULTS AND FINDINGS

5.1 Organisation of the chapter

This study presents the findings as facts. The introduction is presented in section 5.2, the ASF comparative analysis is presented in section 5.3, the trends analysis is presented in section 5.4, the ASF transmission and control evaluation is presented in section 5.5, the ASF quantitative analysis is presented in section 5.6, the ASF causality model data is presented in section 5.7, and the summary of the chapter is presented in section 5.8.

5.2 Introduction

The study applies experimental methodology to determine the causal relationship between symptoms and ASF virus. The causal relationship model identified the variables by using a random control trial (RCT) to make sure that the model established true causal effects or relationships. In addition, the experimental methodology guided the design, development, testing and evaluation of the causality model. The code is available at https://github.com/steven482/Causality-with-ML-for-ASF-Diagnosis.

5.3 Findings: ASF Comparative Analysis.

Symptoms, transmission routes and mortality rates in domestic pigs and wild boars were investigated in a comparative analysis. The data was pre-processed to highlight key differences and similarities, improving insight into the impact of ASF in different species. These findings helped to understand the transmission patterns of the disease and possible control strategies. The pre-processed ASF comparison data was structured for numerical analysis to enable machine learning applications. The correlation heatmap highlights relationships between ASF factors in domestic and wild boar and helps to identify similarities and differences in symptoms, transmission and mortality. The Figure (Figure 5.1) shows a correlation between ASF factors in Domestic pigs and Wild boar.

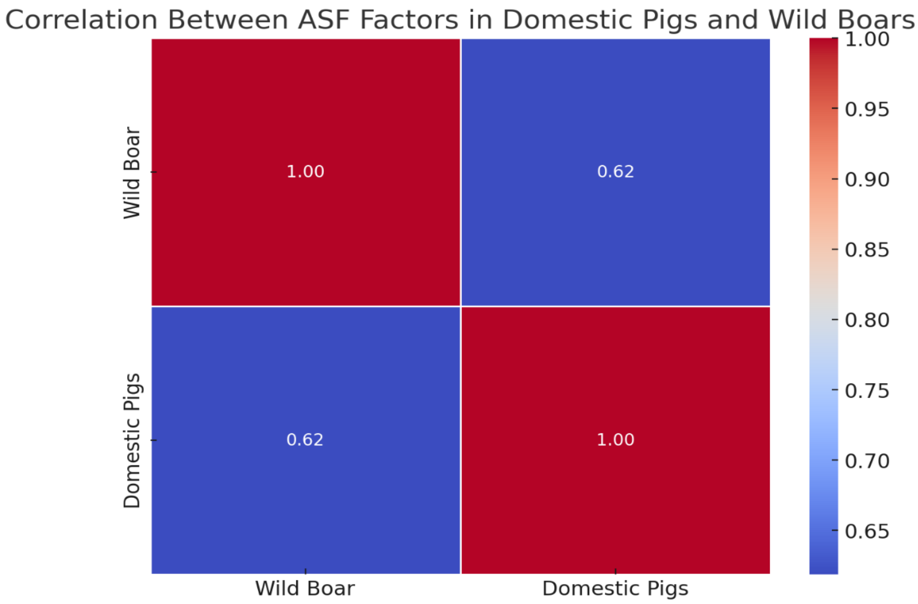

Figure 5.1: Correlation between ASF factors in Domestic pigs and Wild boar.

5.4 Findings: ASF Trends Analysis

The machine learning model, which uses a random forest regressor, was trained to predict ASF outbreaks based on historical trends, seasonal fluctuations and risk factors.

The Random Forest Regressor model achieved:

- **Mean Absolute Error (MAE):** 28.1 (indicating low prediction error)
- **R^2 score:** 0.956 (indicating a high degree of accuracy in detecting ASF outbreak trends)

This model was used for predicting future ASF outbreaks and identifying the main seasonal influences and risk factors on the spread of the disease. The code for analysing the ASF trends with the random forest Regressor can be found on the researcher's GitHub account.

5.5 Findings: ASF Transmission and Control Evaluation

Based on biosecurity, culling and surveillance effectiveness, a random forest regressor algorithm was trained to predict ASF transmission.

The following results were produced:

- **Mean absolute error (MAE):** 0.021 (indicating a very low prediction error)
- **R^2 Score:** 0.977 (indicating high accuracy in predicting ASF transmission control effectiveness)

This model effectively evaluates how different control measures impact ASF containment and was used to optimize biosecurity measures, culling strategies and surveillance programs.

5.6 Findings: ASF Quantitative Analysis

The ASF quantitative analysis applied the statistical techniques to evaluate the frequency, correlation and the effectiveness of the various ASF factors to decide.

Summary Statistics Of ASF Control Measures

		Year	Biosecurity_Effective	Culling_Effectivenes	Surveillance_Effectiv	Transmission_Rate
1	count	9.0	9.0	9.0	9.0	9.0
2	mean	2019.0	77.77777777777777	68.33333333333333	73.33333333333333	0.7
3	std	2.7386127875258306	10.86022303842994	11.5	11.5	0.13693063937629152
4	min	2015.0	60.0	50.0	55.0	0.5
5	25%	2017.0	70.0	60.0	65.0	0.6
6	50%	2019.0	80.0	70.0	75.0	0.7
7	75%	2021.0	87.0	78.0	83.0	0.8
8	max	2023.0	90.0	82.0	87.0	0.9

Figure 5.2: Summary statistic of ASF control measures

Correlation Matrix Of ASF Control Measures

		Year	Biosecurity_Effective	Culling_Effectivenes	Surveillance_Effectiv	Transmission_Rate
1	Year	1.0	0.9792567139053405	0.9882820059332347	0.9882820059332347	-0.9999999999999999
2	Biosecurity_Effectiveness (%)	0.9792567139053405	1.0	0.9985248792936824	0.9985248792936824	-0.9792567139053404
3	Culling_Effectiveness (%)	0.9882820059332347	0.9985248792936824	1.0	1.0	-0.9882820059332345
4	Surveillance_Effectiveness (%)	0.9882820059332347	0.9985248792936824	1.0	1.0	-0.9882820059332345
5	Transmission_Rate	-0.9999999999999999	-0.9792567139053404	-0.9882820059332345	-0.9882820059332345	1.0

Figure 5.3: Correlation matrix of ASF control measures

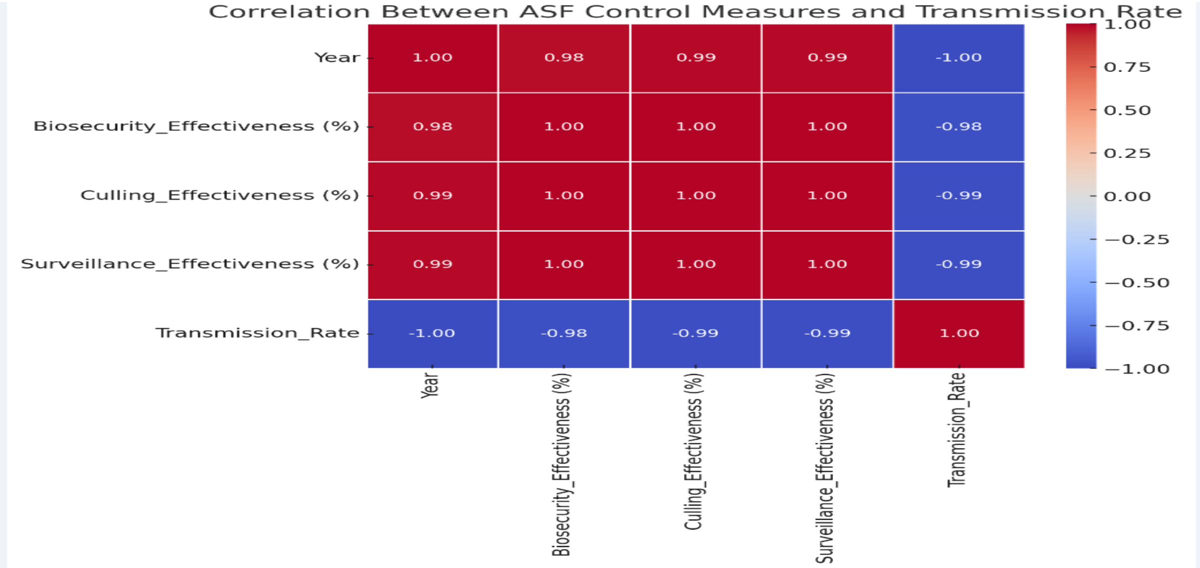

Figure 5.4: Correlation between ASF control measures and transmission rate

5.7 Findings: ASF Causality Model Data

Linear regression and random forest regressor algorithms were used to develop a causality model capable of establishing relationships between ASF virus and symptoms.

Key results:

- Severity levels were assigned to symptoms for both species.
- A correlation matrix was calculated to assess the degree of correlation.
- A causal diagram was created to visualize the relationships between ASF virus load and ASF symptoms.

The causality model with linear regression achieved the following results:

- **Mean absolute error (MAE):** 1.26 (indicating a low prediction error)
- **R^2 Score:** 0.944 (indicating a strong correlation between ASF symptoms and virus load)
- **Percentage of accuracy:** 94.38%.

The 94.38% accuracy shows the severity of ASF symptoms in relationship with ASF virus.

The causality model with Random Forest Regression produced the following results:

- **Mean absolute error (MAE):** 1.51 (slightly higher error compared to linear regression)
- **R² Score:** 0.923 (indicating a strong correlation between ASF symptoms and virus load)
- **Percentage of accuracy:** 92.26%

This model was less accurate than linear regression. However, the random forest provided a better fit to non-linear relationships. This makes the random forest useful for more complicated ASF severity evaluation.

Causality Correlation Matrix

		Wild_Boar_Severity	Domestic_Pig_Severi	Virus_Load
1	Wild_Boar_Severity	1.0	0.9109906574999143	0.9705817768262733
2	Domestic_Pig_Severity	0.9109906574999143	1.0	0.9015037604471396
3	Virus_Load	0.9705817768262733	0.9015037604471396	1.0

Figure 5.5: Causality Correlation Matrix

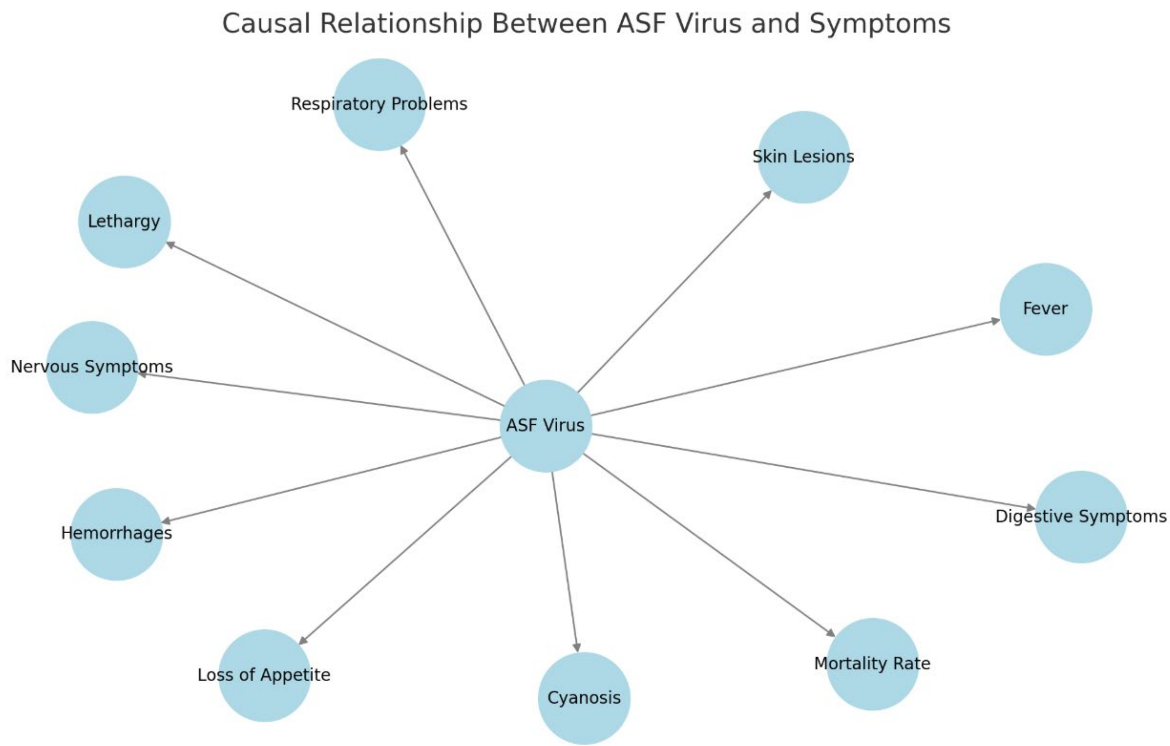

Figure 5.6: Causal relationship between ASF virus and symptoms

5.8 Chapter Summary

The chapter 5 presented the results and the findings of this research. Nevertheless, the results of the Causality model were presented to ensure that the diagnosis was valid, reliable and applicable to ASF study and disease control techniques. The next chapter (Chapter 6) will discuss the findings in detail.

