Discover Food

Research

Feature extraction for artificial intelligence enabled food supply chain failure mode prediction

H. Trollman[1]

Received: 30 October 2023 / Accepted: 22 March 2024
Published online: 06 April 2024

Abstract
The Farm to Fork Strategy of the European Commission is a contingency plan aimed at always ensuring a sufficient and varied supply of safe, nutritious, affordable, and sustainable food to citizens. The learning from previous crises such as COVID-19 indicates that proactive strategies need to span numerous levels both within and external to food networks, requiring both vertical and horizontal collaborations. However, there is a lack of systematic performance management techniques for ripple effects in food supply chains that would enable the prediction of failure modes. Supervised learning algorithms are commonly used for prediction (classification) problems, but machine learning struggles with large data sets and complex phenomena. Consequently, this research proposes a manual approach to feature extraction for artificial intelligence with the aim of reducing dimensionality for more efficient algorithm performance, and improved interpretability/explainability for benefits in terms of ethics and managerial decision-making. The approach is based on qualitative comparative analysis informed by in-depth case knowledge which is refined through Boolean logic, yielding solutions that reflect complex causality as opposed to single failure point modes. Two case exemplars are presented to support the proposed framework for implementation: export readiness of dairy supply chains under the Russia-Ukraine war, and egg supply chain sustainability during COVID-19 lockdown in the United Kingdom.

Keywords Artificial intelligence · COVID-19 · Dairy 4.0 · Export potential · Feature extraction · Food supply chain · Industry 5.0 · Qualitative comparative analysis (QCA) · Ripple effect · Sustainable supply chain · Russia-Ukraine war

1 Introduction

Humanitarian catastrophes concurrent with industrial disruptions require fair allocation of limited resources to both save human lives and stabilize recovery of the industrial sector. An additional complication specific to food supply chains is perishability. Consequently, solutions for non-perishable products such as inventory holding are ineffective in many food supply chains. There may also be varying customer demand and requirements for freshness of products with penalties for decreased freshness or lack of availability [1]. Perishability is one of the intersections of sustainability with resilience in supply chains due to the generation of waste. Other sustainability issues include natural resource consumption, employment rates, etc., necessitating environmental, political, and societal perspectives of impacts. The complexity and volume of the associated data required for the related decision-making indicate a need for support from artificial intelligence. The ideal role of artificial intelligence would be to predict food supply chain failure modes well in advance of occurrence to inform preventative measures.

✉ H. Trollman, ht203@leicester.ac.uk | [1] School of Business, University of Leicester, Leicester, UK.

Food security relies on functioning food supply chains. Food supply chains continue to face numerous challenges in remaining viable largely due to the impacts of significant geo-political and socio-economic shocks such as the COVID-19 pandemic, the Russia-Ukraine war, and climate change. These and other events are also pushing a paradigm shift in consumer behavior [2, 3]. The impact on both developed and developing countries has potentially hindered the global delivery of the Sustainable Development Goals (SDGs) by 2030, including poverty reduction and zero hunger, and those related to climate change [4]. Weak links in supply chain networks can fail, resulting in the ripple/domino effect [1, 5]. Unlike the bullwhip effect which considers low-frequency-high-impact risks that are operational and recurrent, the ripple effect refers to low-frequency-high-impact disruption or exceptional risk. Consequently, there is a need for this theoretical study that focuses on ripple effects in food supply chains in the context of artificial intelligence.

Commonalities exist among some indicators of resilience for almost all industries, but others are industry-specific, highlighting the importance of focusing on critical attributes that could create ripple effects [6]. However, there is a lack of studies on extending supply chain resilience to other performance metrics, the consideration of industry-specific attributes, and elimination of bias due to the use of expert ratings of attributes [6]. Interdisciplinary collaboration employing digitalization is recommended for labor-intensive industries such as construction [7]. E-logistics using the Internet of Things (IoT) in supply chains has benefits for reduced production and logistics costs [8]. This research contributes to existing studies by considering the alternate performance metrics of export readiness and sustainability, respectively, in the context of two case studies: dairy supply chains under the Russia-Ukraine war and egg supply chains during COVID-19 lockdown in the United Kingdom.

The interconnected nature of food supply chains implies that their modelling is complex. Machine learning algorithms can provide real-time analytic insights for proactive data-driven decision-making [9]. Various machine learning algorithms have been used to analyze agricultural supply chains in terms of independent components: preproduction, production, processing, and distribution [9]. However, artificial intelligence spanning the entire supply chain to predict ripple effects is lacking in the literature, although some work has explored antecedents of supply chain viability [10–12]. Supply chain viability goes beyond resilience to also reflect the ability to continuously adapt to situations by altering internal structures [12], and aligns with panarchy theory [13] which enables different levels of meaning to interact with each other through embeddedness [14, 15]. This research contributes to panarchy theory through the lens of ecological embeddedness for sustainability and resilience.

This research builds on the strengths of qualitative comparative analysis (QCA) and advocates application together with domain-specific knowledge for the purpose of feature extraction to the case of supply chain failure mode identification when systemic shocks or disruptions are present. Bocca and Rodrigues [16] demonstrate that even simple strategies such as decomposing weather attributes and detailing fertilization justified by agronomic knowledge can improve crop yield models. The research problem is that data is either sourced from existing databanks or collected based on supposedly being related to the phenomena of interest given resource and/or time constraints, but there is a lack of evaluation of the importance of features before deciding on inclusion [16]. Wieland [13] outlines the need for transitioning supply chain management from the static and reductionist to dynamic and holistic. This transition involves levels on scales of space, time, and meaning: (1) supply chain level, (2) political-economic level, and (3) planetary level [13]. Building on panarchy theory and embeddedness, this research proposes QCA [17–19] as a tool for selecting necessary and/or sufficient features at various levels for use by artificial intelligence to predict failure modes of phenomena that have the potential to create ripple effects.

The aim of this research is to demonstrate the benefit of QCA-based feature extraction in evaluating the importance of features for inclusion, reflecting the complexity of interactions at the aforementioned levels as part of a dynamic and holistic approach to supply chain management.

The main research question is:

How can domain-specific knowledge be used to extract meaningful features for the prediction of complex food supply chain failure modes subject to stresses and shocks?

A subquestion for implementation is:

What is an appropriate framework for implementation?

This research is motivated by the need to avoid reductionist enterprise risk management (ERM) approaches which are often based on single failure points [20]. There is a lack of systematic performance management techniques for the ripple effect in supply chains [1]. Whereas reactive strategies are predominantly immediate tactical decisions relying on simulation techniques, business continuity planning, etc. [21], proactive strategies include digital solutions such as automation, localized sourcing, risk management, and employee upskilling [6]. Proactive management of supply chain disruptions includes failure prediction. To enable failure prediction, supply chain actors need to collaborate to ensure visibility across supply chain echelons to help minimize disturbances and improve robustness [22].

This research informs both academicians and policy-makers in terms of streamlined implementation of artificial intelligence for dynamic and holistic food supply chain failure mode prediction for management of ripple effects. A framework for implementation utilizing expert domain knowledge is presented alongside case studies. At the supply chain level, this work supports informed placement of Industry 4.0 technologies such as IoT sensors and the targeted collection of reliable non-sensor data, and aligns with Industry 5.0 core principles (human-centricity, sustainability, and resilience).

2 Background

All artificial intelligence is currently weak AI, meaning that a single or narrow task is performed with efficiency superior to that of a human. Weak AI applications in food industry include computer vision: image recognition and classification, object tracking and detection, and content-based image retrieval; robotics for materials handling, assembly, and quality inspections; and expert systems that can be trained on data to solve complex problems: uncovering trends and patterns to aid decision-making related to, for example, supply or production planning, preventative maintenance, and customer demand. The data to be input often requires cleaning which means removing or modifying incorrect, incomplete, irrelevant, duplicated, or improperly formatted data. Data cleaning is very important as the accuracy of the artificial intelligence model may be compromised – poor data may lead to poor insights, failure to meet objectives, increased costs, and customer dissatisfaction. Artificial intelligence may be able to aid in some, but not all, forms of data cleaning – the algorithms are unlikely to fully consider context or domain-specific knowledge. Segmentation may be used with artificial intelligence to divide data into groups based on certain characteristics. For example, customer segmentation may improve customer understanding and increase marketing effectiveness. However, if not done properly, segmentation may lead to bias and privacy issues.

From a practical perspective, food industry uses approaches such as Hazard Analysis and Critical Control Points (HACCP), Vulnerability Assessment and Critical Control Points (VACCP), and Threat Assessment and Critical Control Points (TACCP) for the purposes of food safety management. Although artificial intelligence driven HACCP, VACCP and TACCP software projects are beginning to emerge (e.g., Primority – www.primority.com), the software relies on users selecting from lists which are suggested by the system as being known or reasonably foreseen. Then responsible people are added as part of a "neural network map" based on, and limited to, process flow diagrams. Another approach to identifying appropriate solutions for problems in food industry is root cause analysis (RCA) using techniques such as the 5 Whys approach, Fishbone diagrams, and change analysis. When using RCA, it is important to realize that there often are multiple root causes. Artificial intelligence can enhance RCA by using data to detect patterns and anomalies, but models will need specific context and domain customization – effective collaboration with humans is still a requirement.

Features, also known as variables or attributes, are individual measurable properties or characteristics of a data point that may be used for analysis. Three types of data may be collected: observational data, experimental data, and synthetic data [23]. Some examples from supply chains include product type, ambient temperature, transport cost, etc. Feature selection/extraction is used to reduce the dimensionality of data by excluding features that are irrelevant or redundant. An effective choice of features reduces model overfitting (decisions are less likely to be based on noise), improves accuracy (misleading data is removed), and reduces the training time of machine learning. Feature selection may make use of domain knowledge and feature engineering whereas feature extraction may be applied to raw data without feature engineering. Feature extraction is particularly useful when data are in formats that are difficult to analyze directly / are not directly comparable. Feature engineering by experts may be time-consuming and prone to error. Deep learning algorithms do not require feature engineering, but still require careful attention to data cleaning and preprocessing. A task-specific feature extractor employing domain knowledge automated by several convolutional neural networks (deep learning) for fault diagnosis in smart manufacturing has demonstrated usefulness [24]. Gosiewska et al. [25] show that extracting information from complex models may improve the performance of simpler models for prediction.