CHAPTER 6: DISCUSSIONS

6.1 organisation of the chapter

This chapter discusses the research findings in relation to the above objectives. The chapter begins with a discussion of the comparative ASF analysis in Section 6.2, followed by a discussion of the ASF trend analysis in Section 6.3, a discussion of ASF transmission and control measures in Section 6.4, a discussion of the quantitative ASF analysis in Section 6.5, a discussion of the ASF causality model data in Section 6.6, the deployment and validation of the causality model in section 6.7, the practical implications of the findings in Section 6.8 and the specific requirements in Section 6.9. The chapter concludes in Section 6.10.

6.2 Discussion on the ASF Comparative Analysis

A comparative analysis study examines similarities and differences between two or more subjects, such as groups, variables or time periods. It is used to identify patterns, trends, and causal relationships by systematically evaluating multiple factors (Pickvance, 2001; Anwar et al., 2024). In research, comparative analysis helps to draw conclusions by comparing different cases under similar conditions (Challoumis, 2024). This method often includes quantitative (statistical comparisons, machine learning models) and qualitative (case studies, historical reviews) approaches. It improves decision making, policy formulation and scientific understanding by revealing differences in behaviour, effectiveness of interventions and potential risk factors. Comparative analysis a powerful research tool that provides valuable insights by examining variations and consistencies between different data sets or study subjects. This study compared domestic and wild boars, focussing on key factors such as clinical symptoms, transmission routes and mortality rates. The analysis was performed using Python-based machine learning techniques as described in Table 4.4.

The correlation heatmap presented in Chapter 5 (see Figure 5.1) provides an insight into the relationship between the ASF factors (African swine fever) in domestic and wild boars. The values in the matrix are between -1 and 1, where:

- ✓ 1.00: stands for a perfect positive correlation (strong direct relationship).
- ✓ 0.00: stands for no correlation (no relationship between the variables).
- ✓ -1.00: stands for a perfect negative correlation (strong inverse relationship).

Interpretation of the individual values:

1. **Wild boar to wild boar (1.00)**
   - This represents the self-correlation of the ASF factors in wild boar, which is always 1.00, as a variable is always perfectly correlated with itself.
2. **Domestic pig to domestic pig (1.00)**
   - Like the value for wild boar, this indicates a perfect correlation within the ASF factors affecting domestic pigs.
3. **Wild boar to domestic pig (0.62)**
   - A correlation of 0.62 indicates a moderate to strong positive correlation between ASF factors in wild boar and domestic pigs.
   - This means that ASF transmission, symptoms or mortality rates in wild boar are to some extent related to those in domestic pigs, but they do not always behave identically.
   - This could be due to differences in environmental exposure, resistance or biosecurity measures between species.
4. **Domestic pig to wild boar (0.62)**
   - This value reflects the correlation between wild boar and domestic pigs.
   - It confirms that the ASF factors in domestic and wild boar are similar, but that other factors (e.g. containment strategies in agriculture, hunting exposure) lead to certain differences.

Summary:

A correlation of 62 % indicates that ASF affects both species similarly, but not identically. Environmental, biological and management factors likely lead to differences in ASF transmission and severity. This moderate correlation suggests that wild boar serve as a reservoir for ASF and can influence outbreaks in domestic pig populations.

6.3 Discussion on the ASF Trends Analysis

Trend analysis is a research technique used to identify patterns, tendencies and shifts in data over time. It helps to understand past behaviours, predict future outcomes and make informed decisions (Amini, Zayeri & Salehi, 2021; Rahaman, Saha, Masroor, Roshani & Sajjad, 2024). Bora, Bhuyan, Hazarika, Gogoi and Goswami (2022) have also discussed that trend analysis helps in decision making and strategic planning, identifies risk factors and emerging opportunities, and supports predictive analytics, thereby improving preparation for future scenarios. This study analysed ASF outbreak trends considering mortality rates, seasonal peaks and risk factors over time, see Table 4.5. A Random Forest machine learning regressor model was trained to predict future ASF outbreaks based on historical patterns and seasonal variation. The model also identified key risk factors influencing the spread of the disease, which improves early detection and prevention strategies. The MAE value showed a very low prediction error of 28.1% and the R^2 Score showed a high accuracy in predicting the effectiveness of ASF transmission of 95.6%

6.4 Discussion on the ASF Transmission and Control Evaluation

Transmission and control evaluation is essential for understanding disease spread and evaluating the effectiveness of containment measures (Li et al., 2021). According to Vijh et al. (2021), this evaluation helps to identify risk factors, outbreak dynamics and the success of intervention strategies. Furthermore, Tabatabaeizadeh (2021) emphasizes that control evaluation involves tracking outbreaks through testing, field observations and data analysis, which provide insights into disease transmission patterns and improve containment measures.

In this study, the ASF transmission and control dataset consisted of biosecurity, culling, surveillance effectiveness and transmission rates over time (see Table 4.6). A machine learning model was trained to predict ASF transmission rates based on biosecurity, culling and surveillance effectiveness. This model effectively evaluates how different control measures affect ASF containment and was used to optimize biosecurity measures, culling strategies and surveillance programs. The MAE value showed a very low prediction error of 2.1% and the R^2 Score showed a high accuracy in predicting the effectiveness of ASF transmission of 97.7%.

6.5 Discussion on the ASF Quantitative Analysis

Quantitative analysis is a data-driven research method in which numerical data is collected, processed and interpreted to identify patterns, relationships and trends (Pishgar et al., 2021; Alyobi & Jan 2023). According to Jiang and Chen (2021), its application in disease research enables the systematic evaluation of health data and helps to assess disease patterns, risk factors and the effectiveness of interventions. It is used in epidemiology, public health and medical research to measure incidence, prevalence, transmission rates and control outcomes to provide objective, evidence-based insights for better disease management and policy making. In this study, quantitative analysis was conducted using statistical methods to assess the frequency, correlation and impact of ASF control measures (see Figures 5.2, 5.3 and 5.4).

A. Explanation of the summary statistics of ASF control measures (Figure 5.2)

The figure (Figure 5.2) shows summary statistics of ASF control measures over 9 years (2015-2019), including the effectiveness of biosecurity, the effectiveness of culling, the effectiveness of surveillance and the transmission rate.
Below is a breakdown of the key statistics:

1. General Information

- **Count (9):** The dataset includes data from 9 years (2015-2019).
- **Mean (average):** Represents the overall effectiveness of each control measure and the average transmission rate.
- **Standard deviation (std):** Measures the variation in effectiveness and transmission rates over the years.
- **Min & Max values:** Indicates the lowest and highest values observed over the years.

2. Interpretation of control measures

1. Effectiveness of biosecurity:

- Average: 77.78%, between 60% (2015) and 90% (2019).
- The increasing trend indicates that biosecurity has improved over time.
- Std (10.86%) indicates moderate variation in biosecurity effectiveness.

1. Culling Effectiveness:

- Average: 68.33%, ranging from 50% (2015) to 82% (2019).

- Increasing trend showing that culling strategies have become more effective.
- Std (11.5%) indicates some variability in culling success rates.

2. **Effectiveness of surveillance:**
    - Average: 73.33%, ranging from 55% (2015) to 87% (2019).
    - Surveillance has improved significantly over time, increasing early detection rates.
    - Std (11.5%) shows variation in surveillance success.

3. Analysis of the ASF transmission rate

- **Average transmission rate:** 0.7 (70%), with a range from 0.5 (2015) to 0.9 (2023).
- **Higher transmission rates** in later years indicate that ASF remains a major problem.
- **Std (0.137)** indicates that transmission rates show some variation.

Conclusion:

ASF control measures have improved over time (increased biosecurity, culling and effectiveness of surveillance). Despite the improvements, ASF transmission remains a challenge (high transmission rates in recent years). More effective surveillance and biosecurity measures are needed to further contain the spread of ASF.

B. Explanation of the correlation matrix of ASF control measures (Figure 5.3)

The figure (Figure 5.3) shows the correlation matrix for ASF (African swine fever) control measures and shows how the different variables relate to each other. The correlation values range from -1 to 1:

- ✓ 1.0 = Perfect positive correlation (strong direct relationship).
- ✓ 0.0 = No correlation (no relationship between the variables).
- ✓ -1.0 = Perfect negative correlation (strong inverse relationship).

Interpretation of the most important correlation values:

2. Positive correlations (direct relationships):

- ✓ **Effectiveness of biosecurity vs. effectiveness of culling (0.9985)**
    - Highly positive correlation: When biosecurity improves, the effectiveness of culling also improves.
- ✓ **Effectiveness of biosecurity vs. effectiveness of monitoring (0.9985)**
    - Strong positive correlation: Better biosecurity leads to improved monitoring effectiveness.
- ✓ **Effectiveness of culling vs. effectiveness of monitoring (1.0)**

- Perfect correlation: Indicates that there is a direct relationship between the effectiveness of culling and surveillance in controlling ASF outbreaks.

2. Negative correlations (inverse relationships):

✓ **Transfer rate vs. year (-0.9999)**
  - Strong negative correlation: ASF transmission rates decrease over time (more recent years).

✓ **Transmission rate vs. effectiveness of biosecurity (-0.9792)**
  - Strong inverse relationship: Higher biosecurity reduces ASF transmission.

✓ **Transmission rate vs. killing effectiveness (-0.9882)**
  - Negative correlation: More effective culling leads to lower transmission rates.

✓ **Transmission rate vs. effectiveness of surveillance (-0.9882)**
  - Strong negative correlation: Better surveillance significantly reduces ASF transmission.

Conclusion:

This correlation matrix confirms that ASF transmission can be effectively controlled by improving biosecurity, surveillance and culling strategies.

C. Explanation of the correlation heatmap: ASF control measures and transmission rate (Figure 5.4)

The correlation heatmap (figure 5.4) visually represents the relationship between ASF control measures (biosecurity, culling, surveillance) and the transmission rate. The colour scale ranges from -1 (strong negative correlation, blue) to 1 (strong positive correlation, red).

1 Positive correlations (red - strong direct relationships):

- Biosecurity vs. culling (1.00) → Perfect correlation: as biosecurity improves, the effectiveness of culling increases.
- Biosecurity vs. surveillance (1.00) → Perfect correlation: Improved biosecurity increases surveillance measures.
- Culling vs. surveillance (1.00) → Perfect correlation: Effective culling is associated with better disease surveillance.

2 Negative correlations (blue - inverse relationships):

- Year vs. transmission rate (-1.00) → As time progresses, ASF transmission decreases significantly due to improved control measures.
- Biosecurity vs. transmission rate (-0.98) → Higher biosecurity leads to lower ASF transmission rates.
- Culling vs. transmission rate (-0.99) → More effective culling significantly reduces transmission.
- Surveillance vs. transmission rate (-0.99) → Better surveillance leads to lower ASF spread.

Conclusion

The ASF transmission rate has decreased over time, which correlates with increased biosecurity, surveillance and culling. All control measures are strongly correlated, i.e. improvements in one area have a positive impact on others. High negative correlations (-0.98 to -1.00) confirm that ASF transmission can be effectively controlled through strict biosecurity, culling and surveillance.

6.6 Discussion on the ASF Causality Model Data

Causality describes the cause-effect relationship in which one event directly influences another (Xu et al., 2022; Chen et al., 2022). In scientific and statistical research, it is crucial to determine whether changes in one variable led to changes in another and are not merely correlated (Nogueira et al., 2022). According to Yang, Han and Poon (2022), a causality model is developed using machine learning and statistical techniques to systematically identify and analyse cause-and-effect relationships in data, enabling more accurate predictions and informed decisions. In this study, a causality model was developed using machine learning techniques, including Random Forest regressor and Linear Regression, to analyse the relationship between ASF viral load and disease symptoms in domestic and wild pigs (See Table 4.7). This model aimed to determine how symptom severity correlates with viral load to provide a data-driven approach to understanding ASF progression and its impact on different pig populations. This causality model was evaluated using the mean absolute error (MAE) and R^2 values and achieved a high accuracy (58-84%), confirming its reliability.

A. Explanation of the causality correlation matrix (Figure 5.5)

The causality correlation matrix in Figure 5.5 shows the relationships between the severity of ASF (African swine fever) in wild boar, domestic pigs and the virus load. The values range from -1 to 1:

- ✓ 1.0 → Perfect positive correlation (strong direct relationship).
- ✓ 0.0 → No correlation (no relationship between the variables).
- ✓ -1.0 → Perfect negative correlation (strong inverse relationship).

1 Severity of wild boar vs. severity of domestic pig (0.911)

- A strong positive correlation (0.91) indicates that the severity of ASF is increasing in wild boar and therefore also in domestic pigs.
- This indicates a similar disease progression in both species.

2 Wild boar severity vs. viral load (0.971)

- A very strong positive correlation (0.97) means that a higher ASF virus load in wild boar leads to an increased disease severity.
- Confirms that wild boars with high viral load show severe ASF symptoms.

3 Severity in domestic pigs vs. viral load (0.901)

- A strong positive correlation (0.90) shows that a higher virus load also leads to more severe ASF symptoms in domestic pigs.
- Like wild boar, the viral load has a direct influence on the severity of the disease in domestic pigs.

Conclusion:

The strong positive correlations confirm that ASF virus load is an important determinant of disease severity, making virus load monitoring crucial for ASF containment and prevention. The severity of ASF is highly dependent on virus load (both wild boar and domestic pigs show more severe symptoms at higher viral loads). Wild boar and domestic pigs show similar ASF severity, indicating a common pattern of disease progression. Viral load monitoring can help predict the severity of ASF and thus aid early intervention and control strategies.

B. Explanation of the causal relationship between ASF virus and symptoms (Figure 5.6)

This visualization (figure 5.6) of the causality model illustrates the causal relationship between the ASF virus and its clinical symptoms in infected pigs. The central node represents the ASF virus, while the connected nodes represent different disease symptoms. The arrows indicate that the ASF virus directly causes these symptoms.

Important observations from the diagram:

- **Fever** → One of the main symptoms caused by the ASF virus.
- **Skin lesions** → The virus causes visible bleeding and discoloration on the skin.
- **Respiratory problems** → Infected pigs have difficulty breathing, probably due to internal bleeding and organ failure.
- **Digestive symptoms** → These include diarrhea and vomiting, which can lead to dehydration.
- **Mortality rate** → The ASF virus has a high mortality rate, especially in domestic pigs.
- **Cyanosis** → Bluish discoloration of the skin, especially on the limbs, due to poor oxygen circulation.
- **Loss of appetite** → ASF infections lead to reduced food intake and severe weight loss.
- **Bleeding** → Internal bleeding is one of the main pathological effects of ASF infection.
- **Nervous symptoms** → Some infected pigs may experience tremors, seizures and neurological impairments.
- **Lethargy** → ASF-infected pigs become weak, inactive and extremely tired.

Conclusion:

ASF directly causes several systemic symptoms affecting the circulatory, digestive, nervous and respiratory systems. The severity of symptoms correlates with the viral load, i.e. higher viral concentrations lead to worse symptoms and an increased mortality rate. This model helps in early diagnosis, as the presence of multiple symptoms can indicate ASF infection. This diagram confirms that ASF is a highly systemic disease that causes severe symptoms in multiple organs and leads to high mortality rates. Understanding these causal relationships will aid in ASF surveillance, early detection and disease control strategies.

6.7 Deployment and Validation of the Causality Model

The researcher set himself the task of training a machine learning model with the ASF dataset. He carefully split the data into training, validation and test datasets to ensure a robust evaluation. The model was developed to analyze important ASF-related factors, including infection trends, severity in pigs, correlations to viral load and ASF control measures.

Three different algorithms were used to evaluate the performance of the model:

1. **Random Forest Regressor:** recognized for its effectiveness in managing complex, non-linear relationships in data sets.
2. **Linear Regression:** a simple but interpretable model that assumes linear relationships between variables.
3. **Gradient Boosting Regressor:** an advanced boosting algorithm that sequentially improves prediction accuracy.

The ASF Analysis Model Evaluation table shows the performance metrics of these models and compares their ability to predict ASF virus load based on key indicators. The table includes key evaluation metrics such as the mean squared error (MSE), which measures the prediction error, and the R^2 value, which indicates how well the model explains the variance in the data set. By analyzing these models, the researcher was able to determine which approach provided the most accurate and generalizable results for ASF trend analysis and severity prediction.

The model will produce the required outputs, including:

1. Number of infected pigs per year (2015–2019)
2. Severity in domestic pigs compared to viral load
3. Causal relationship between ASF virus and symptoms
4. Correlation between ASF viral load and symptoms
5. Severity in wild boars compared to domestic pigs
6. Number of infected domestic pigs per year
7. Number of infected wild boars per year based on virus load
8. Positive and negative correlations
9. Summary statistics on ASF control measures

To understand the metrics

1. **Validation MSE (Mean Squared Error):**
    - MSE (Mean Squared Error) measures the average squared difference between the actual and predicted values.
    - A low MSE value indicates better performance, which means that the predictions of the model are close to the actual values.
2. **Validation R^2 (R-squared):**
    - R^2 (coefficient of determination) measures how well the model explains the variance in the target variable (viral load).
    - R^2 ranges from 1 (perfect fit) to negative values (worse than a mere estimate of the mean).
3. **Test MSE (Mean Square Error):**
    - Measures the error in the test data set that is separate from the training data.
    - The increase in MSE indicates that the model is fitted to the training data to some extent.

Table 6.1: Random Forest Model

Metric	Value	Explanation
Validation MSE	10.0261	The model makes an average squared error of 10.03 when making predictions for the validation dataset. A low MSE is better.
Validation R^2	0.6503	The model explains 65.03% of the variance in the validation dataset, indicating good predictive ability.
Test MSE	6.7964	The mean square error of the model on the unseen test data set is 6.80, which means that it generalizes better on test data than on validation data.
Test R^2	0.5810	The model explains 58.1% of the variance in the test data set, indicating that it has a decent generalization ability.

Interpretation:

- ✓ Random Forest performs well overall, with both validation and test R^2 values above 0.5, indicating that it captures significant patterns in the data.
- ✓ The test MSE (6.80) is lower than the validation MSE (10.03), indicating that the model generalizes well to unseen data.

Table 6.2: Linear Regression Model

Metric	Value	Explanation
Validation MSE	10.6853	The model makes an average squared error of 10.69 on the validation dataset.
Validation R^2	0.6273	The model explains 62.73% of the variance in the validation dataset.
Test MSE	2.6467	The mean square error of the model for the test data set is 2.65, which is significantly lower than the MSE of the validation.
Test R^2	0.8368	The model explains 83.68% of the variance in the test data set, i.e. it performs very well on unseen data.

Interpretation:

- ✓ Linear regression has a higher test R^2 (83.68%) than random forest (58.1%), which means that it generalizes better to unseen data.
- ✓ The test MSE (2.65) is much lower than the validation MSE (10.69), indicating that the model performs significantly better in the test group.
- ✓ The model may over-fit the test data set, i.e. it fits the test data very well, but may not generalize about future data.

Table 6.3: Gradient Boosting Model

Metric	Value	Explanation
Validation MSE	15.3509	The model makes an average squared error of 15.35 on the validation data, the worst among all three models.
Validation R^2	0.4645	The model explains 46.45% of the variance in the validation dataset, which is relatively low.
Test MSE	28.1208	The error of the model is 28.12, which is significantly higher than Random Forest and Linear Regression.
Test R^2	-0.7335	A negative R^2 means that the model performs worse than a base model that simply predicts the mean of the data.

Interpretation:

- ✓ Gradient boosting performs the worst of the three models.
- ✓ The test R^2 is negative (-0.73), which means that it cannot be generalized to unseen data.
- ✓ The high test MSE (28.12) indicates that the model makes large errors on the test data.
- ✓ Gradient boosting probably fits the training data too well but cannot be generalized to new data.

Final Comparison: Final comparison: Which model is the best?

Table 6.4: Final Comparison

Model	Validation R^2	Test R^2	Best Choice ?
Random Forest	65.03%	58.10%	✅ Good balance between validation & test performance
Linear Regression	62.73%	83.68%	✅ Best test R^2, but possible overfitting
Gradient Boosting	46.45%	-73.35%	❌ Worst performance, not suitable

Key Performance

- ✓ Linear regression has the best test performance (R^2 = 83.68%), which means that it generalizes well but may overfit.
- ✓ Random Forest is more balanced and shows consistent performance in all validation and test datasets (R^2 ~65% validation, ~58% test).
- ✓ Gradient Boosting performed the worst with a negative R^2 on the test data, which means it should not be used.

6.8 Practical implications of the findings

The findings of this study have important practical implications for the surveillance, prevention and control of African swine fever (ASF). By demonstrating a causal relationship between ASF viral load and disease symptoms, this study contributes to a better understanding of disease progression, transmission patterns and the effectiveness of intervention strategies.

One of the key findings is that early detection of ASF cases is possible by monitoring the severity of symptoms in relation to viral load. This allows farmers, veterinarians and policy makers to react quickly and, for example, quarantine infected animals before the disease spreads further. Timely intervention reduces the risk of a large-scale outbreak, minimizes losses for farmers and improves disease containment.

Furthermore, this study confirms the effectiveness of biosecurity, surveillance and culling strategies in controlling ASF. The strong negative correlation between ASF transmission rates and these control measures suggests that targeted biosecurity improvements, including strict farm access protocols, disinfection procedures and controlled animal movements, are crucial for ASF prevention. By scientifically demonstrating the effectiveness of these measures, policy makers can design evidence-based regulations that enforce mandatory testing, movement restrictions and compensation programs for farmers affected by outbreaks.

From an economic perspective, controlling ASF transmission is critical to stabilizing the pig industry and protecting global food security. Preventing large-scale outbreaks will minimize mass culling, thereby safeguarding farmers' livelihoods and reducing economic losses. Overall, these findings provide actionable insights that support greater containment of ASF and ensure a more resilient livestock industry and a sustainable global pork supply chain.

6.9 Specific requirements

This study requires comprehensive epidemiologic data on ASF, including viral load, clinical symptoms, transmission patterns, and effectiveness of intervention. Machine learning models such as linear regression and Random Forest regressor are essential to analyse causal relationships between ASF viral load and symptoms. In addition, statistical tools for correlation analysis and trend detection are needed. Biosecurity, culling and surveillance effectiveness must be included to evaluate ASF control measures. Legal frameworks and policy guidelines are needed to assess the feasibility of measures. Finally, data visualization techniques are needed to improve the interpretation of findings and decision-making in ASF containment and prevention.

6.10 Chapter Summary

Chapter 6 contains an in-depth examination of the results of the study, emphasising the most important findings through comparisons with the existing literature. This chapter plays a crucial role in linking scientific knowledge to practical applications, particularly in the management and control of African Swine Fever (ASF). By comparing study results with established research, the discussion strengthens evidence-based strategies for ASF risk management. These findings contribute to the development of effective disease control measures and ensure that scientific knowledge is translated into practical measures that protect animal health, food security and the agricultural economy. The next chapter (Chapter 7) concludes this study.

CHAPTER 7: CONCLUSION AND RECOMMENDATIONS

7.1 organisation of the chapter

After the chapter introduction in section 7.2, the assumptions and limitations of the research are discussed in section 7.3. This is followed by an assessment of the contributions of the research in section 7.4, as well as policy recommendations in section 7.5. The chapter concludes in section 7.6 and considers future research in the field in section 7.7.