QCA requires that each case is deconstructed into a set of "features" (termed causal conditions). Interest in QCA reflects the need for methodologies that address causality in complex systems. A key characteristic of complex systems is "emergence" [26] which means that effects may accrue from combinations of components, in contingent ways, that cannot be reduced to any one level. QCA [17, 18] combines quantitative and qualitative methodologies to address complex causation. Data on outcomes may arise from primary or secondary qualitative and/or quantitative sources. QCA enables comparison among cases whilst offering detailed understanding of the complexity of each case. Although

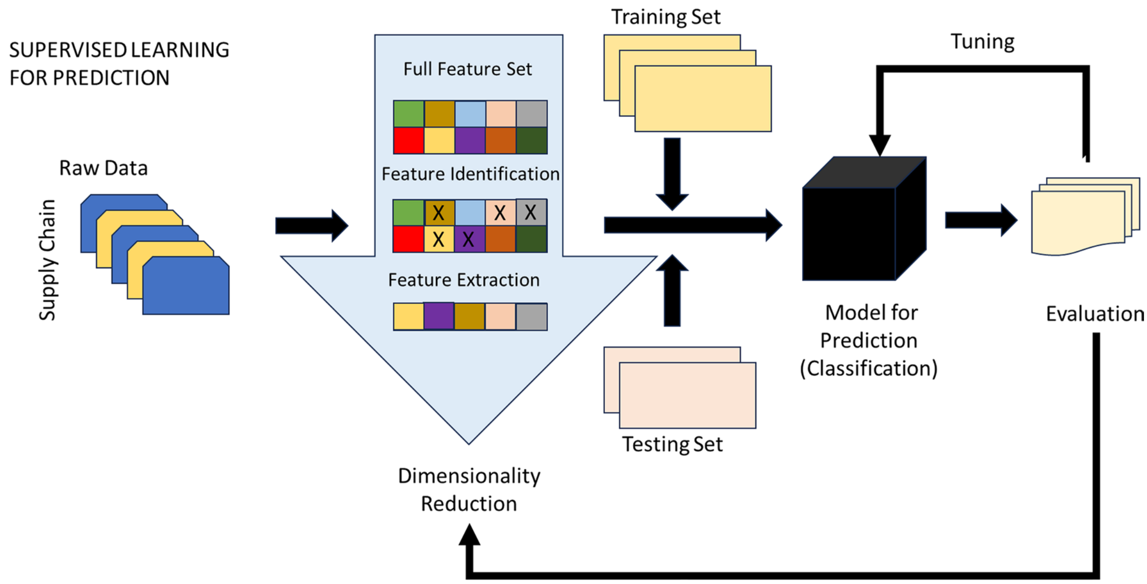

Fig. 1 Supervised learning model for prediction. See Nguyen et al. [34] for a flowchart

traditionally used for small- or medium-N, large-N samples may also be analyzed [27]. Boolean logic is employed in QCA instead of traditional correlation methods to establish causal conditions. The configurational approach of QCA is based on an analysis of necessary and sufficient causes for the outcome. Namely, a condition is necessary if it is present in all instances of the outcome, but there may be multiple sufficient causes [28]. QCA is a method that emphasizes diversity, including the identification of more than one causal pathway to an outcome (equifinality). Causal asymmetry means that explanations of success do not imply that their absence leads to failure. Consistency (how well cases with the same causal conditions align with a proposed subset relation) and coverage (how well a configuration represents instances of an outcome in the empirical cases) are used to test the results of QCA. Consequently, QCA aligns with Wieland's [13] suggestion of testing any set of levels for saliency in terms of whether the research phenomena are necessary to explain the selected levels; whether the smallest and largest levels are sufficient to delimit the research context; whether adding additional levels explains the phenomena better; and if any relevant levels are missing between the selected levels.

Theoretical information consists of background knowledge, basic modeling assumptions, processing assumptions, and cognitive conclusions obtained with appropriate strategies of analysis. Deep learning is a machine learning technique that uses artificial neural networks to mimic human learning processes. Methods used in deep learning may be either supervised, semi-supervised or unsupervised. Figure 1 illustrates a typical supervised learning model used for prediction. The aim of unsupervised learning is to analyze and cluster unlabeled datasets primarily for dimensionality reduction [23, 29]. There is a lack of consensus about how deep learning moves from training to inference. Deep neural networks, convolutional neural networks, and recurrent neural networks (adding time dimensionality) are being used, requiring considerable processing, memory, and storage requirements. Clustering groups similar observations in the same cluster. Clustering may be done under uncertainty such as probabilistic, fuzzy, possibilistic, rough, and granular [30]. Subspace clustering finds clusters within the subspaces of the original data [31]. Semi-supervised clustering uses "background knowledge" to guide otherwise unsupervised learning processes. However, human error and bias may be introduced in the collection and labeling of data for training purposes. Additionally, labelling large data is very difficult. Reducing or eliminating the need for labeling training data has thus been a focus of recent research [32]. In practical real-world scenarios, data are distributed in numerous different domains which may be approached by, for example, extracting both domain-specific and category-specific features [33].

In this work, we assume exploratory data analysis has been completed so that the data distributions, values, and their range are known, and preprocessing has been done to ensure that the data fits any models and is ready for further analysis. Data reliability would be addressed through various approaches such as deep web data extraction [35] and the information gain ratio approach [36]. Feature extraction then aims to find the most compact and informative set of features for reliable classification. Due to issues of interpretability/explainability as well as ethical and legal issues, manual feature extraction remains important in certain fields, especially where a transfer learning approach is not applicable. Explainable Artificial Intelligence (XAI) methods aim to increase understandability for the benefit of human decision-making, for example, in the case of time series data for predicting product quality in production [37]. QCA therefore contributes to XAI for understandable feature extraction in reducing the dimensionality of complex food supply chain related data.

3 Literature review

Many researchers claim that artificial intelligence has yet to be fully exploited for the full benefit of supply chains [38]. Potential applications of artificial intelligence in agri-food industry include cognitive computing for complex problems [39], sentiment analysis [40], human–machine collaboration providing, for example, improved marketing strategies that reduce waste and overproduction [41], autonomous learning systems which self-program, self-organize and self-optimize, and digital twins supported by rapid-response artificial intelligence algorithms [42]. The key enablers for the adoption of artificial intelligence in food grain supply chains in India have been explored, finding 16 significant enablers [43].

The application of artificial intelligence in supply chains has been found to reduce the bullwhip effect [44–46], enhance resilience [47], and support supply chain efficiency and responsiveness [38]. Food supply chains are better able to respond to disruptions caused by events such as the COVID-19 pandemic if their resilience is improved [48]. In general, there is scant research on the resilience and robustness of food supply chains [49]. Domain-specific knowledge on agri-food supply chains indicates that many models focus on the agricultural part of the supply chain as opposed to all the nodes of the supply chain, and do not consider decisions that should be made simultaneously such as when to plant, harvest, not serve demand, employ laborers, etc. [50].

The prediction of supply chain risks is ever more important, but especially due to recent events with global impacts such as the COVID-19 pandemic and the war in Ukraine. Failure Mode Effects Analysis (FMEA) methodology supports the importance of preventive food safety measures across the whole supply chain to address potential problems before an adverse event occurs [51]. FMEA is a method that has been applied to food supply chains supported by artificial intelligence to reduce human subjectivity [52]. However, FMEA generally does not consider combinations, just individual elements that can lead to a failure. Assessing multi-dimensional risk factors in supply chains for broken connections may help avoid a food catastrophe. In this context, five-dimensional sustainability criteria have been used to evaluate risk factors and prioritize mitigation strategies [53]. Furthermore, perishable food supply chains are subject to unique supply chain disturbances [54].

There is appreciation that qualitative and quantitative elements should be combined in a model for understanding critical challenges related to, for example, food storage management [55]. However, machine learning algorithms struggle with the classification of large and real-world datasets. Fuzzy approaches such as fuzzy clustering [56] are being used in finance including supply chain financing [57] to reduce noise in large, real-world datasets. A coexistent resilience index using fuzzy logic has been proposed to evaluate supply chain resilience of industries [6]. Feature extraction techniques such as Term Frequency – Inverse Document Frequency (TF-IDF), Bigrams, Counter Vector [58], genetic algorithm [59], and information gain ratio [36] are generally not applicable to heterogeneous data types and do not make use of domain-specific knowledge which may then lead to noise and redundancy as the extracted features may not be informative.

The Internet of Things (IoT) applied to supply chains generates vast quantities of real-time data [60] which may be combined with domain-specific data such as regulatory changes and financial data. The analysis of the large amount of IoT sensor data for real-time use in supply chain forecasting by industrial users would require techniques such as important feature extraction for the construction of lightweight deep learning techniques for resource-constrained devices and applications [61]. Incorporating domain knowledge into deep learning models for smart decision-making as well as adaptive and context-aware systems (e.g., spatial, temporal, social and environmental contexts) constitutes another important challenge. Hybrid intelligent systems are popular given their ability to solve complicated real-world problems [62].

Feature extraction techniques have been proposed for COVID-19 restrictions for supply chains [63], but these depended on two main variants of deep learning models. Work to predict possible risks in the distribution of vaccines used machine learning on only two management system cases: vaccine demand forecasting and vaccine review sentiment analysis [64]. A hybrid deep learning approach has been proposed to lessen the impact of natural disasters on shipping operations which relies on a convolutional neural network composed of a convolutional layer and a pooling layer to recognize the most significant features, however, it is noted that the logistics industry would benefit from a more nuanced approach to supply chain risks rather than that presented which relies exclusively on binary predictions [65]. The exploration of additional approaches and combinations of deep learning techniques is recommended for more reliable outcomes as well as for increasing the quantity and quality of the database, including the use of real-world case studies and external datasets.