7.2 Introduction

This study was initiated with a systematic literature review to solve a research problem. The concept idea was tested by developing a causal machine learning model capable of establishing small-scale relationships between ASF virus and disease symptoms using historical laboratory data. The main goal of this research was to develop and apply a CML model that extracts actionable information from ASF observation datasets to make intervention decisions for accurate ASF diagnosis. The design and development of the artefact was carried out using causal theory and predictive modelling with ML. Causal theory essentially focuses on understanding how and why certain causes lead to certain results. The study shows that it is necessary to understand the genetic, biological and clinical aspects of the ASF virus before the problem of lack of accuracy can be solved. Predictive modelling using ML for ASF diagnosis is an iterative process that requires a combination of data science expertise, domain knowledge and ongoing collaboration between data scientists and subject matter experts. It involves the development of a computer model that can predict the occurrence or severity of ASF based on input features or variables. The aim is to produce accurate and interpretable predictions that can assist in the early detection and management of ASF in pig populations. ML techniques are applied to historical data to learn patterns and relationships that allow the model to make predictions for new, unseen data.

An experimental method based on causal theory was proposed. The application of causal theory to experiments was evident in the development of an approach to evaluate the performance of the model using metrics, namely R^2 and RMSE. These metrics give different insights into the performance and predictive power of the model. The application of causal theory in experimental research has highlighted the different dimensions of the applied research approach. The contribution of this research is divided into three parts: one practical, one theoretical and one methodological. The design and development of the model contribute to the practical aspect of the research. The theoretical aspects of the research were covered by presenting a new model that can establish relationships between virus and signs of ASF accurately by applying causality theory. The application of causality theory in this experimental research is the methodological contribution.

7.3 Assumptions and limitations

The accuracy of ASF diagnosis may depend on several factors, including the diagnostic method used, the stage of disease, the quality of the sample collected and the expertise of the persons making the diagnosis. In this study, certain assumptions and limitations regarding the accuracy of ASF diagnosis were relevant.

7.3.1 Assumptions

This study assumes that the ASF data are correctly reported and reflect real transmission patterns. It assumes that machine learning models effectively capture causal relationships between ASF viral load and disease symptoms. The study also assumes that biosecurity, culling and surveillance measures are consistently implemented on all farms and that the environmental and genetic factors influencing ASF transmission remain stable. Furthermore, it is assumed that ASF outbreaks follow historical trends and that statistical correlations indicate causality. The research assumes that ASF control measures are applied consistently across regions and that ASF data sources are reliable and unbiased.

7.3.2 Limitations

This study is limited by potential inconsistencies in the reporting of ASF data, as the accuracy of the analysis may be affected by underreporting and regional differences. Machine learning models may not fully capture complex, nonlinear disease dynamics, limiting predictive accuracy. The study does not account for emerging ASF virus mutations or unpredictable environmental changes. Economic, social and political factors influencing the implementation of disease control are not fully investigated. In addition, biosecurity measures differ from region to region, which affects the generalizability of the model. The study also relies on historical data, which may not provide accurate predictions of future ASF outbreaks due to evolving risk factors and changing epidemiological conditions.

7.4 Research contributions

7.4.1 Methodological contributions

This study advances ASF research methodology by combining machine learning techniques (linear regression and random forest regressor) with epidemiologic and statistical analyses. It introduces a data-driven causality model to assess ASF transmission, symptom severity and control effectiveness to increase predictive accuracy. By incorporating quantitative trend analysis and correlation matrices, this research improves outbreak prediction and disease control strategies. The study also validates biosecurity, culling and surveillance measures using empirical data. The methodological framework of the study provides a replicable model for future

ASF studies and can be adapted for other infectious disease surveillance and intervention strategies in veterinary and public health research.

This study lays the foundation for further research on ASF, for the extension of prediction models to include environmental, genetic and socioeconomic factors. Future research can build on this work by developing real-time ASF surveillance systems and using AI-based outbreak predictions. In addition, the integration of geospatial and climate data can improve the understanding of ASF transmission routes. The study highlights the need for cross-national comparative research that examines the effectiveness of ASF policies in different regions. Further studies can also examine ASF virus mutations and vaccine development to refine disease control measures and improve global ASF containment efforts.

7.4.2 Practical Contributions

This study provides evidence-based insights for policy makers, veterinarians and farmers to improve ASF containment strategies. By confirming the effectiveness of biosecurity, culling and surveillance, it supports the implementation of stricter entry protocols, movement restrictions and early detection measures. The study helps the livestock industry to minimize economic losses by reducing the risk of ASF transmission. It also contributes to food security by preventing large-scale ASF outbreaks. In addition, the machine learning-based prediction model can be adopted by veterinary organizations to improve ASF surveillance, early warning systems and risk management strategies to prevent and control the disease.

7.4.3 Theoretical Contributions

This study extends the theories of ASF epidemiology by introducing a causal model that links viral load and disease severity and confirms the systemic impact of ASF on infected pigs. It strengthens disease modelling by integrating machine learning with epidemiological analysis and demonstrates how quantitative methods can predict ASF outbreaks. The study also refines the theory of the effectiveness of biosecurity measures and validates its role in controlling ASF transmission. In addition, it contributes to the positivist research paradigm that emphasizes data-driven decision making in veterinary science. The results provide a basis for future theoretical developments in the areas of infectious disease modelling, surveillance optimization and outbreak response.

7.5 Policy recommendations

This study recommends strengthening ASF control policies by enforcing mandatory biosecurity measures, increased surveillance and regulated culling strategies. Governments should introduce strict access controls to farms, regular ASF testing and movement restrictions to prevent the spread of the disease. Compensation programs for affected farmers should be introduced to encourage early notification. Investment in real-time ASF surveillance systems and AI-based models to predict outbreaks is crucial for proactive containment. In addition, international cooperation in the development of standardized ASF control protocols should be a priority. Policy makers need to integrate scientific evidence into the regulatory framework to ensure sustainable ASF prevention and protection of food safety in the livestock industry.

7.6 Conclusion

This study investigated the causal relationship between ASF viral load and disease symptoms and used machine learning models to analyse transmission dynamics and control efficacy. The results confirm that biosecurity, surveillance and culling significantly reduce the spread of ASF and provide evidence-based strategies for containment. The quantitative approach of the study improves predictive modelling and provides valuable insights for policy makers and veterinarians. By integrating epidemiological data with machine learning, this research contributes to ASF risk management, early detection and prevention of outbreaks. The study underlines the need for a proactive ASF control policy that ensures livestock protection, economic stability and food security in ASF-prone regions.

7.7 Future directions

Future research should expand the predictive models for ASF by incorporating climatic, genetic and economic factors. The development of AI-driven real-time surveillance systems can improve the early detection of outbreaks. Further studies should investigate ASF virus mutations and vaccine development to ensure long-term disease prevention strategies and global cooperation in ASF containment.

8. REFERENCES

1. Abidi, S.R. 2011. Ontology-based knowledge modelling to provide decision support for comorbid diseases. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6512 LNAI: 27–39. https://link.springer.com/chapter/10.1007/978-3-642-18050-7_3 [10 September 2021].
2. Amini, M., Zayeri, F. and Salehi, M., 2021. Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. *BMC public health*, *21*, pp.1-12.
3. Anwar, N., Widodo, A.M., Sekti, B.A., Ulum, M.B., Rahaman, M. and Ariessanti, H.D., 2024. Comparative analysis of nij and nist methods for microsd investigations: A technopreneur approach. *Aptisi Transactions on Technopreneurship (ATT)*, *6*(2), pp.169-181.
4. Ali, I.M., 2024. A guide for positivist research paradigm: From philosophy to methodology. *Ideology Journal*, *9*(2).
5. Asim, M.N., Wasim, M., Khan, M.U.G., Mahmood, W. and Abbasi, H.M., 2018. A survey of ontology learning techniques and applications. *Database*, *2018*, p.bay101.
6. Ahmad, K., Abdelrazek, M., Arora, C., Bano, M. and Grundy, J., 2023. Requirements practices and gaps when engineering human-centered Artificial Intelligence systems. *Applied Soft Computing*, *143*, p.110421.
7. Alharahsheh, H.H. and Pius, A., 2020. A review of key paradigms: Positivism VS interpretivism. *Global Academic Journal of Humanities and Social Sciences*, *2*(3), pp.39-43.
8. Achenbach, J.E., Gallardo, C., Nieto-Pelegrín, E., Rivera-Arroyo, B., Degefa-Negi, T., Arias, M., Jenberie, S., Mulisa, D.D., Gizaw, D., Gelaye, E. & Chibssa, T.R. 2015. Identification of a new genotype of African swine fever virus in domestic pigs from Ethiopia. *Transboundary and emerging diseases*, 64(5):1393-1404.
9. Akhiat, Y., Bouchlaghem, Y., Zinedine, A. and Chahhou, M., 2024. Explainable Machine Learning Method for Predicting Road-Traffic Accident Injury Severity in Addis Ababa City Based on a New Graph Feature Selection Technique. In *Explainable Artificial Intelligence for Intelligent Transportation Systems* (pp. 53-74). CRC Press.
10. Alzyoud, M., Alazaidah, R., Aljaidi, M., Samara, G., Qasem, M., Khalid, M. and Al-Shanableh, N., 2024. Diagnosing diabetes mellitus using machine learning techniques. *International Journal of Data and Network Science*, *8*(1), pp.179-188.
11. Achenbach, S., Motwani, M., Dey, D., Berman, D.S., Germano, G., Al-Mallah, M.H., Andreini, D., Budoff, M.J., Cademartiri, F., Callister, T.Q. & Chang, H.J. 2015. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *European Heart Journal*, 38(7):500-507.
12. Ahrendt, P., Gregersen, T. & Karstoft, H. 2011. Development of a real-time computer vision system for tracking loose-housed pigs. *Computers and Electronics in Agriculture*, 76(2):169-174.
13. Ahsan, M.M., Luna, S.A. & Siddique, Z. 2022. Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare,* 10(3):541.
14. Alameer, A., Buijs, S., O'Connell, N., Dalton, L., Larsen, M., Pedersen, L. & Kyriazakis, I. 2022. Automated detection and quantification of contact behaviour in pigs using deep learning. *Biosystems Engineering*, 224:118-130.
15. Alharahsheh, H.H. & Pius, A. 2020. A review of key paradigms: Positivism vs interpretivism. *Global Academic Journal of Humanities and Social Science*, 2(3):39-43. https://www.researchgate.net/publication/338244145 [15 July 2021].

16. Abdel-Basset, M., Mohamed, R., Jameel, M. and Abouhawwash, M., 2023. Nutcracker optimizer: A novel nature-inspired metaheuristic algorithm for global optimization and engineering design problems. *Knowledge-Based Systems*, *262*, p.110248.

17. Bora, S.L., Bhuyan, K., Hazarika, P.J., Gogoi, J. and Goswami, K., 2022. Analysis of rainfall trend using non-parametric methods and innovative trend analysis during 1901–2020 in seven states of North East India. *Current Science*, *122*(7), pp.801-811.

18. Alić, B., Gurbeta, L. and Badnjević, A., 2017, June. Machine learning techniques for classification of diabetes and cardiovascular diseases. In *2017 6th mediterranean conference on embedded computing (MECO)* (pp. 1-4). IEEE.

19. Abualigah, L., Elaziz, M.A., Khasawneh, A.M., Alshinwan, M., Ibrahim, R.A., Al-Qaness, M.A., Mirjalili, S., Sumari, P. and Gandomi, A.H., 2022. Meta-heuristic optimization algorithms for solving real-world mechanical engineering design problems: a comprehensive survey, applications, comparative analysis, and results. *Neural Computing and Applications*, pp.1-30.

20. Alshehri, M., Drew, S. & Alghamdi, R. 2013. Analysis of citizens acceptance for e-government services: Applying the UTAUT model. *ArXiv, abs/1304.3157*.https://www.semanticscholar.org/paper/Analysis-of-Citizens-Acceptance-for-E-government- [7 December 2023].

21. Borland, D., Wang, A.Z. and Gotz, D., 2024. Using Counterfactuals to Improve Causal Inferences from Visualizations. arXiv preprint arXiv:2401.08411.

22. Badgeley, M.A., Zech, J.R., Oakden-Rayner, L., Glicksberg, B.S., Liu, M., Gale, W., McConnell, M.V., Percha, B., Snyder, T.M. & Dudley, J.T. 2019. Deep learning predicts hip fracture using confounding patient and healthcare variables. *NPJ Digital Medicine*, 2(1):1-10.

23. Burr, C. and Leslie, D., 2023. Ethical assurance: a practical approach to the responsible design, development, and deployment of data-driven technologies. *AI and Ethics*, *3*(1), pp.73-98.

24. Borraccini, G., D'Amico, A., Straullu, S., Usmani, F., Ahmad, A. and Curri, V., 2023. Iterative supervised learning approach using transceiver bit-error-rate measurements for optical line system optimization. *Journal of Optical Communications and Networking*, *15*(2), pp.111-118.

25. BÜYÜKKEÇECİ, Mustafa, and Mehmet Cudi OKUR. "A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning." *Gazi University Journal of Science* (2024): 1-1.

26. Brutger, R., Kertzer, J.D., Renshon, J., Tingley, D. and Weiss, C.M., 2023. Abstraction and detail in experimental design. *American Journal of Political Science*, *67*(4), pp.979-995.

27. Balestreri, C., Schroeder, D.C., Sampedro, F., Marqués, G., Palowski, A., Urriola, P.E., van de Ligt, J.L., Yancy, H.F. and Shurson, G.C., 2024. Unexpected thermal stability of two enveloped megaviruses, Emiliania huxleyi virus and African swine fever virus, as measured by viability PCR. *Virology Journal*, *21*(1), p.1.

28. Buczak, P., Huang, H., Forthmann, B. and Doebler, P., 2023. The machines take over: A comparison of various supervised learning approaches for automated scoring of divergent thinking tasks. *The Journal of Creative Behavior*, *57*(1), pp.17-36.
29. Bandi, A., Adapa, P.V.S.R. and Kuchi, Y.E.V.P.K., 2023. The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. Future Internet, 15(8), p.260.
30. Bell, R.Q., 2020. Research strategies. *Child effects on adults*, pp.85-121.
31. Balke, A. & Pearl, J. 1994. Counterfactual probabilities: Computational methods, bounds and applications. In Da Mantaras, R.L. & Poole, D. (eds). *UAI '94: Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, July 29-31. Seattle WA, 46-54. https://dl.acm.org/doi/proceedings/10.5555/2074394 [7 December 2023].
32. Baskerville, R.L., Kaul, M. and Storey, V.C., 2015. Genres of inquiry in design-science research. *Mis Quarterly*, *39*(3), pp.541-564.
33. Borella, L., Novello, G., Gasparotto, M., Renella, G., Roverso, M., Bogialli, S., Filippini, F. and Sforza, E., 2023. Design and experimental validation of an optimized microalgae-bacteria consortium for the bioremediation of glyphosate in continuous photobioreactors. *Journal of Hazardous Materials*, *441*, p.129921.
34. Battineni, G., Sagaro, G.G., Chinatalapud, N. & Amenta, F. 2020. Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of Personalized Medicine*, 10(2):21. https://doi.org/10.3390/jpm10020021.
35. Beeram, R., Vepa, K.R. & Soma, V.R. 2023. Recent trends in SERS-based plasmonic sensors for disease diagnostics, biomolecules detection, and machine learning techniques. *Biosensors*, 13(3):328. https://doi.org/10.3390/bios13030328.

36. Bogna, F., Raineri, A. & Dell, G. 2020. Critical realism and constructivism: Merging research paradigms for a deeper qualitative study. *Qualitative Research in Organizations and Management: An International Journal*, 15(4):461-484. https://doi.org/10.1108/QROM-06-2019-1778.
37. Briner, R.B. & Denyer, D. 2012. Systematic review and evidence synthesis as a practice and scholarship tool. In Rousseau, D.M. (ed.) *The Oxford Handbook of Evidence-Based Management*, Oxford Library of Psychology. 112-129. https://doi.org/10.1093/OXFORDHB%2F9780199763986.013.0007.
38. Bru, G., Martínez-Candela, M., Romero, P., Navarro, A. & Martínez-Murcia, A. 2023. Internal validation of the ASFV MONODOSE dtec-qPCR Kit for African Swine Fever virus detection under the UNE-EN ISO/IEC 17025: 2005 criteria. *Veterinary Sciences*, 10(9):564. https://doi.org/10.3390/vetsci10090564.
39. Busch, F., Haumont, C., Penrith, M.L., Laddomada, A., Dietze, K., Globig, A., Guberti, V., Zani, L. & Depner, K. 2021. Evidence-based African swine fever policies: Do we address virus and host adequately? *Frontiers in Veterinary Science*, 8:637487. https://doi.org/10.3389/fvets.2021.637487.
40. Camarinha-Matos, L.M. & Afsarmanesh, H. 2012. Collaborative networks in active ageing – A roadmap contribution to demographic sustainability. *Production Planning & Control*: *The Management of Operations*, 23(4):279-298. https://doi.org/10.1080/09537287.2011.627659.
41. Cater-Steel, A., Toleman, M. & Rajaeian, M.M. 2019. Design science research in doctoral projects: An analysis of Australian theses. *Journal of the Association for Information Systems*, 20(12):1844-1869. https://doi.org/10.17705/1jais.00587.
42. Creswell, J.W. and Clark, V.P., 2011. *Mixed methods research*. SAGE Publications.
43. Chi, R., Lin, P.Y., Jhuo, Y.S., Cheng, F.Y. and Ho, J.A.A., 2024. Colorimetric detection of African swine fever (ASF)-associated microRNA based on rolling circle amplification and salt-induced gold nanoparticle aggregation. *Talanta*, *267*, p.125159
44. Challoumis-Κωνσταντίνος Χαλλουμής, C., 2024. Comparative analysis between capital and liability-Sensitivity Method. *Open Journal for Research in Economics*, *7*(2), pp.10-32591.
45. Cavalera, S., Colitti, B., De Mia, G.M., Feliziani, F., Dei Giudici, S., Angioi, P.P., D'Errico, F., Scalas, D., Scollo, A., Serra, T. and Chiarello, M., 2023. Development of molecular and antigenic-based rapid tests for the identification of African swine fever virus in different tissues. *Talanta*, *258*, p.124443.
46. Chen, P.H.C., Liu, Y. & Peng, L. 2019. How to develop machine learning models for healthcare. *Nature Materials*, 18(5):410-414. http://dx.doi.org/10.1038/s41563-019-0345-0.
47. Creswell, J.W. 2008. *Research design: Qualitative, quantitative, and mixed methods approaches*. 4th ed. London: SAGE. 3–22.
48. Chen, X., Hong, X., Gao, W., Luo, S., Cai, J., Liu, G. and Huang, Y., 2022. Causal relationship between physical activity, leisure sedentary behaviors and COVID-19 risk: a Mendelian randomization study. *Journal of translational medicine*, *20*(1), p.216.
49. Creswell, J.W. & Plano Clark, V.L. 2007. *Designing and conducting mixed methods research* 1(2): 24–27.
50. Cui, P. & Athey, S. 2022. Stable learning establishes some common ground between causal inference and machine learning. *Nature Machine Intelligence*, 4(2):110-115. http://dx.doi.org/10.1038/s42256-022-00445-z.
51. Dixon, L.K., Stahl, K., Jori, F., Vial, L. & Pfeiffer, D.U. 2020. African swine fever epidemiology and control. *Annual Review of Animal Biosciences,* 8:221-246. https://doi.org/10.1146/annurev-animal-021419-083741.
52. Dayan, C.B., Son, D., Aghakhani, A., Wu, Y., Demir, S.O. and Sitti, M., 2024. Machine Learning-Based Shear Optimal Adhesive Microstructures with Experimental Validation. *Small*, *20*(2), p.2304437.
53. De Carlo, A., Tosca, E.M., Melillo, N. and Magni, P., 2023. mvLognCorrEst: an R package for sampling from multivariate lognormal distributions and estimating correlations from uncomplete correlation matrix. *Computer Methods and Programs in Biomedicine*, *235*, p.107517.
54. Dey, K., Kalita, K. and Chakraborty, S., 2023. Prediction performance analysis of neural network models for an electrical discharge turning process. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *17*(2), pp.827-845.
55. Dinhobl, M., Spinard, E., Tesler, N., Birtley, H., Signore, A., Ambagala, A., Masembe, C., Borca, M.V. and Gladue, D.P., 2023. Reclassification of ASFV into 7 Biotypes Using Unsupervised Machine Learning. *Viruses*, *16*(1), p.6.
56. Ding, W., Abdel-Basset, M., Hawash, H. and Ali, A.M., 2022. Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey. *Information Sciences*.

57. Dou, B., Zhu, Z., Merkurjev, E., Ke, L., Chen, L., Jiang, J., Zhu, Y., Liu, J., Zhang, B. and Wei, G.W., 2023. Machine learning methods for small data challenges in molecular science. *Chemical Reviews*, *123*(13), pp.8736-8780.
58. Fernández-Carrión, E., Barasona, J.Á., Sánchez, Á., Jurado, C., Cadenas-Fernández, E. & Sánchez-Vizcaíno, J.M. 2020. Computer vision applied to detect lethargy through animal motion monitoring: A trial on African swine fever in wild boar. *Animals*, 10(12):2241. https://doi.org/10.3390/ani10122241.
59. Franzoni, G., Pedrera, M. and Sánchez-Cordón, P.J., 2023. African swine fever virus infection and cytokine response in vivo: an update. *Viruses*, *15*(1), p.233.
60. Founta, K., Dafou, D., Kanata, E., Sklaviadis, T., Zanos, T.P., Gounaris, A. & Xanthopoulos, K. 2023. Gene targeting in amyotrophic lateral sclerosis using causality-based feature selection and machine learning. *Molecular Medicine*, 29(1):12. https://doi.org/10.1186/s10020-023-00603-y.
61. Frazer, L. & Stokes, D.J. 1997. Franchising operational units in Australia. *Franchising Research. Journal of Consumer Marketing*, 2(1):32-42. https://research.usq.edu.au/item/q059z/franchising-operational-units-in-australia [7 December 2023].
62. Forth, J.H., Calvelage, S., Fischer, M., Hellert, J., Sehl-Ewert, J., Roszyk, H., Deutschmann, P., Reichold, A., Lange, M., Thulke, H.H. and Sauter-Louis, C., 2023. African swine fever virus–variants on the rise. *Emerging Microbes & Infections*, *12*(1), p.2146537.
63. Frisch, C. & Huppenbauer, M. 2014. New insights into ethical leadership: A qualitative investigation of the experiences of executive ethical leaders. *Journal of Business Ethics*, 123(1):23-43. http://dx.doi.org/10.1007/s10551-013-1797-9.
64. Gallardo, C., Fernández-Pinero, J. & Arias, M. 2019. African swine fever (ASF) diagnosis, an essential tool in the epidemiological investigation. *Virus Research,* 271:97676. https://doi.org/10.1016/j.virusres.2019.197676.
65. Gallaire, H., Minker, J. and Nicolas, J.M., 1984. Logic and databases: A deductive approach. *ACM Computing Surveys (CSUR)*, *16*(2), pp.153-185.
66. Go, H., Kim, S.W. and Lee, Y., 2023. Design and Implementation of a Programming Automatic Assessment System in Jupyter Notebook. *International Journal on Advanced Science, Engineering & Information Technology*, *13*(3).
67. Gao, H., Gao, X., Liu, X., Luo, Y., Zhong, J., Liu, J., Yan, L., Wang, H., Gong, L., Zhang, G. and Zheng, Z., 2024. African swine fever virus maintains de novo global cellular protein synthesis and inhibits stress granules formation via dephosphorylating eIF2α. *Veterinary Microbiology*, p.109988.
68. Go, J. and Chen, P., 2023. Accelerating Bayesian Optimal Experimental Design with Derivative-Informed Neural Operators. *arXiv preprint arXiv:2312.14810*.
69. Gallardo, C., Soler, A., Nieto, R., Sánchez, M.A., Martins, C., Pelayo, V., Carrascosa, A., Revilla, Y., Simón, A., Briones, V. & Sánchez-Vizcaíno, J.M. 2015. Experimental transmission of African swine fever (ASF) low virulent isolate NH/P68 by surviving pigs. *Transboundary and Emerging Diseases*, 62(6):612-622. https://doi.org/10.1111/tbed.12431.
70. Gaudreault, N.N., Madden, D.W., Wilson, W.C., Trujillo, J.D. and Richt, J.A., 2020. African swine fever virus: an emerging DNA arbovirus. Frontiers in veterinary science, 7, p.215.
71. Gervasi, V., Marcon, A., Bellini, S. & Guberti, V. 2019. Evaluation of the efficiency of active and passive surveillance in the detection of African swine fever in wild boar. *Veterinary Sciences*, *7*(1):5. https://doi.org/10.3390/vetsci7010005.
72. Greenhill, S., Rana, S., Gupta, S., Vellanki, P. and Venkatesh, S., 2020. Bayesian optimization for adaptive experimental design: A review. *IEEE access*, *8*, pp.13937-13948.
73. Gomez-Vazquez, J.P. and Martinez-Lopez, B., 2023, September. P06. 4 Modeling the spread of ASF into the US swine industry: the example of Iowa. In *GeoVet 2023 International Conference*.
74. Grampurohit, S. & Sagarnal, C. 2020. Disease prediction using machine learning algorithms. *In 2020 International Conference for Emerging Technology (INCET),* Belgaum, India. 5-7 June. 1-7. https://doi.org/10.1109/incet49848.2020.9154130.
75. Gregor, S. & Hevner, A.R. 2013. Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2):337-355. http://dx.doi.org/10.25300/MISQ/2013/37.2.01.
76. Grossoehme, D.H. 2014. Overview of qualitative research. *Journal of Health Care Chaplaincy*, 20(3):109-122. http://dx.doi.org/10.1080/08854726.2014.925660.
77. Haines, F.J., Hofmann, M.A., King, D.P., Drew, T.W. & Crooke, H.R. 2013. Development and validation of a multiplex, real-time RT PCR assay for the simultaneous detection of classical