A multi-agent system relying on artificial intelligence and edge computing has been designed as an alternative consumer-based price-making mechanism to mitigate food speculation whilst improving the sustainability of agricultural production [66]. However, there is a scarcity of case studies into the application of deep learning techniques in supply chain management [65]. Consequently, this research aims to address the gap by considering perishable food supply chains by incorporating both qualitative and quantitative (heterogeneous) sources of data to identify multi-dimensional failure modes via two case studies in a manner that is efficient and effective for large real-world datasets. Although Stone et al. [67] provide an overview of resilience factors in food supply chains, it is important to note the "no free lunch" (NFL) theory for artificial intelligence: there is no single best machine learning algorithm for predictive modeling problems. The implications of the NFL theorem to this theoretical study are that the proposed framework for feature extraction based on QCA should be adapted and tuned to specific cases.

4 Methodology

The main method used in this analysis was QCA. Specifically, crisp set QCA (csQCA) was used on two case studies to demonstrate applicability of the method to the complex issues related to two different outcomes causally related to multi-level data for the purpose of feature extraction. The two case studies were selected based on familiarity and availability of data [68–70]: dairy supply chains under the Russia-Ukraine war with the outcome of interest identified as "export readiness"; and egg supply chains during COVID-19 lockdown in the United Kingdom with the outcome of interest identified as "sustainability". The "sustainability" outcome was interpreted based on the triple bottom line [71, 72] i.e., considering updated economic, social, and environmental aspects.

There are six main stages of QCA:

1. Building the data table.
2. Constructing a 'truth table'.
3. Resolving contradictory configuration.
4. Boolean minimization.
5. Consideration of 'logical remainders' cases.
6. Interpretation.

For csQCA, there are usually 9 distinct steps [73]:

1. Specify the outcome for investigation.
2. Identify the research population and select sufficiently varying cases.
3. Determine significant conditions which may contribute to explaining the outcome.
4. Make the conditions and outcome binary.
5. Code each condition for each case to construct a data table.
6. Analyze the data table by resolving contradictions and specifying an explanatory model.
7. Produce a truth table.
8. Use Boolean minimization to generate the most parsimonious explanation.
9. Interpret the resulting explanatory models (one for the presence of the outcome, and one for its absence).

The building of the data table relied on domain-specific knowledge for each case combined with relevant literature. The truth table was constructed for representative dairy and egg supply chains, respectively, from "farm to fork" i.e., each case represents the supply chain of a different farm originating at the time of supplying a quantity/batch to the next node of the supply chain. Data represented both real-time supply chain sensor data and information from grey literature sources related to sustainability and governance, e.g., presence or absence of tariffs, panic buying, etc. Time was not specifically incorporated into the truth table.

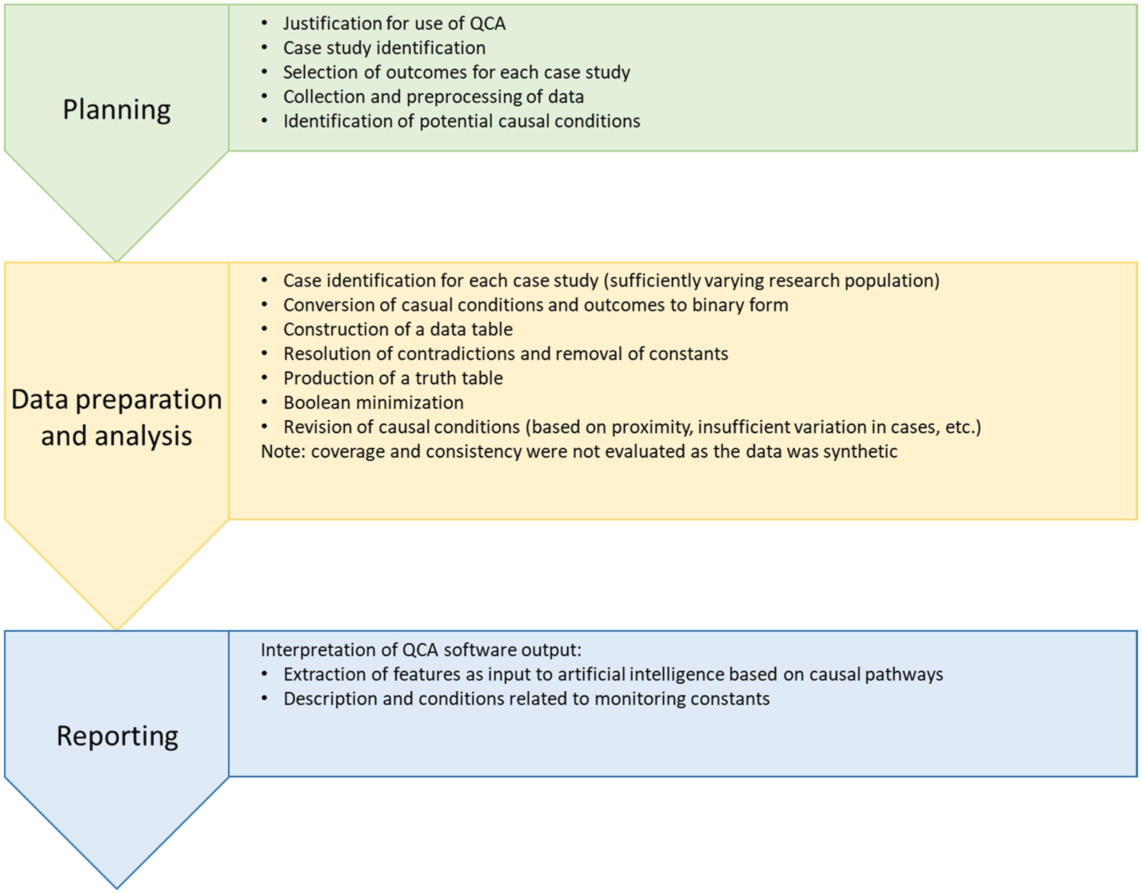

Fig. 2 Research framework for application of QCA to case studies

The remaining steps of QCA were applied as standard (steps 1–9 above). Tosmana version 1.61 software[1] was used to perform the minimizations as recommended by Rihoux and De Meur [74]. The research framework is shown in Fig. 2.

5 Case studies

5.1 Case study 1: export readiness of Ukrainian dairy products under conditions of war

The dairy industry is a food sector that is very vulnerable to climate change and other food system shocks [75]. Dairy supply chains are increasingly adopting IoT and other Industry 4.0 technologies [76, 77]. A wide array of sensor devices that can connect to the IoT are available including smart collars for cow behavior monitoring [78] and estrus events [79]. A low-cost IoT solution for milk-quality monitoring has been proposed [80] as well as IoT-based disease diagnosis in cows [81]. The various factors that affect quality such as pH, temperature, odor, turbidity, color, fat, taste, and the presence of additives can also be monitored using IoT technology to make quality-based classifications for determining optimal price [82].

The export of dairy products from Ukraine is an example of a long food supply chain with a complex food path from producer to consumer. The Ukrainian dairy supply chain is predominantly based on family farm production of less than ten cows leading to a large number of cases of raw material (milk) suppliers (farm to fork). Due to the war with Russia, domestic demand for dairy products is low because of lower disposable incomes and substantial population displacement [69]. Challenges facing the export of Ukrainian products include securing reliable partners in other countries.

[1] Cronqvist L. Tosmana – Tool for cs-, mv-, and fsQCA Version 1.6.1.0. Tosmana. University of Trier. www.tosmana.net

Table 1 Identified initial features (labels) for feature extraction using csQCA for dairy supply chains in Ukraine under conditions of war

Feature	Feature code (label)	Source (domain knowledge/reference)
Family farm	FF (1 = yes, 0 = no)	Domain knowledge
Labor/capacity shortage	LS (1 = yes, 0 = no)	[5, 67]
Adequate working conditions	WC (1 = yes, 0 = no)	[5]
Sufficient raw material available (e.g., feed for cows)	RM (1 = yes, 0 = no)	Domain knowledge
Adequate cow health	CH (1 = yes, 0 = no)	Domain knowledge
Adequate milk quality (scrap/rework)	MQ (1 = yes, 0 = no)	[67]
Located in active war zone	WZ (1 = yes, 0 = no)	Domain knowledge
Logistics access to production facility approved for EU exports (assessed in real-time)	LA (1 = yes, 0 = no)	Domain knowledge and [67]
Viable export route (logistics out of Ukraine)	ER (1 = yes, 0 = no)	Domain knowledge
Low (local) domestic demand (displacement of people)	DD (1 = yes, 0 = no)	Domain knowledge
Favorable (low or no) export duty	ED (1 = yes, 0 = no)	Domain knowledge
Favorable export demand	FD (1 = yes, 0 = no)	Domain knowledge
Adequate supply chain visibility	SV (1 = yes, 0 = no)	[10] (causal robustness)
Low network complexity	NC (1 = yes, 0 = no)	[10] (causal robustness)
High level of uncertainty	LU (1 = yes, 0 = no)	[10] (causal robustness)

Both Great Britain (from 26 April 2022) and the EU (from 4 June 2022) abolished import duties for all Ukrainian products including dairy, and simplified trans-shipment procedures to support Ukraine's economy. There are 50 dairy production facilities approved for EU exports in Ukraine. Furthermore, as Russian forces either advance or retreat, different parts of Ukraine become more or less affected by the war.

The most common failures in terms of food safety and quality at dairy processing plants, and transportation and cold storage during retail, have been identified by Aleksic et al. [51]. These may be combined with food supply chain failure modes adapted to the situation of war [68] for a comprehensive description of relevant features.