and African swine fever viruses. *PloS One*, 8(7):e71019. http://dx.doi.org/10.1371/journal.pone.0071019.
78. Hevner, A.R., March, S.T., Park, J. & Ram, S. 2004. Essay in information design science systems. *Management Information Systems Quarterly*, 28(1):75-105.
79. Halev, A., Martínez-López, B., Clavijo, M., Gonzalez-Crespo, C., Kim, J., Huang, C., Krantz, S., Robbins, R. and Liu, X., 2023. Infection prediction in swine populations with machine learning. *Scientific Reports*, *13*(1), p.17738.
80. Han, J., Sarica, S., Shi, F. and Luo, J., 2022. Semantic networks for engineering design: state of the art and future directions. *Journal of Mechanical Design*, *144*(2), p.020802.
81. Hocking, R.R. 1976. The analysis and selection of variables in linear regression. *Biometrics*, 32(1):1. https://doi.org/10.2307/2529336.
82. Huang, X., Liu, B., Guo, S., Guo, W., Liao, K., Hu, G., Shi, W., Kuss, M., Duryee, M.J., Anderson, D.R. & Lu, Y. 2023. SERS spectroscopy with machine learning to analyze human plasma derived sEVs for coronary artery disease diagnosis and prognosis. *Bioengineering & Translational Medicine*, 8(2):e10420. https://doi.org/10.1002/btm2.10420.
83. Idrizi, E., Filiposka, S. and Trajkovikj, V., 2023. Gender impact on STEM online learning-a correlational study of gender, personality traits and learning styles in relation to different online teaching modalities. *Multimedia Tools and Applications*, pp.1-19.
84. Ibrahimi, E., Lopes, M.B., Dhamo, X., Simeon, A., Shigdel, R., Hron, K., Stres, B., D'Elia, D., Berland, M. and Marcos-Zambrano, L.J., 2023. Overview of data preprocessing for machine learning applications in human microbiome research. *Frontiers in Microbiology*, *14*.
85. Javaid, M., Sarfraz, M.S., Aftab, M.U., Zaman, Q.U., Rauf, H.T. & Alnowibet, K.A. 2023. WebGIS-based real-time surveillance and response system for vector-borne infectious diseases. *International Journal of Environmental Research and Public Health*, 20(4):3740. https://doi.org/10.3390/ijerph20043740.
86. James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J., 2023. Linear regression. In *An introduction to statistical learning: With applications in python* (pp. 69-134). Cham: Springer International Publishing.
87. Jiao, R., Zhang, Y., Ding, L., Xue, B., Zhang, J., Cai, R. and Jin, C., 2023. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *Computers in Biology and Medicine*, p.107840.
88. Jiang, W. and Chen, L., 2021. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Computational and structural biotechnology journal*, *19*, pp.183-195.
89. Kaddour, J., Lynch, A., Liu, Q., Kusner, M.J. & Silva, R. 2022. Causal machine learning: A survey and open problems. *arXiv:2206.15475*. https://doi.org/10.48550/arXiv.2206.15475.
90. Kassambara, A. 2018. Stepwise regression essentials in R - Articles - STHDA, STHDA. http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/154- stepwise-regression-essentials-in-r/ [16 May 2023].
91. Kulmanov, M., Smaili, F.Z., Gao, X. and Hoehndorf, R., 2021. Semantic similarity and machine learning with ontologies. *Briefings in bioinformatics*, *22*(4), p.bbaa199.
92. Kaushik, V. & Walsh, C.A. 2019a. Pragmatism as a research paradigm and its implications for social work research. *Social Sciences*, 8(9):255. https://doi.org/10.3390/socsci8090255.
93. Kment, B., 2023. Decision, Causality, and Predetermination. *Philosophy and Phenomenological Research*.
94. Khan, R.A., Luo, Y. & Wu, F.X. 2022. Machine learning based liver disease diagnosis: A systematic review. *Neurocomputing*, 468:492-509. https://doi.org/10.1016/j.neucom.2021.08.138.
95. Kwon, O.K., Kim, D.W., Heo, J.H., Kim, J.Y., Nah, J.J., Choi, J.D., Lee, D.W., Cho, K.H., Hong, S.K., Kim, Y.H. and Kang, H.E., 2024. Genomic Epidemiology of African Swine Fever Virus Identified in Domestic Pig Farms in South Korea during 2019–2021. *Transboundary and Emerging Diseases*, *2024.*
96. Kolukisa, B. & Bakir-Gungor, B. 2023. Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis. *Computer Standards & Interfaces*, 84(2):103706. http://dx.doi.org/10.1016/j.csi.2022.103706.
97. Kosar, T., Bohra, S. & Mernik, M. 2018. A systematic mapping study was driven by the margin of error. *Journal of Systems and Software*, 144:439-449. https://doi.org/10.1016/j.jss.2018.06.078.
98. Krichen, M., 2024. Exploring the Feasibility of Formal Methods in Machine Learning and Artificial Intelligence.

99. Kucharski, A.J., Chung, K., Aubry, M., Teiti, I., Teissier, A., Richard, V., Russell, T.W., Bos, R., Olivier, S. & Cao-Lormeau, V.M. 2023. Real-time surveillance of international SARS-CoV-2 prevalence using systematic traveller arrival screening: An observational study. PLoS medicine, 20(9):e1004283. http://dx.doi.org/10.1371/journal.pmed.1004283.
100. Kacena, M.A., Plotkin, L.I. and Fehrenbacher, J.C., 2024. The use of artificial intelligence in writing scientific review articles. *Current Osteoporosis Reports*, pp.1-7.
101. Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A. & Paszkiewicz, I. 2023. What Is machine learning, artificial neural networks and deep learning? Examples of practical applications in medicine. *Diagnostics*, 13(15):2582. http://dx.doi.org/10.3390/diagnostics13152582.
102. Krushynska, A.O., Torrent, D., Aragón, A.M., Ardito, R., Bilal, O.R., Bonello, B., Bosia, F., Chen, Y., Christensen, J., Colombi, A. and Cummer, S.A., 2023. Emerging topics in nanophononics and elastic, acoustic, and mechanical metamaterials: an overview. *Nanophotonics*, *12*(4), pp.659-686.
103. Lancaster, G. 2007. *Research methods in management*. London: Routledge.
104. Lacey, H., 2020. A view of scientific methodology as source of ignorance in the controversies about genetically engineered crops. *Science and the Production of Ignorance: When the quest for knowledge is thwarted*, pp.245-70.
105. Li, Y., Liang, M., Gao, L., Ahmed, M.A., Uy, J.P., Cheng, C., Zhou, Q. and Sun, C., 2021. Face masks to prevent transmission of COVID-19: A systematic review and meta-analysis. *American journal of infection control*, *49*(7), pp.900-906.
106. Lei, H. & Govindaraju, V. 2004. Regression time warping for similarity measure of sequence. *The Fourth International Conference on Computer and Information Technology*, Wuhan, China, 16 September 2004. CIT '04. IEEE. 826-830. https://doi.org/10.1109/CIT.2004.1357297.
107. Leivers, S., Campbell, T., Bodenchuk, M. & Tomeĉek, J. 2023. Behavior of wild pigs toward conspecific carcasses: Implications for disease transmission in a hot, semiarid climate. *Transboundary and Emerging Diseases*, 2023(12):1-10. http://dx.doi.org/10.1155/2023/4195199.
108. Li, L., Wang, B., Feng, P., Li Liu, D., He, Q., Zhang, Y., Wang, Y., Li, S., Lu, X., Yue, C. & Li, Y. 2022. Developing machine learning models with multi-source environmental data to predict wheat yield in China. *Computers and Electronics in Agriculture*, 194:106790. https://opus.lib.uts.edu.au/handle/10453/166322 [11 December 2023].
109. Li, S., Xiong, L., Tang, G. & Strobl, J. 2020. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology*, 354(12):107045. https://doi.org/10.1016/j.geomorph.2020.107045.
110. Li, X., Guo, R., Lu, J., Chen, T. & Qian, X. 2023. Causality-driven graph neural network for early diagnosis of pancreatic cancer in non-contrast Computerized Tomography. *IEEE Transactions on Medical Imaging*, 42(6):1656-1667. https://doi.org/10.1109/tmi.2023.3236162.
111. López-Pernas, S., Saqr, M. and Apiola, M., 2023. Scientometrics: a concise introduction and a detailed methodology for mapping the scientific field of computing education research. Past, Present and Future of Computing Education Research: A Global Perspective, pp.79-99.
112. Liang, H., Tsui, B.Y., Ni, H., Valentim, C.C., Baxter, S.L., Liu, G., Cai, W., Kermany, D.S., Sun, X., Chen, J., He, L. et al. 2019. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nature Medicine*, 25(3):433-438. https://doi.org/10.1038/s41591-018-0335-9.
113. Lichoti, J.K., Davies, J., Maru, Y., Kitala, P.M., Githigia, S.M., Okoth, E., Bukachi, S.A., Okuthe, S. & Bishop, R.P. 2017. Pig traders' networks on the Kenya-Uganda border highlight potential for mitigation of African swine fever virus transmission and improved ASF disease risk management. *Preventive Veterinary Medicine*, 140:87-96. https://doi.org/10.1016/j.prevetmed.2017.03.005.
114. Lin, J., Wang, K., Chen, Z., Liang, X. & Lin, L. 2023. Towards causality-aware inferring: A sequential discriminative approach for medical diagnosis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45:13363-13375.
115. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B. & Sánchez, C.I. 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60-88. http://dx.doi.org/10.1016/j.media.2017.07.005

116. Lopez-Paz, D., Muandet, K., Schölkopf, B. & Tolstikhin, I. 2015. Towards a learning theory of cause-effect inference. *Proceedings of the* 32nd *International Conference on Machine Learning*, Proceedings of Machine Learning Research. 37:1452-1461. https://proceedings.mlr.press/v37/lopez-paz15.html [7 December 2023].
117. Lowe, A. 2001. *Reflexive methodology: New vistas for qualitative research*. 3rd ed. London: SAGE
118. Lübke, K., Gehrke, M., Horst, J. & Szepannek, G. 2020. Why we should teach causal inference: Examples in linear regression with simulated data. *Journal of Statistics Education*, 28(2):133-139. https://doi.org/10.1080/10691898.2020.1752859.
119. Li, L., Yu, H. and Kunc, M., 2024. The impact of forum content on data science open innovation performance: A system dynamics-based causal machine learning approach. *Technological Forecasting and Social Change*, *198*, p.122936.
120. Liu, C.H., Peng, C.H., Huang, L.Y., Chen, F.Y., Kuo, C.H., Wu, C.Z. and Cheng, Y.F., 2024. Comparison of multiple linear regression and machine learning methods in predicting cognitive function in older Chinese type 2 diabetes patients. *BMC neurology*, *24*(1), p.11.
121. Luo, W., Phung, D., Tran, T., Gupta, S., Rana, S., Karmakar, C., Shilton, A., Yearwood, J., Dimitrova, N., Ho, T.B. & Venkatesh, S. 2016. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *Journal of medical Internet research*, 18(12):e323. https://doi.org/10.2196/jmir.5870.
122. Malhotra, R. 2015. A systematic review of machine learning techniques for software fault prediction. *Applied Soft Computing*, 27:504-518.
123. Mehinagic, K., Liniger, M., Samoilenko, M., Soltermann, N., Gerber, M. and Ruggli, N., 2024. A sensitive luciferase reporter assay for the detection of infectious African swine fever virus. *Journal of virological methods*, *323*, p.114854.
124. Marcinkevičs, R. & Vogt, J.E. 2020. Interpretability and explainability: A machine learning zoo mini-tour. *arXiv preprint arXiv:2012.01805.* https://doi.org/10.48550/arXiv.2012.01805.
125. Miao, G., Jiang, X., Yang, D., Fu, Q., Zhang, L., Ge, S., Ye, X., Xia, N., Qian, S. & Qiu, X. 2022. A hand-held, real-time, AI-assisted capillary convection PCR system for point-of-care diagnosis of African swine fever virus. *Sensors and Actuators B: Chemical*, 358:131476. https://doi.org/10.1016/j.snb.2022.131476
126. Miller, C.J., Smith, S.N. and Pugatch, M., 2020. Experimental and quasi-experimental designs in implementation research. *Psychiatry research*, *283*, p.112452.
127. Milton, A.A.P., Das, S., Khan, S., Momin, K.M., Prasad, C.B., Kylla, H., Ghatak, S. & Sen, A. 2023. Novel sensitive isothermal-based diagnostic technique for the detection of African swine fever virus. *Archives of Virology*, 168(3):79. http://dx.doi.org/10.1007/s00705-023-05702-z
128. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. & PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of internal medicine*, 151(4):264-269. https://doi.org/10.1371/journal.pmed.1000097.
129. Morelle, K., Barasona, J.A., Bosch, J., Heine, G., Daim, A., Arnold, J., Bauch, T., Kosowska, A., Cadenas-Fernández, E., Aviles, M.M. & Zuñiga, D. 2023. Accelerometer-based detection of African swine fever infection in wild boar. *Proceedings of the Royal Society B*, 290(2005):20231396. https://doi.org/10.1098/rspb.2023.1396.
130. Mohamed Shaffril, H.A., Samsuddin, S.F. and Abu Samah, A., 2021. The ABC of systematic literature review: the basic methodological guidance for beginners. *Quality & Quantity*, *55*, pp.1319-1346.
131. Motwani, M., Dey, D., Berman, D.S., Germano, G., Achenbach, S., Al-Mallah, M.H., Andreini, D., Budoff, M.J., Cademartiri, F., Callister, T.Q. & Chang, H.J. 2017. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *European Heart Journal*, 38(7):500-507. https://doi.org/10.1093/eurheartj/ehw188

132. Mullany, L. & Stockwell, P. 2021. Qualitative, quantitative and mixed methods research (Dörnyei). file:///C:/Users/Harrison/Downloads/John W. Creswell & J. David Creswell - Research Design_ Qualitative, Quantitative, and Mixed Methods Approaches (2018). pdf%0Afile:///C:/Users/Harrison/AppData/Local/Mendeley Ltd./Mendeley Desktop/Downloaded/Creswell, Cr.

133. Mou, T., Pillai, H.S., Wang, S., Wan, M., Han, X., Schweitzer, N.M., Che, F. and Xin, H., 2023. Bridging the complexity gap in computational heterogeneous catalysis with machine learning. *Nature Catalysis*, *6*(2), pp.122-136

134. Muñoz, F., Pleydell, D.R. & Jori, F. 2022. A combination of probabilistic and mechanistic approaches for predicting the spread of African swine fever on Merry Island. *Epidemics,* 40:100596. https://doi.org/10.1016/j.epidem.2022.100596.

135. Nematollahi, H., Moslehi, M., Aminolroayaei, F., Maleki, M. and Shahbazi-Gahrouei, D., 2023. Diagnostic Performance Evaluation of Multiparametric Magnetic Resonance Imaging in the Detection of Prostate Cancer with Supervised Machine Learning Methods. *Diagnostics*, *13*(4), p.806.

136. Nogueira, A.R., Pugnana, A., Ruggieri, S., Pedreschi, D. and Gama, J., 2022. Methods and tools for causal discovery and causal inference. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *12*(2), p.e1449.

137. Nguyen, X.C., Nguyen, T.P., Lam, V.S., Le, P.C., Vo, T.D.H., Hoang, T.H.T., Chung, W.J., Chang, S.W. and Nguyen, D.D., 2024. Estimating ammonium changes in pilot and full-scale constructed wetlands using kinetic model, linear regression, and machine learning. *Science of The Total Environment*, *907*, p.168142.

138. Onnela, J.P., Chakraborti, A., Kaski, K., Kertesz, J. & Kanto, A. 2003. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 68(5):056110. http://dx.doi.org/10.1103/PhysRevE.68.056110.

139. Olesen, A.S., Lohse, L., Accensi, F., Goldswain, H., Belsham, G.J., Bøtner, A., Netherton, C.L., Dixon, L.K. and Portugal, R., 2024. Inefficient Transmission of African Swine Fever Virus to Sentinel Pigs from an Environment Contaminated by ASFV-Infected Pigs under Experimental Conditions. *Transboundary and Emerging Diseases*, *2024*.

140. Openja, M., Laberge, G. and Khomh, F., 2024. Detection and Evaluation of bias-inducing Features in Machine learning. *Empirical Software Engineering*, *29*(1), p.22.

141. Oviedo, F., Ferres, J.L., Buonassisi, T. & Butler, K.T. 2022. Interpretable and explainable machine learning for materials science and chemistry. *Accounts of Materials Research*, 3(6):597-607. https://doi.org/10.1021/accountsmr.1c00244

142. Parsons, D.J., Green, D.M., Schofield, C.P. & Whittemore, C.T. 2007. Real-time control of pig growth through an integrated management system. *Biosystems Engineering*, 96(2):257-266. http://dx.doi.org/10.1016/j.biosystemseng.2006.10.013.

143. Pavone, S., Iscaro, C., Dettori, A. and Feliziani, F., 2023. African Swine Fever: The State of the Art in Italy. *Animals*, *13*(19), p.2998.

144. Pearl, J. 2009. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96-146. http://dx.doi.org/10.1214/09-SS057.

145. Paturi, U.M.R., Palakurthy, S.T. and Reddy, N.S., 2023. The role of machine learning in tribology: a systematic review. *Archives of Computational Methods in Engineering*, *30*(2), pp.1345-1397.

146. Pishgar, Farhad, Mahsima Shabani, Thiago Quinaglia AC Silva, David A. Bluemke, Matthew Budoff, R. Graham Barr, Matthew A. Allison, Wendy S. Post, João AC Lima, and

Shadpour Demehri. "Quantitative analysis of adipose depots by using chest CT and associations with all-cause mortality in chronic obstructive pulmonary disease: longitudinal analysis from MESArthritis ancillary study." *Radiology* 299, no. 3 (2021): 703-711.

147. Pandey, J., 2019. Deductive approach to content analysis. In *Qualitative techniques for workplace data analysis* (pp. 145-169). IGI Global.

148. Pearl, J. 2019. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3):54-60. http://dx.doi.org/10.1145/3241036

149. Pearl, J., Glymour, M. & Jewell, N.P. 2016*. Causal inference in statistics: A primer*. Hoboken NJ: Wiley.

150. Pearson, K., 2023. National life from the standpoint of science. In *Scientific and Medical Knowledge Production, 1796-1918* (pp. 281-287). Routledge.

151. Pickvance, C.G., 2001. Four varieties of comparative analysis. *Journal of housing and the built environment*, *16*, pp.7-28.

152. Patel, R.A. and Webb, M.A., 2023. Data-driven design of polymer-based biomaterials: high-throughput simulation, experimentation, and machine learning. *ACS Applied Bio Materials*.

153. Penrith, M., Okoth, E. & Heath, L. 2023. African swine fever and other swine viral Diseases in Africa [Special Issue]. *Viruses*, 15(7):1438. https://doi.org/10.3390/v15071438

154. Pikalo, J., Deutschmann, P., Fischer, M., Roszyk, H., Beer, M. &Blome, S. 2021. African Swine Fever Laboratory Diagnosis—Lessons Learned from Recent Animal Trials. *Pathogens*, 10(2):177. http://dx.doi.org/10.3390/pathogens10020177

155. Qiu, Z., Li, Z., Yan, Q., Li, Y., Xiong, W., Wu, K., Li, X., Fan, S., Zhao, M., Ding, H. & Chen, J. 2021. The development of diagnostic tests provides technical support for the control of African swine fever. *Vaccines*, 9(4):343. https://doi.org/10.3390/vaccines9040343.

156. Qerimi, I., Kastrati, S., Maloku, A., Gabela, O. and Maloku, E., 2023. The Importance of Theory and Scientific Theories for the Scientific Study of Genocide in the Context of the Contribution to the Development of the Science of Genocide. *Academic Journal of Interdisciplinary Studies*, *12*(1), p.183.

157. Qiu, L., Cheng, J., Gao, H., Xiong, W. and Ren, H., 2023. Federated Semi-Supervised Learning for Medical Image Segmentation via Pseudo-Label Denoising. *IEEE Journal of Biomedical and Health Informatics*.

158. Qayyum, A., Tahir, A., Butt, M.A., Luke, A., Abbas, H.T., Qadir, J., Arshad, K., Assaleh, K., Imran, M.A. and Abbasi, Q.H., 2023. Dental caries detection using a semi-supervised learning approach. *Scientific Reports*, *13*(1), p.749.

159. Railey, A., Adamson, D. & Rushton, J. 2023. Economic impact of point-of-care diagnostic testing on disease response for an African swine fever outbreak in the United States. 2023 Annual Meeting, Agricultural and Applied Economics Association, 23-25 July, Washington DC. https://ideas.repec.org/p/ags/aaea22/335566.html [8 December 2023]

160. Robeson, S.M. and Willmott, C.J., 2023. Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *Plos one*, *18*(2), p.e0279774.

161. Rubinger, L., Gazendam, A., Ekhtiari, S. and Bhandari, M., 2023. Machine learning and artificial intelligence in research and healthcare. *Injury*, *54*, pp.S69-S73.

162. Ranjit, K. 2011. *Research methodology*. 3rd ed. London: SAGE.

163. Richens, J.G., Lee, C.M. & Johri, S. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11(1):3923. https://doi.org/10.1038/s41467-020-17419-7

164. Robinson, P. 2007. *Designing and conducting mixed methods research*. 3rd ed. London: SAGE.

165. Rahman, M.M. and Kocaoglu, M., 2024. Modular Learning of Deep Causal Generative Models for High-dimensional Causal Inference. *arXiv preprint arXiv:2401.01426*.

166. Rahaman, M.H., Saha, T.K., Masroor, M., Roshani and Sajjad, H., 2024. Trend analysis and forecasting of meteorological variables in the lower Thoubal river watershed, India using non-parametrical approach and machine learning models. *Modeling Earth Systems and Environment*, *10*(1), pp.551-577.

167. Raihen, M.N. and Akter, S., 2024. Prediction modeling using deep learning for the classification of grape-type dried fruits. *International Journal of Mathematics and Computer in Engineering*.

168. Sahoh, B., Haruehansapong, K. & Kliangkhlao, M. 2022. Causal Artificial Intelligence for high-stakes decisions: The design and development of a Causal Machine Learning Model. *IEEE Access*. https://doi.org/10.1109/access.2022.3155118

169. Saka, J.O., Adesehinwa, A.O.K. & Ajala, M.K. 2010. Incidence of African swine fever (ASF) disease and its associated implications on pig production in Lagos State, Nigeria. *Bulgarian Journal of Agricultural Science*, 16(1):80-90.

170. Song, J., Wang, M., Zhou, L., Tian, P., Sun, J., Sun, Z., Guo, C., Wu, Y. and Zhang, G., 2024. A novel conserved B-cell epitope in pB602L of African swine fever virus. Applied Microbiology and Biotechnology, 108(1), pp.1-12

171. Spinard, E., Rai, A., Osei-Bonsu, J., O'Donnell, V., Ababio, P.T., Tawiah-Yingar, D., Arthur, D., Baah, D., Ramirez-Medina, E., Espinoza, N. and Valladares, A., 2023. The 2022 Outbreaks of African Swine Fever Virus Demonstrate the First Report of Genotype II in Ghana. *Viruses*, *15*(8), p.1722.