Export readiness, much like fairness in the supply chain and corporate social responsibility (CSR) failure, are outcomes that may be caused by a complex set of factors at varying degrees of impact rather than a single decisive one (although this may also be the case). The factors generally work together to yield the outcome so that predicting failure is complex. Furthermore, equifinality is likely present, meaning that different paths may yield the same outcome. To simplify the initial selection of features, thresholds may be used to assign values of either 0 or 1 to run a csQCA which can later be refined by the artificial intelligence model. Table 1 presents the initial features in which domain knowledge could be sourced from IoT sensors, access to digital records, etc. including sustainability dimensions such as working conditions (social), quality (indicative of waste for the environmental dimension), and economic indicators. The initial features may or may not be comprehensive and the extent to which this is the situation is determined by assessing the results of csQCA via coverage and consistency.

Synthetic data based on realistic cases are used due to the difficulty of accessing data on supply chains during the war. Each case (CASE ID) represents a raw milk supplier with the relevant data table shown in Table 2. The causal conditions used are as shown by the abbreviations (labels) in column 2 of Table 1 above. Note that LA should be identical to the outcome as farms without access to production facilities approved for EU exports should not have the potential to export under regulatory regimes, however, reality on the ground could vary (CASE ID 9E has a positive outcome).

5.2 Case study 2: sustainability of egg supply chains in the UK during COVID-19 lockdown

Eggs are an important part of the diet in the UK with about 197 eggs consumed per person in the pre-pandemic year 2019. COVID-19 lockdown brought about demand-induced scarcity, resulting in people having difficulty accessing a nutritious diet without resorting to unsustainable behavior [70]. Demand-induced scarcity means that eggs were available, but unable to reach retail consumers in sufficient quantity. In 2018, about one-third of eggs were sourced by food manufacturers from outside the UK (1 billion eggs annually). The low supply of eggs during COVID-19 lockdown had roots in multiple events: "Pancake Day" just before the pandemic hit (25th February 2020) depleted egg stocks; panic buying and hoarding was evident in consumer behavior; farm-gate sales of eggs increased 400% or more so that some

Table 2 Raw milk supplier binary data for labelled features and the relevant outcome (Outcome: 0 = product not exported; 1 = product exported)

CASE ID	FF	LS	WC	RM	CH	MQ	WZ	LA	ER	DD	ED	FD	SV	NC	LU	Outcome
1e	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0
2E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
3e	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0
4E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
5e	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0
6E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
7e	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0
8E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
9E	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	1
10E	1	0	1	1	0	1	0	1	1	0	1	1	0	0	1	1
11e	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0
12E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
13e	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0
14E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
15e	1	0	1	1	0	0	0	0	0	0	1	1	0	0	1	0
16E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
17e	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0
18E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
19e	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0
20E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
21e	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0
22e	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0
23e	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0
24e	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0
25e	1	0	0	0	0	1	1	0	0	1	1	1	0	0	1	0
26e	1	1	0	0	0	0	1	0	0	1	1	1	0	0	1	0
27e	1	1	0	0	0	0	1	0	0	1	1	1	0	0	1	0
28e	0	1	0	0	0	1	1	0	0	1	1	1	1	0	0	0
29E	0	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1
30E	0	0	1	0	1	1	0	1	1	0	1	1	1	0	0	1

farms installed egg vending machines thereby circumventing normal supply chains; packaging, logistics and contractual agreements caused issues with redirecting eggs from businesses (hotels, restaurants, catering industry) to retail; alternate packaging alongside consumer education were not explored; and the poultry sector had difficulty accessing raw materials including bedding, fuel, parts and feed caused by temporary closures of saw-mills and small stores as well as port disruptions in China and India (causing delays in raw feed materials shipments) [70]. Consumers sought substitutes for eggs, some of which increased environmental impact and/or decreased nutritional benefit. Some consumers sought eggs on the black/grey market, further subverting normal supply chains and undermining production, although this effect was not significantly observed during the relatively short period of the COVID-19 lockdown and neither was consumption renaturalization [70]. However, these effects could become more pronounced had the lockdown period been longer.

Table 3 presents the initial features in which domain knowledge would be sourced from grey literature, sales data, etc. including sustainability dimensions such as consumer behavior (social), waste due to hoarding (environmental dimension), and economic effect (presence of a black/grey market). The initial features may or may not be comprehensive and the extent to which this is the case would be determined by assessing the results of csQCA via coverage and consistency. The "UK farm source" (first feature in Table 3) would aggregate issues related to logistics either nationally or internationally, which could be decomposed as necessary.

Synthetic data based on realistic cases are used to represent egg suppliers at the national and international levels who form part of the egg supply network for the UK. Each case (CASE ID) represents one egg supplier with the relevant data table shown in Table 4. The "Outcome" variable is "0 = unsustainable supply chain", and "1 = sustainable supply chain" such that sustainable supply chains are those which are "coordinated through the voluntary integration of economic,

Table 3 Identified initial features (labels) for feature extraction using csQCA for egg supply chains in the UK under lockdown during COVID-19 (demand-induced scarcity)

Feature	Feature code (label)	Source (domain knowledge/reference)
UK farm source	UF (1 = yes, 0 = no)	Domain knowledge
Recent increased consumption (e.g., Pancake day)	IC (1 = yes, 0 = no)	Domain knowledge
Panic buying / hoarding	PB (1 = yes, 0 = no)	Domain knowledge
Farmgate sales	FS (1 = yes, 0 = no)	Domain knowledge
Commercial to Retail transfer possible	CT (1 = yes, 0 = no)	Domain knowledge
Alternate packaging available	AP (1 = yes, 0 = no)	Domain knowledge
Consumer education available	CE (1 = yes, 0 = no)	Domain knowledge
Raw materials available	RM (1 = yes, 0 = no)	Domain knowledge
Unsustainable consumer behavior (waste)	CW (1 = yes, 0 = no)	Domain knowledge
Unsustainable consumer behavior (nutrition)	CN (1 = yes, 0 = no)	Domain knowledge
Significant black/grey market impact	BM (1 = yes, 0 = no)	Domain knowledge
Significant consumption renaturalization	CR (1 = yes, 0 = no)	Domain knowledge
Adequate supply chain visibility	SV (1 = yes, 0 = no)	[10] (causal robustness)
Low network complexity	NC (1 = yes, 0 = no)	[10] (causal robustness)
High level of uncertainty	LU (1 = yes, 0 = no)	[10] (causal robustness)

Table 4 Egg supplier binary data for labelled features and the relevant outcome (Outcome: 0 = unsustainable supply chain; 1 = sustainable supply chain)

CASE ID	UF	IC	PB	FS	CT	AP	CE	RM	CW	CN	BM	CR	SV	NC	LU	Outcome
1S	1	1	1	0	0	0	0	0	1	0	0	0	1	1	0	1
2S	1	1	1	0	0	0	0	1	1	0	0	0	1	1	0	1
3S	1	1	1	0	0	0	0	0	1	0	0	0	1	1	0	1
4S	1	1	1	0	0	0	0	1	1	0	0	0	1	1	0	1
5s	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0
6S	1	1	1	0	0	0	0	1	0	1	0	0	1	1	0	1
7s	1	1	1	0	0	0	0	0	0	1	0	0	1	1	0	0
8S	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1
9s	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
10S	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1
11s	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
12S	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1
13s	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
14S	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1
15s	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
16S	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1
17s	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0
18S	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1
19s	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0
20S	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1
21S	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1
22S	0	0	1	0	0	0	0	1	1	0	1	0	1	0	1	1
23 s	0	0	1	0	0	0	0	0	1	1	1	0	1	0	1	0
24S	0	0	1	1	0	0	0	1	0	1	1	0	1	0	1	1
25S	0	0	1	1	0	0	0	0	0	0	0	1	1	0	1	1
26S	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1
27s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
28S	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1
29s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
30S	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1

Truth-Table:

CASEID	FF	LS	WC	RM	CH	MQ	WZ	LA	ER	DD	ED	FD	SV	NC	LU	EXPORT
30E	0	0	1	0	1	1	0	1	1	0	1	1	1	0	0	1
28e	0	1	0	0	0	1	1	0	0	1	1	1	1	0	0	0
29E	0	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1
21e, 22e, 23e, 24e	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0
25e	1	0	0	0	0	1	1	0	0	1	1	1	0	0	1	0
5e, 13e	1	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0
15e	1	0	1	1	0	0	0	0	0	0	1	1	0	0	1	0
10E	1	0	1	1	0	1	0	1	1	0	1	1	0	0	1	1
1e, 3e, 7e, 11e, 17e, 19e	1	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0
2E, 4E, 6E, 8E, 12E, 14E, 16E, 18E, 20E	1	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1
26e, 27e	1	1	0	0	0	0	1	0	0	1	1	1	0	0	1	0

Created with Tosmana Version 1.61

Fig. 3 Tosmana 1.61 QCA software truth table output following removal of case 9E due to the presence of a contradictory configuration

environmental, and social considerations with key interorganizational business systems designed to efficiently and effectively manage the material, information, and capital flows associated with the procurement, production, and distribution of products in order to meet stakeholder requirements and improve the profitability, competitiveness, and resilience over the short- and long-term" as adapted from Ahi and Searcy [83] – an outcome that would be evaluated independently of the features based on CSR data. Contradictory configurations might then indicate the correctness of this evaluation. Sustainable supply chains would be resilient (recoverability from impacts) in the presence of ripple effects. The causal conditions (labels) used are as shown by the abbreviations in column 2 of Table 3 above. Ten cases were selected to not be UK farm sources to reflect the one-third international sourcing of eggs and these did not celebrate Pancake Day.

The next section uses the data tables (Tables 2 and 4) to construct and analyze truth tables which then inform Boolean logic to identify causal pathways for feature extraction.