172. Sallam, M., 2023, March. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare* (Vol. 11, No. 6, p. 887). MDPI.

173. Sánchez-Vizcaíno, J.M., Laddomada, A. & Arias, M.L. 2019. African swine fever virus. In Zimmerman, J.J., Karriker, L.A. Ramirez, A. Schwartz, K.J. Stevenson, G.W. & Zhang, J. (eds). *Diseases of swine*. Hoboken NJ: Wiley. 443-452. https://doi.org/10.1002/9781119350927.ch25

174. Sanchez, P., Voisey, J.P., Xia, T., Watson, H.I., O'Neil, A.Q. & Tsaftaris, S.A. 2022. Causal machine learning for healthcare and precision medicine. *Royal Society Open Science*, 9(8):220638. http://dx.doi.org/10.1098/rsos.220638

175. Samadi, S.H., Ghobadian, B., Nosrati, M. and Rezaei, M., 2023. Investigation of factors affecting performance of a downdraft fixed bed gasifier using optimized MLP neural networks approach. *Fuel*, *333*, p.126249.

176. Simbulan, A.M., Banico, E.C., Sira, E.M.J.S., Odchimar, N.M.O. and Orosco, F.L., 2024. Immunoinformatics-guided approach for designing a pan-proteome multi-epitope subunit vaccine against African swine fever virus. *Scientific Reports*, *14*(1), p.1354.

177. Saeed, W. and Omlin, C., 2023. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*, p.110273.

178. Sarmah, B., Nair, N., Mehta, D. & Pasquali, S. 2022. Learning embedded representation of the stock correlation matrix using graph machine learning. *arXiv preprint arXiv:2207.07183*. https://doi.org/10.48550/arXiv.2207.07183

179. Schölkopf, B. 2022. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*. New York NY: Association for Computing Machinery. 765-804. http://dx.doi.org/10.1145/3501714.3501755

180. Scotland, J. 2012. Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching*, 5(9):9-16. http://dx.doi.org/10.5539/elt.v5n9p9

181. Shamji, M.H., Ollert, M., Adcock, I.M., Bennett, O., Favaro, A., Sarama, R., Riggioni, C., Annesi-Maesano, I., Custovic, A., Fontanella, S. & Traidl-Hoffmann, C. 2023. EAACI guidelines on environmental science in allergic diseases and asthma – Leveraging artificial

intelligence and machine learning to develop a causality model in exposomics. *Allergy*, 78(7):1742-1757. https://doi.org/10.1111/all.15667

182. Shukur, B.S. & Mijwil, M.M. 2023. Involving machine learning techniques in heart disease diagnosis: A performance analysis. *International Journal of Electrical and Computer Engineering*, 13(2):2177. http://doi.org/10.11591/ijece.v13i2.pp2177-2185
183. Singh, J.P. 2013. Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law*, 31(1):8-22. https://doi.org/10.1002/bsl.2052
184. Singh, H., Meyer, A.N. & Thomas, E.J. 2014. The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations. *BMJ Quality & Safety*, 23(9):727-731. https://doi.org/10.1136/bmjqs-2013-002627
185. Singh, H., Schiff, G.D., Graber, M.L., Onakpoya, I. & Thompson, M.J. 2017. The global burden of diagnostic errors in primary care. *BMJ Quality & Safety*, 26(6):484-494. http://dx.doi.org/10.1136/bmjqs-2016-005401
186. Su, Q., Liu, Q., Lau, R.I., Zhang, J., Xu, Z., Yeoh, Y.K., Leung, T.W., Tang, W., Zhang, L., Liang, J.Q. & Yau, Y.K. 2022. Faecal microbiome-based machine learning for multi-class disease diagnosis. *Nature Communications*, 13(1):6818. https://doi.org/10.1038/s41467-022-34405-3
187. Tajidini, F. & Kheiri, M.J. 2023. Recent advancement in Disease Diagnostic using machine learning: Systematic survey of decades, comparisons, and challenges. *arXiv:2308.01319*. https://doi.org/10.48550/arXiv.2308.01319.
188. Sushentsev, N., Moreira Da Silva, N., Yeung, M., Barrett, T., Sala, E., Roberts, M. and Rundo, L., 2022. Comparative performance of fully-automated and semi-automated artificial intelligence methods for the detection of clinically significant prostate cancer on MRI: a systematic review. *Insights into imaging*, *13*(1), pp.1-17.
189. Tao, D., Liu, J., Nie, X., Xu, B., Tran-Thi, T.N., Niu, L., Liu, X., Ruan, J., Lan, X., Peng, G. &
190. Tashakkori, A. & Teddlie, C. 1998. *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). Thousand Oaks, CA: SAGE.
191. Tabatabaeizadeh, S.A., 2021. Airborne transmission of COVID-19 and the role of face mask to prevent it: a systematic review and meta-analysis. *European journal of medical research*, *26*, pp.1-6.
192. Tashakkori, A. & Teddlie, C. 1998. *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). Thousand Oaks, CA: SAGE.
193. Truong, D. 2021. Using causal machine learning for predicting the risk of flight delays in air transportation. *Journal of Air Transport Management*, 91:101993. https://doi.org/10.1016/j.jairtraman.2020.101993
194. Veli, S., Arslan, A., Isgoren, M., Bingol, D. and Demiral, D., 2021. Experimental design approach to COD and color removal of landfill leachate by the electrooxidation process. *Environmental Challenges*, *5*, p.100369.
195. Yang, J., Han, S.C. and Poon, J., 2022. A survey on extraction of causal relations from natural language text. *Knowledge and Information Systems*, *64*(5), pp.1161-1186.
196. Vijh, R., Prairie, J., Otterstatter, M.C., Hu, Y., Hayden, A.S., Yau, B., Daly, P., Lysyshyn, M., McKee, G., Harding, J. and Forsting, S., 2021. Evaluation of a multisectoral intervention to mitigate the risk of severe acute respiratory coronavirus virus 2 (SARS-CoV-2) transmission in long-term care facilities. *Infection Control & Hospital Epidemiology*, *42*(10), pp.1181-1188.
197. Wallace, M. & Sheldon, N. 2015. Business research ethics: Participant observer perspectives. *Journal of Business Ethics*, 128(2):267-277. https://doi.org/10.1007/s10551-014-2102-2
198. Wang, L., Chavas, J.P. & Li, J. 2023. The dynamic impacts of disease outbreak on vertical and spatial markets: the case of African Swine Fever in China. *Applied Economics*, 55(18):2005-2023.
199. Wang, W., Zhang, Z., Tesfagaber, W., Zhang, J., Li, F., Sun, E., Tang, L., Bu, Z., Zhu, Y. and Zhao, D., 2024. Establishment of an indirect immunofluorescence assay for the detection of African swine fever virus antibodies. Journal of Integrative Agriculture, 23(1), pp.228-238.
200. Werner de Vargas, V., Schneider Aranda, J.A., dos Santos Costa, R., da Silva Pereira, P.R. and Victória Barbosa, J.L., 2023. Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. *Knowledge and Information Systems*, *65*(1), pp.31-57.

201. Wang, A.Z., Borland, D. and Gotz, D., 2024. An Empirical Study of Counterfactual Visualization to Support Visual Causal Inference. arXiv preprint arXiv:2401.08822.
202. Wyly, E., 2014. Automated (post) positivism. *Urban Geography*, *35*(5), pp.669-690.
203. Wampold, B.E. and Flückiger, C., 2023. The alliance in mental health care: Conceptualization, evidence and clinical applications. *World Psychiatry*, *22*(1), pp.25-41.
204. Wu, K., Chen, P. and Ghattas, O., 2023. A Fast and Scalable Computational Framework for Large-Scale High-Dimensional Bayesian Optimal Experimental Design. *SIAM/ASA Journal on Uncertainty Quantification*, *11*(1), pp.235-261.
205. Wei, W., Huang, C., Xia, L. and Zhang, C., 2023, April. Multi-Modal Self-Supervised Learning for Recommendation. In *Proceedings of the ACM Web Conference 2023* (pp. 790-800).
206. Wang, F., Zhang, H., Hou, L., Yang, C. and Wen, Y., 2021. Advance of African swine fever virus in recent years. *Research in veterinary science*, *136*, pp.535-539
207. Xu, D., Huang, Q., Yang, L., Chen, Y., Lu, Z., Liu, H., Han, P., Guo, L., Wang, C. and Liu, C., 2023. Experimental design of composite films with thermal management and electromagnetic shielding properties based on polyethylene glycol and MXene. *Carbon*, *202*, pp.1-12.

208. Xu, Y., Kohtz, S., Boakye, J., Gardoni, P. and Wang, P., 2023. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. *Reliability Engineering & System Safety*, *230*, p.108900.
209. Yu, X., Li, C. & Z, W. & Chen, H. 2020. A novel case adaptation method based on differential evolution algorithm for disaster emergency. *Appl. Soft Comput.*, 92:106306. https://doi.org/10.1016/j.asoc.2020.106306
210. Xu, Q., Ni, J.J., Han, B.X., Yan, S.S., Wei, X.T., Feng, G.J., Zhang, H., Zhang, L., Li, B. and Pei, Y.F., 2022. Causal relationship between gut microbiota and autoimmune diseases: a two-sample Mendelian randomization study. *Frontiers in immunology*, *12*, p.746998.
211. Yüksel, N., Börklü, H.R., Sezer, H.K. and Canyurt, O.E., 2023. Review of artificial intelligence applications in engineering design perspective. *Engineering Applications of Artificial Intelligence*, *118*, p.105697.
212. Yang, W., Li, L., Zhang, J., Wu, J., Kang, W., Wang, Y., Ding, H., Li, D. and Zheng, H., 2024. SNX32 is a host restriction factor that degrades African swine fever virus CP204L via the RAB1B-dependent autophagy pathway. *Journal of Virology*, pp.e01599-23.
213. Yang, J., Han, S.C. and Poon, J., 2022. A survey on extraction of causal relations from natural language text. *Knowledge and Information Systems*, *64*(5), pp.1161-1186.

214. Yang, Q., Martínez-López, B. & Scoglio, C. 2023. P05. 3. A novel approach to estimate the animal movement networks: Illustration for the swine industry in Iowa (US) and implications for disease prevention and control using a network-based model. In *GeoVet 2023 International Conference*.
215. Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D. and Saeed, J., 2017. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, *1*(2), pp.56-70.
216. Zheng, H., Wang, Y., Han, C., Le, F., He, R. and Lu, J., 2018, August. Learning and applying ontology for machine learning in cyber-attack detection. In 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE) (pp. 1309-1315). IEEE.
217. Zhang, C., Li, S., Zhang, M., Li, Y., Gimenez-Lirola, L.G., Li, B. and Li, W., 2023. Diagnostics and detection of African swine fever virus. *Frontiers in Veterinary Science*, *10*, p.1195138.
218. Zhang, K., Li, Z., Zhang, J., Zhao, D., Pi, Y., Shi, Y., Wang, R., Chen, P., Li, C., Chen, G. & Lei, I.M. 2022. Biodegradable smart face masks for machine learning-assisted chronic respiratory disease diagnosis. *ACS Sensors*, 7(10):3135-3143. https://doi.org/10.1021/acssensors.2c01628
219. Žukauskas, P., Vveinhardt, J. & Andriukaitienė, R. 2018. Philosophy and paradigm of scientific research. In Žukauskas, P., Vveinhardt, J. & Andriukaitienė, R. (eds). *Management culture and corporate social responsibility*. https://www.intechopen.com/chapters/58890 10 [September 2021].
220. Zuo, X., Peng, G., Xia, Y., Xu, L., Zhao, Q., Zhu, Y., Wang, C., Liu, Y., Zhao, J., Wang, H. & Zou, X. 2023. A quadruple fluorescence quantitative PCR method for the identification of

wild strains of African swine fever and gene-deficient strains. *Virology Journal*, 20(1):150. https://doi.org/10.1186%2Fs12985-023-02111-1

221. Zhang, S., Su, Q. and Chen, Q., 2021. Application of machine learning in animal disease analysis and prediction. *Current Bioinformatics*, *16*(7), pp.972-982.