6 Results

6.1 Case study 1: export readiness of Ukrainian dairy products under conditions of war

From the truth table generated by the software (Tosmana 1.61), it was noted that there were seven cases representing a contradictory configuration. The reason was noted in the case study above: that 9E has export potential. This could be an error in data collection or some other cause for further investigation. Consequently, case 9E was removed from the analysis. The resulting Truth Table has no contradictory configurations (Fig. 3).

In the first run of the software (after removing 9E and performing 4 minimizations as recommended by Rihoux & De Meur [74]), MQ was found to be too proximate to WC, WZ and DD (output shown in Fig. 4 below) based on a minimization of "1" outcomes including logical remainders (non-observed cases as handled by the software). Consequently, WC, WZ and DD were removed from analysis. Minimization of 0 outcomes including remainders also indicated that LA and ER were the main features for determining export readiness or lack thereof implying the fundamental nature of logistics availability in determining export readiness (also shown in Fig. 4). LA and ER were therefore combined into one new feature (VL) representing Viable Logistics. This yielded a new truth table, but VL was found to be too granular (the only causal condition leading to the outcome due to low level of detail). VL was therefore identified as a feature for input to artificial intelligence and then removed from the truth table to identify the next level of features as shown in Fig. 5.

Minimization of the 0 outcome (not export ready) including logical remainders led to an elegant solution of interest for identifying failure modes:

$$MQ(0) + RM(0)CH(0) \rightarrow EXPORT(0) \tag{1}$$

Fig. 4 Software output showing "proximate" result for WC, WZ, and DD. Also, LA and ER identified as main features for "1" outcome – not of particular interest for failure mode identification

Result(s):

LA{1}
(2E,4E,6E,8E,12E,14E,16E,18E,20E+10E+29E+30E)

ER{1}
(2E,4E,6E,8E,12E,14E,16E,18E,20E+10E+29E+30E)

WC{1}MQ{1}
(2E,4E,6E,8E,12E,14E,16E,18E,20E+10E+29E+30E)

MQ{1}WZ{0}
(2E,4E,6E,8E,12E,14E,16E,18E,20E+10E+29E+30E)

MQ{1}DD{0}
(2E,4E,6E,8E,12E,14E,16E,18E,20E+10E+29E+30E)

Fig. 5 The final truth table for export readiness of Ukrainian dairy products under conditions of war

Truth-Table:

CASEID	FF	LS	RM	CH	MQ	ED	FD	SV	NC	LU	EXPORT
30E	0	0	0	1	1	1	1	1	0	0	1
28e	0	1	0	0	1	1	1	1	0	0	0
29E	0	1	1	1	1	1	1	1	0	0	1
21e, 22e, 23e, 24e	1	0	0	0	0	1	1	0	0	1	0
25e	1	0	0	0	1	1	1	0	0	1	0
5e, 13e	1	0	0	1	0	1	1	0	0	1	0
15e	1	0	1	0	0	1	1	0	0	1	0
10E	1	0	1	0	1	1	1	0	0	1	1
1e, 3e, 7e, 11e, 17e, 19e	1	0	1	1	0	1	1	0	0	1	0
2E, 4E, 6E, 8E, 12E, 14E, 16E, 18E, 20E	1	0	1	1	1	1	1	0	0	1	1
26e, 27e	1	1	0	0	0	1	1	0	0	1	0

Created with Tosmana Version 1.61

Therefore, in addition to VL (viable logistics) as a feature, there are two causal pathways that could lead to a lack of export readiness: lack of MQ (milk quality) or ("+" indicates "or" in this notation) lack of RM (raw material) and (logical "and") CH (cow health). Consequently, four features have been identified for input into artificial intelligence as opposed to the original 15.

Note that ED, FD, and NC are constants (must be present) and would need to be monitored for change, but do not need to be used to train the artificial intelligence.

6.2 Case study 2: sustainability of egg supply chains in the UK during COVID-19 lockdown

The truth table generated by the software contained no contradictions as shown in Fig. 6. The table shows that CT, AP, and CE are constants and so should be monitored for change, but do not need to be used to train the artificial intelligence.

The results for minimizing 0 outcomes with remainders indicate multiple causal pathways as shown in Fig. 7. In this situation, there are no clear proximity issues. However, it is known that BM and CR were not issues in the UK, so for lockdown situations of comparable or lesser length, these features should be monitored but not used to train the artificial intelligence until that length of time has elapsed. Consequently, these two features (BM and CR) are eliminated from further analysis under the assumption of short lockdown period.

By not including BM and CR, a more parsimonious result is obtained as shown in Eq. 2.

Fig. 6 Tosmana 1.61 QCA software truth table output for egg suppliers

Truth-Table:

CASE ID	UF	IC	PB	FS	CT	AP	CE	RM	CW	CN	BM	CR	SV	NC	LU	Outcome
27s, 29s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
26S, 28S, 30S	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1
21S	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1
23s	0	0	1	0	0	0	0	0	1	1	1	0	1	0	1	0
22S	0	0	1	0	0	0	0	1	1	0	1	0	1	0	1	1
25S	0	0	1	1	0	0	0	0	0	0	0	1	1	0	1	1
24S	0	0	1	1	0	0	0	1	0	1	1	0	1	0	1	1
9s, 11s, 13s, 15s	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
7s	1	1	1	0	0	0	0	0	0	1	0	0	1	1	0	0
1S, 3S	1	1	1	0	0	0	0	0	1	0	0	0	1	1	0	1
5s	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0
8S, 10S, 12S, 14S	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1
6S	1	1	1	0	0	0	0	1	0	1	0	0	1	1	0	1
2S, 4S	1	1	1	0	0	0	0	1	1	0	0	0	1	1	0	1
17s, 19s	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0
16S, 18S, 20S	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1

Created with Tosmana Version 1.61

Fig. 7 Software output showing multiple causal pathways for egg supply chains to be unsustainable

Result(s):

RM{0}CN{1} + RM{0}SV{0}
(5s+7s+23s) (9s,11s,13s,15s+17s,19s+27s,29s)

RM{0}CN{1} + RM{0}CW{0}CR{0}
(5s+7s+23s) (7s+9s,11s,13s,15s+17s,19s+27s,29s)

RM{0}CN{1} + RM{0}BM{0}CR{0}NC{0}
(5s+7s+23s) (9s,11s,13s,15s+17s,19s+27s,29s)

RM{0}CN{1} + RM{0}BM{0}CR{0}LU{1}
(5s+7s+23s) (9s,11s,13s,15s+17s,19s+27s,29s)

CW{1}CN{1} + RM{0}CW{0}CR{0}
(5s+23s) (7s+9s,11s,13s,15s+17s,19s+27s,29s)

$$\mathrm{RM(0)CN(1)} + \mathrm{RM(0)SV(0)} \rightarrow \mathrm{SUSTAIN(0)} \tag{2}$$

The result is interpreted as two causal pathways: lack of RM (Raw Material) and presence of CN (Unsustainable Consumer Behavior (nutrition)) or lack of RM and lack of adequate supply chain visibility (SV) both lead to an outcome of having an unsustainable supply chain. Consequently, the features that should be input into artificial intelligence include those of the pathways: RM, CN, and SV i.e., three of the original 15 features with BM (black market) and CR (consumption renaturalization) to be added for longer lockdown periods.

7 Discussion

Ongoing developments in engineering technology such as Industry 4.0 offer opportunities for managing the ripple effect, but also create new challenges for its analysis. For example, disruptions to information systems and networked cloud-based digital supply chain environments need to be avoided for the effective deployment of such technological solutions which may prove problematic under conditions of stresses and shocks to food systems. The identification of features for use by artificial intelligence in predicting food supply chain failure modes is highly dependent on quality input

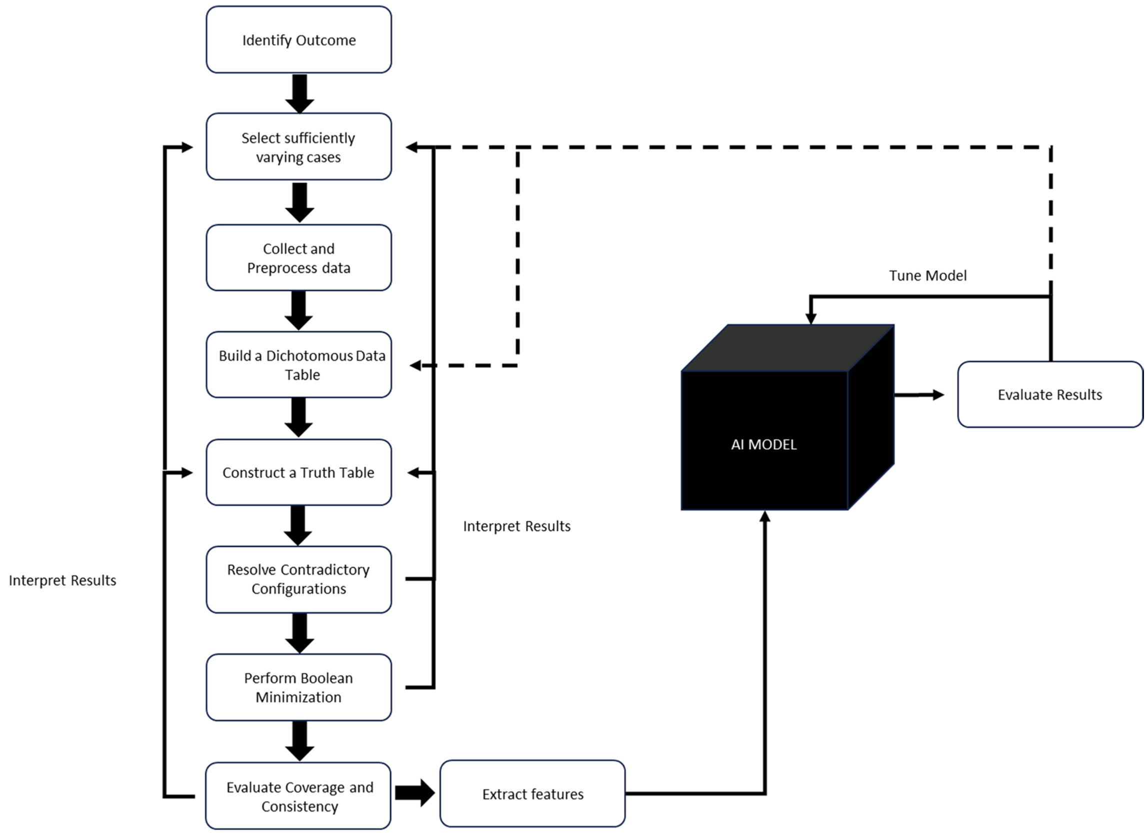

Fig. 8 Framework for implementation of csQCA-based feature extraction

data. Both automatic tools and domain experts may be utilized for feature extraction from available data, but invariably, quality models are based on the right features, requiring knowledge of the domain and use cases, e.g., Mende et al. [37] found that manually extracted features not only result in better prediction but also better explain physical parameters.

This research presents a feasible approach to addressing the problem of feature extraction for use by artificial intelligence algorithms in predicting complex food supply chain failure modes subject to stresses and shocks using domain-specific knowledge. The proposed solution is to apply csQCA following the framework for implementation shown in Fig. 8 as derived from the case studies. The main research question is addressed by the framework in that domain-specific knowledge is applied in identification of the outcome, selection of the initial cases, resolution of contradictory cases, and interpretation of the results of Boolean minimization, with the latter two feeding back to case selection and truth table construction. The research subquestion is reflected by the remainder of the framework which incorporates the artificial intelligence element feeding back into the selection of cases and the dichotomous data table as needed. This part of the framework is run only if the extraction of features was successful, meaning that consistency and coverage were found to be within acceptable parameters. If this is not the case, feedback loops to case selection and Truth Table construction are used to remedy the relevant shortcomings.

However, for this research, the results of the cases have not been evaluated. Consistency and coverage are the two main measurements used to assess QCA outputs [84]. Consistency is measured by adding up the number of cases where both conditions and outcome are present divided by the sum of cases where the outcome is present. Coverage indicates the percentage of the sum of membership scores in the outcome condition that can be explained by the causal pathways. As both case studies are synthetic combinations of real-world data, these analyses are not performed, however, they should be in the case of actual data to further inform the suitability of the features selected using the feedback loop in the proposed framework (Fig. 8). QCA has various techniques that may be applied to resolve consistency and coverage deficiencies such as adding features, etc. [74].

In addition, the supply chain cases used are deliberately low in granularity. Each of the selected features represents a synthesis of various data sources (quantitative and qualitative) and could be broken down further. Machine learning usually follows the "10 times rule" meaning that the amount of input data should be ten times more that the number of degrees of freedom (features) of the model: a dataset with 10 columns should have at least 100 rows. For the case of reliable wireless broadcast, Nguyen et al. [34] employ historical feedback data and six extracted communication link features (to address the correlation of some features), showing that the accuracy of their classifier increases proportionately to the training dataset size. However, retaining all features is not feasible for complex problems, nor is it possible to experimentally vary features. Consequently, the use of a flexible and adaptable feature extraction approach such as QCA is desirable.

The use of specifically csQCA is another simplification for the purposes of modelling. The membership of each set is binary whereas ideally fuzzy-set QCA (fsQCA) could be employed to assign membership values based on the degree of membership rather than categorical memberships. For example, the degree to which labor force is available, product quality, and consumer satisfaction would vary from 0 to 1. However, rather than using potentially error-prone expert assessments of membership, these determinations are better left to the artificial intelligence algorithm and then fed back into the feature extraction. Such feature extraction updating could be based on real-time data as opposed to historical data. An example of the use of fsQCA outside of supply chain is bankruptcy prediction which uses fsQCA and convolutional neural networks (a hybrid model) [85], but this work examined various feature selection techniques as opposed to feature extraction.

An interesting outcome from the synthetic data is that global features such as network complexity and uncertainty which were motivated by literature [10] are not as important as case specific ones. This is promising for the relative importance of the incorporation of IoT technologies into food supply chains, including that such applications are the focus of considerable research (e.g., [76–82]). However, some of the more granular features such as Viable Logistics (VL) should be decomposed into critical nodes, and in relevant cases, intra- and inter-organizational variables [10].

Finally, the importance of examining contradictory configurations needs to be emphasized for the purposes of implementation. The application of QCA must be robust in terms of case familiarity, and any contradictions should be logically resolved as opposed to simply fed into the artificial intelligence algorithm which could then lead to garbage in, garbage out scenarios.

8 Conclusion

The present work examined food supply chain failure modes under systemic stresses and shocks by presenting a solution in the form of csQCA for feature extraction which addresses numerous challenges including:

(i) Food supply chain complexity and interconnectedness including perishability and diversity of data types. QCA is able to handle both quantitative and qualitative data at various levels with applicability to various outcomes.
(ii) Extending supply chain resilience to other performance metrics such as export readiness and sustainability which consider industry-specific attributes whilst minimizing bias due to expert ratings as they are evaluated by Boolean logic and feedback mechanisms. CSR failure and (un)fairness in supply chains would be interesting outcomes to investigate in terms of performance as future work.
(iii) Informing the lack of systematic performance management techniques for ripple effect in supply chains by providing a proactive approach to predicting failure modes. QCA incorporates both commonalities among industries as well as industry-specific indicators enabling cautious generalization as well as specific managerial solutions. These solutions are based on complex causality as opposed to the evaluation of single failure points.
(iv) The lack of evaluation of feature importance prior to use by artificial intelligence algorithms. Most feature extraction is automated whereas the QCA approach requires domain-specific knowledge to inform the initial selection of features.
(v) Interpretability/explainability and ethical/legal issues in terms of the benefits that manual feature extraction brings in addition to the usual reduction of dimensionality thereby improving algorithm efficiency and performance.

8.1 Theoretical contribution

The research presented builds on panarchy theory and embeddedness to provide a dynamic and holistic approach to complex ripple effect phenomena in food supply chains. The main theoretical outcome of this research is to demonstrate how complex social-ecological systems that are the subject of panarchy theory can be captured by QCA as part of a cross-scale structure, and dynamically interpreted to identify failure modes with artificial intelligence. The multiple scales align with emergence as emergent phenomena are generally manifested at upper levels arising from lower levels.

8.2 Practical contribution

The real-world applications of this research potentially extend beyond supply chains to any dynamic multi-level social-ecological system prone to domino/ripple effects. The overall aim is to improve system resilience. The importance of identifying failure modes in time is to avoid situations in which even a small disturbance can lead to collapse because there is a lack of capacity to buffer or adapt to shock. Examples of such real-world applications include the emergence of novel diseases, city-scale food systems, and soil systems. In these contexts, the ability to integrate both qualitative and quantitative data sources is valuable and may enable the imputation of missing data (a common problem) from, for example, social media and satellites.

QCA may be considered a manual approach to feature extraction for the initial evaluation of levels (supply chain level, political-economic level, planetary level) which may subsequently be updated by artificial intelligence as part of a mostly automated feedback-informed process for specific cases. At the level of supply chain nodes, such an approach could be used to inform the location of IoT sensors. At the other levels, domain-specific knowledge would enhance the collection of non-sensor data. Overall, the synergistic application of expert domain knowledge combined with the strengths of Boolean logic / artificial intelligence reflects the ambitions of Industry 5.0.

Author contributions H.T. sole authorship.

Funding This research was supported by a Research England grant administered by Universities UK International (project reference number 11155).

Declarations

Competing interests The Author declares being a Guest Editor on the Special Issue. The Author declares having published work with co-Guest Editor Dr. Sandeep Jagtap. The Author has no other conflicts of interest or competing interests. The author confirms that all data generated or analysed during this study are included in this published article.

References

1. Dolgui A, Ivanov D, Sokolov B. Ripple effect in the supply chain: an analysis and recent literature. Int J Prod Res. 2018;56(1–2):414–30. https://doi.org/10.1080/00207543.2017.1387680.
2. Mehta S, Saxena T, Purohit N. The new consumer behaviour paradigm amid COVID-19: permanent or transient? J Health Manag. 2020;22(2):291–301. https://doi.org/10.1177/0972063420940834.
3. Bazzoni E, Jacob M, Land S, Rupp M, Welcherng S. 9 charts that show how inflation and the Ukraine war are impacting European consumers. World Economic Forum. 2022. weforum.org/agenda/2022/06/how-inflation-and-the-conflict-in-ukraine-are-impacting-european-consumers/. Accessed 6 Oct. 2023.
4. Fenner R, Cernev T. The implications of the COVID-19 pandemic for delivering the sustainable development goals. Futures. 2021;128:12726. https://doi.org/10.1016/j.futures.2021.102726.
5. Aldrighetti R, Battini D, Ivanov D, Zennaro I. Costs of resilience and disruptions in supply chain network design models: a review and future research directions. Int J Prod Econ. 2021;235: 108103. https://doi.org/10.1016/j.ijpe.2021.108103.
6. Narassima MS, Gedam V, Gunasekaran A, Anbuudayasankar SP, Dwarakanath M. A novel coexistent resilience index to evaluate the supply chain resilience of industries using fuzzy logic. Supply Chain Manag. 2023. https://doi.org/10.1108/SCM-05-2023-0218.
7. Liao L, Yang C, Quan L. Construction supply chain management: a systematic literature review and future development. J Clean Prod. 2023;382: 135230. https://doi.org/10.1016/j.jclepro.2022.135230.
8. Ashima R, Haleem A, Bahl S, Nandan D, Javaid M. Automation of AM via IoT towards implementation of e-logistics in supply chain for industry 4.0. In: Pradhan P, Pattanayak B, Das HC, Mahanta P, editors. Recent Advances in Mechanical Engineering: Lecture Notes in Mechanical Engineering. Springer: Singapore; 2023. p. 181–9 (**10.1007/978-981-16-9057-0_20**).

9. Sharma R, Kamble SS, Gunasekaran A, Kumar V, Kumar A. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Comput Oper Res. 2020;119: 104926. https://doi.org/10.1016/j.cor.2020.104926.
10. Durach CF, Wieland A, Machuca JAD. Antecedents and dimensions of supply chain robustness: a systematic literature review. Int J Phys Distrib Logist Manag. 2015;45(1/2):118–37. https://doi.org/10.1108/IJPDLM-05-2013-0133.
11. Kumari S, Jeble S, Venkatesh VG, Nagarajan C, Shi Y. Antecedents of agriculture supply chain performance during COVID-19: an emerging economy perspective. Oper Manag Res. 2023;16:489–510. https://doi.org/10.1007/s12063-022-00295-3.
12. Misbauddin SM, Alam MJ, Karmaker CL, Nabi MNU, Hasan MM. Exploring the antecedents of supply chain viability in a pandemic context: an empirical study on the commercial flower supply chain of an emerging economy. Sustainability. 2023;15(3):2146. https://doi.org/10.3390/su15032146.
13. Wieland A. Dancing the supply chain: toward transformative supply chain management. J Supply Chain Manag. 2021;57:58–73. https://doi.org/10.1111/jscm.12248.
14. Granovetter M. Economic action and social structure: the problem of embeddedness. Am J Sociol. 1985;91:481–510. https://doi.org/10.1086/228311.
15. Trollman H, Garcia-Garcia G, Jagtap S, Trollman F. Blockchain for ecologically embedded coffee supply chains. Logist. 2022;6(3):43. https://doi.org/10.3390/logistics6030043.
16. Bocca FF, Rodrigues LHA. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. Comput Electron Agric. 2016;128:67–76. https://doi.org/10.1016/j.compag.2016.08.015.
17. Ragin C. The comparative method. Berkeley: University of California Press; 1987.
18. Ragin C. Fuzzy-set social science. Chicago: University of Chicago Press; 2000.
19. Roig-Tierno N, Gonzalez-Cruz TF, Llopis-Martinez J. An overview of qualitative comparative analysis: A bibliometric analysis. J Innov Knowl. 2017;2(1):15–23. https://doi.org/10.1016/j.jik.2016.12.002.
20. Kowalska A, Lingham S, Maye D, Manning L. Food insecurity: is leagility a potential remedy? Foods. 2023;12(16):3138. https://doi.org/10.3390/foods12163138.
21. Belhadi A, Kamble S, Jabbour CJC, Gunasekaran A, Ndubisi NO, Venkatesh M. Manufacturing and service supply chain resilience to the COVID-19 outbreak: lessons learned from the automobile and airline industries. Technol Forecast Soc Change. 2021;163: 120447. https://doi.org/10.1016/j.techfore.2020.120447.
22. Wieland A, Marcus WC. Dealing with supply chain risks: linking risk management practices and strategies to performance. Int J Phys Distrib Logist Manag. 2012;42(10):887–905. https://doi.org/10.1108/09600031211281411.
23. De Bock KW, Coussement K, De Caigny A, Słowiński R, Baesens B, Boute RN, Choi T-M, Delen D, Kraus M, Lessmann S, Maldonado S, Martens D, Óskarsdóttir M, Vairetti C, Verbeke W, Weber R. Explainable AI for Operational Research: A defining framework, methods, applications, and a research agenda. Eur J Oper Res. 2023. https://doi.org/10.1016/j.ejor.2023.09.026.
24. Feng J, Yao Y, Lu S, Liu Y. Domain Knowledge-Based Deep-Broad Learning Framework for Fault Diagnosis. In: IEEE Transactions on Industrial Electronics. 2021;68(4): 3454–3464. https://doi.org/10.1109/TIE.2020.2982085
25. Gosiewska A, Kozak A, Biecek P. Simpler is better: Lifting interpetability-performace trade-off via automated feature engineering. Decis Support Syst. 2021;150: 113556. https://doi.org/10.1016/j.dss.2021.113556.
26. Byrne D. Evaluating complex social interventions in a complex world. Eval. 2013;19(3):217–28. https://doi.org/10.1177/1356389013495617.
27. Emmenegger P, Schraff D, Walter A. QCA, the truth table analysis and large-N survey data: The benefits of calibration and the importance of robustness tests. (Compasss Working Paper 2014–79) compasss.org/wpseries/EmmeneggerSchraffWalter2014.pdf. Accessed 13 Oct 2023.
28. Ragin C. Redesigning social inquiry: Fuzzy sets and beyond. Chicago: University of Chicago Press; 2008.
29. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015;349(6245):255–60. https://doi.org/10.1126/science.aaa8415.
30. D'Urso P. Informational paradigm, management of uncertainty and theoretical formalisms in the clustering framework: a review. Inf Sci. 2017;400–401:30–62. https://doi.org/10.1016/j.ins.2017.03.001.
31. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data. Data Min Knowl Discov. 2005;11(1):5–33. https://doi.org/10.1007/s10618-005-1396-1.
32. Ryu J, Yang M-H, Lim J. Unsupervised feature learning for self-tuning neural networks. Neural Netw. 2021;133:103–11. https://doi.org/10.1016/j.neunet.2020.10.011.
33. Peng X, Li Y, Saenko K. Domain2Vec: Domain Embedding for Unsupervised Domain Adaptation. In: European Conference on Computer Vision (ECCV) 2020 Proceedings; 2020. pp. 756–774. https://doi.org/10.1007/978-3-030-58539-6_45
34. Nguyen D, Nguyen C, Duong-Ba T, Nguyen H, Nguyen A, Tran T. Joint network coding and machine learning for error-prone wireless broadcast. In: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2017. pp. 1–7.
35. Banu BA, Chitra M. Deep Web Data Extraction Based on URL and Domain Classification. ISACA J. 2015;4. https://www.isaca.org/resources/isaca-journal/issues/2015/volume-4/deep-web-data-extraction-based-on-url-and-domain-classification. Accessed 16 Oct 2023.
36. Alsanad A. Hoeffding Tree Method with Feature Selection for Forecasting Daily Demand Orders. In: 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taipei, Taiwan. 2020: 223–227. https://doi.org/10.1109/TAAI51410.2020.00048. Accessed 24 Oct 2023.
37. Mende H, Frye M, Vogel P-A, Kiroriwal S, Schmitt RH, Bergs T. On the importance of domain expertise in feature engineering for predictive product quality in production. Proc CIRP. 2023;118:1096–101. https://doi.org/10.1016/j.procir.2023.06.188.
38. Younis H, Sundarakani B, Alsharairi M. Applications of artificial intelligence and machine learning within supply chains: systematic review and future research directions. J Model Manag. 2022;17(3):916–40. https://doi.org/10.1108/JM2-12-2020-0322.
39. Adel A. Future of industry 5.0 in society: human-centric solutions, challenges and prospective research areas. J Cloud Comput. 2022. https://doi.org/10.1186/s13677-022-00314-5.

40. Espina-Romero L, Guerrero-Alcedo J, Goñi Avila N, Noroño Sánchez JG, Gutiérrez Hurtado H, Quiñones LA. Industry 5.0: tracking scientific activity on the most influential industries, associated topics, and future research agenda. Sustainability. 2023;15(6):5554. https://doi.org/10.3390/su15065554.
41. Alojaiman B. Technological modernizations in the industry 5.0 era: a descriptive analysis and future research directions. Processes. 2023;11(5):1318. https://doi.org/10.3390/pr11051318.
42. Leng J, Sha W, Wang B, Zheng P, Zhuang C, Liu Q, Wuest T, Mourtzis D, Wang L. Industry 5.0: prospect and retrospect. J Manuf Syst. 2022;65:279–95. https://doi.org/10.1016/j.jmsy.2022.09.017.
43. Das S, Barve A, Sahu NC, Muduli K. Enabling artificial intelligence for sustainable food grain supply chains: an agri 5.0 and circular economy perspective. Oper Manag Res. 2023;16:2104–24. https://doi.org/10.1007/s12063-023-00390-z.
44. Beheshtinia MA, Salmabadi N, Rahimi S. A robust possibilistic programming model for production-routing problem in a three-echelon supply chain. J Model Manag. 2021. https://doi.org/10.1108/JM2-06-2020-0162.
45. Gupta S, Saxena A. Operations-based classification of the bullwhip effect. J Model Manag. 2020;17(1):134–53. https://doi.org/10.1108/JM2-01-2020-0029.
46. Singh LP, Challa RT. Integrated forecasting using the discrete wavelet theory and artificial intelligence techniques to reduce the bullwhip effect in a supply chain. Glob J Flex Syst Manag. 2016;17(2):157–69. https://doi.org/10.1007/s40171-015-0115-z.
47. Akbari M, Do TNA. A systematic review of machine learning in logistics and supply chain management: current trends and future directions. Benchmarking Int J. 2021. https://doi.org/10.1108/BIJ-10-2020-0514.
48. Marusak A, Sadeghiamirshahidi N, Krejci CC, Mittal A, Beckwith S, Cantu J, Morris M, Grimm J. Resilient regional food supply chains and rethinking the way forward: key takeaways from the COVID-19 pandemic. Agric Syst. 2021;190: 103101. https://doi.org/10.1016/j.agsy.2021.103101.
49. Gurrala K, Hariga M. Key food supply chain challenges: a review of the literature and research gaps. Oper Supply Chain Manag Int J. 2022;15(4):441–60. https://doi.org/10.31387/oscm0510358.
50. Esteso A, Alemany MME, Ortiz A. Sustainable agri-food supply chain planning through multi-objective optimisation. J Decis Syst. 2023. https://doi.org/10.1080/12460125.2023.2180138.
51. Aleksic B, Djekic I, Miocinovic J, Miloradovic Z, Memisi N, Smigic N. The application of failure mode effects analysis in the long supply chain—a case study of ultra filtrated milk cheese. Food Control. 2022;138: 109057. https://doi.org/10.1016/j.foodcont.2022.109057.
52. Peddi S, Lanka K, Gopal PRC. Modified FMEA using machine learning for food supply chain. Mater Today Proc. 2023. https://doi.org/10.1016/j.matpr.2023.04.353.
53. Das S, Myla AY, Barve A, Kumar A, Sahu NC, Muduli K, Luthra S. A systematic assessment of multi-dimensional risk factors for sustainable development in food grain supply chains: a business strategic prospective analysis. Bus Strateg Environ. 2023;32(8):5536–62. https://doi.org/10.1002/bse.3435.
54. Sharma H, Shanker S, Barve A, Muduli K, Kumar A, Luthra S. Interval-valued intuitionistic fuzzy digraph-matrix approach with PERMAN algorithm for measuring COVID-19 impact on perishable food supply chain. Environ Dev Sustain. 2022. https://doi.org/10.1007/s10668-022-02487-0.
55. Das S, Barve A, Sahu NC, Muduli K, Kumar A, Luthra S. Analysing the challenges to sustainable food grain storage management: a path to food security in emerging nations. Int J Food Sci Technol. 2023;58:5501–9. https://doi.org/10.1111/ijfs.16437.
56. Almasi ON, Rouhani M. Fast and de-noise support vector machine training method based on fuzzy clustering method for large real world datasets. Turk J Elec Eng Comp Sci. 2016;24(1):219–33. https://doi.org/10.3906/elk-1304-139.
57. Liu Y, Huang L. Supply chain finance credit risk assessment using support vector machine-based ensemble improved with noise elimination. Int J Distrib Sens. 2020. https://doi.org/10.1177/1550147720903631.
58. Jassim MA, Abd DH, Omri MN. Machine learning-based new approach to films review. Soc Netw Anal Min. 2023;13:40. https://doi.org/10.1007/s13278-023-01042-7.
59. Khezripour H, Mozaffari SP, Reshadi M, Zarrabi H. Classification of electrocardiogram signals using deep learning based on genetic algorithm feature extraction. Biomed Phys Eng Express. 2023;9(5): 055014. https://doi.org/10.1088/2057-1976/acdc2a.
60. Jagtap S, Duong L, Trollman H, Bader F, Garcia-Garcia G, Skouteris G, Li J, Pathare P, Martindale W, Swainson M, Rahimifard S. Chapter 5—IoT technologies in the food supply chain. In: Galanakis CM, editor. Food technology disruptions. Academic Press; 2021. p. 175–211 (**10.1016/B978-0-12-821470-1.00009-4**).
61. Sarker IH. Deep Learning: A comprehensive overview on techniques, taxonomy, applications and research directions. Sn Comput Sci 221;2:420. Doi:https://doi.org/10.1007/s42979-021-00815-1
62. Olatunji SO, Selamat A, Abdulraheem A. A hybrid model through the fusion of type-2 fuzzy logic systems and extreme learning machines for modelling permeability prediction. Inf Fusion. 2014;16:29–45. https://doi.org/10.1016/j.inffus.2012.06.001.
63. Bassiouni MM, Chakrabortty RK, Hussain OK, Rahman HF. Advanced deep learning approaches to predict supply chain risks under COVID-19 restrictions. Expert Syst Appl. 2023;211: 118604. https://doi.org/10.1016/j.eswa.2022.118604.
64. Hu H, Xu J, Liu M, Lim MK. Vaccine supply chain management: an intelligent system utilizing blockchain, IoT and machine learning. J Bus Res. 2023;156: 113480. https://doi.org/10.1016/j.busres.2022.113480.
65. Alzahrani A, Asghar MZ. Intelligent risk prediction system in IoT-based supply chain management in logistics sector. Electronics. 2023;12(13):2760. https://doi.org/10.3390/electronics12132760.
66. Trollman H, Jagtap S, Trollman F. Crowdsourcing food security: introducing food choice derivatives for sustainability. Food Secur. 2023;15:953–65. https://doi.org/10.1007/s12571-023-01363-7.
67. Stone J, Rahimifard S, Woolley E. An overview of resilience factors in food supply chains. 11th Biennial Conference of the European Society for Ecological Economics. 2015. https://core.ac.uk/download/pdf/288374238.pdf. Accessed 25 Oct 2023.
68. Jagtap S, Trollman H, Tollman F, Garcia-Garcia G, Parra-López C, Duong L, Martindale W, Munekata PES, Lorenzo JM, Hdaifeh A, Hassoun A, Salonitis K, Afy-Aharah M. The Russia-Ukraine conflict: Its implications for the global food supply chains. Foods. 2022;11(14):2098. https://doi.org/10.3390/foods11142098.
69. Lou C, Gorobec R, Samoilyk I, Trollman H. Traceable dairy supply chain implementation in ukraine for improved export potential. Eng Proc. 2023;40(1):14. https://doi.org/10.3390/engproc2023040014.

70. Trollman H, Jagtap S, Garcia-Garcia G, Harastani R, Colwill J, Trollman F. COVID-19 demand-induced scarcity effects on nutrition and environment: investigating mitigation strategies for eggs and wheat flour in the United Kingdom. Sustain Prod Consum. 2021;27:1255–72. https://doi.org/10.1016/j.spc.2021.03.001.
71. Elkington J. Cannibals with Forks: the Triple Bottom Line of 21st Century Business. Gabriola Island, BC; Stony Creek: New Society Publishers; 1998.
72. Elkington J. 25 Years Ago I Coined the Phrase "Triple Bottom Line." Here's Why It's Time to Rethink It. Harv Bus Rev 2018; Hbr.org/2018/06/25years-ago-i-coined-the-phrase-triple-bottom-line-heres-why-im-giving-up-on-it. Accessed 17 Oct 2023.
73. Marx A, Dusa A. Crisp-set qualitative comparative analysis (csQCA), contradictions and consistency benchmarks for model specification. Methodol Innov Online. 2011;6(2):103–48. https://doi.org/10.4256/mio.2010.0037.
74. Rihoux B, De Meur G. Crisp-set qualitative comparative analysis (csQCA). In: Rihoux B, Ragin C, editors. Configurational Comparative Methods. USA; Sage; 2009. pp. 33–68.
75. Hassoun A, Garcia-Garcia G, Trollman H, Jagtap S, Parra-López C, Cropotova J, Bhat Z, Centobelli P, Aït-Kaddour A. Birth of dairy 4.0: Opportunities and challenges in adoption of fourth industrial revolution technologies in the production of milk and its derivatives. Curr Res Food Sci. 2023;7:100535. https://doi.org/10.1016/j.crfs.2023.100535.
76. da Rosa RR, Goldschmidt G, Kunst R, Deon C, da Costa CA. Towards combining data prediction and internet of things to manage milk production on dairy cows. Comput Electron Agric. 2020. https://doi.org/10.1016/j.compag.2019.105156.
77. Karouani Y, Elgarej M. Milk-run monitoring system using the internet of things based on swarm intelligence. Int J Inf Syst Supply Chain Manag. 2022. https://doi.org/10.4018/IJISSCM.290018.
78. Pratama YP, Basuki DK, Sukaridhoto S, Yusuf AA, Yulianus H, Faruq F, Putra FB. Designing of a smart collar for dairy cow behaviour monitoring with application monitoring in microservices and internet of things-based systems. In: IES 2019 – International Electronics Symposium: the Role of Techno-Intelligence in Creating an Open Energy System towards Energy Democracy, Proceedings; 2019. pp. 527–533. https://doi.org/10.1109/ELECSYM.2019.8901676
79. Arago N, Robles RR, Alvarex C, Mabale A, Legista C, Repiso N, Amado T, Jorda RJ, Thio-ac A, Tolentino LK, Velasco J. Smart dairy cattle farming and in-heat detection through the internet of things (IoT). Int J Integrated Eng. 2022;14(1):157–72. https://doi.org/10.30880/ijie.2022.14.01.014.
80. Yavari A, Georgakopoulos D, Agrawal H, Korala H, Jayaraman P, Milovac J. Internet of things milk spectrum profiling for industry 4.0 dairy and milk manufacturing. In: International Conference on Information Networking; 2020. pp. 342–347. https://doi.org/10.1109/ICOIN48656.2020.9016608
81. Faruq, Syarif I, Ahsan AS, Al Rasyid MUH. Health monitoring and early disease detection on dairy cow based on internet of things and intelligent systems. In: IES 2019—International Electronics Symposium: the Role of Techno-Intelligence in Creating an Open Energy System towards Energy Democracy. 2019;183–188. https://doi.org/10.1109/ELECSYM.2019.8901527
82. Habsari W, Udin F, Arkeman Y. An analysis and design of fresh milk smart grading system based on internet of things. In: IOP Conference Series: Earth and Environmental Science. 2022; 1063:012059. https://doi.org/10.1088/1755-1315/1063/1/012059. Accessed 24 Oct 2023.
83. Ahi P, Searcy C. A comparative literature analysis of definitions for green and sustainable supply chain management. J Clean Prod. 2013;52:329–41. https://doi.org/10.1016/j.jclepro.2013.02.018.
84. Jordan E, Gross ME, Javernick-Will AM, Gavin MJ. Use and misuse of qualitative comparative analysis. Constr Manag Econ. 2011;29(11):1159–73. https://doi.org/10.1080/01446193.2011.640339.
85. Jabeur SB, Serret V. Bankruptcy prediction using fuzzy convolutional neural networks. Res Int Bus Finance. 2023;64: 101844. https://doi.org/10.1016/j.ribaf.2022.101844.