Reliability Engineering and System Safety 204 (2020) 107208

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Operational limits for aquaculture operations from a risk and safety perspective

Xue Yang[a,*], Ramin Ramezani[c], Ingrid Bouwer Utne[a], Ali Mosleh[b], Pål Furset Lader[a]

[a] Department of Marine Technology, Norwegian University of Science and Technology (NTNU), NO 7491, Trondheim, Norway
[b] The B. John Garrick Institute for the Risk Sciences, Department of Materials Science & Engineering, University of California, Los Angeles, CA, United States
[c] Department of Computer Science, University of California, Los Angeles, CA, United States

ARTICLE INFO

Keywords:
Machine learning
Operational decisions
Risk analysis
Operational limits
Aquaculture operation

ABSTRACT

Current decision making regarding whether to abort a high-risk aquaculture operation in a Norwegian fish farm is mainly experience-driven. The on-site personnel decides whether to start/delay/abort operations primarily based on their subjective judgement about whether they can handle the situation. The risk is considered implicitly as "gut feelings". There are no explicit operational limits nor a structured process to derive these for high-risk operations. In this research, a predefine safety-critical attributes have been identified from major accident scenarios to guide machine learning process to define operational limits based on multi-source data. Bayesian network, Tree Augmented Naïve Bayes (TAN) search algorithms were selected to build up prediction model so that operational limits upon a given condition can be decided. The paper concludes that machine learning techniques have great potential to be used to support safe decision-making in high-risk aquaculture operation, and the risk-based operational limits facilitates better understanding of operational context, and comprehension of the meaning of several deviations which may indicate a dangerous situation.

1. Introduction

1.1. Norwegian aquaculture

Aquaculture in Norway has been identified as the sector with significant potential for further growth. The Norwegian fish farming industry is expected to grow fivefold by 2050 [1] compared to 2010. The most recent figures from Statistics Norway show that in 2018, the sector produced 1.35 million tons of fish for human consumption, with a first-hand value of almost €6.5 billion, of which, Atlantic salmon made up 95% of the total [2]. Despite the positive prediction, the industry is facing challenges of a lack of sheltered coastal sites and increasing negative ecological consequences due to sea lice, fish escapes and farm waste left on the seabed [3]. The industry is also experiencing technological innovations in more exposed locations. The severe wave and current conditions, irregular wind and sheer remoteness, and uncertainties in new technologies amplify the risk to both personnel and the fish [4, 5]. It is especially challenging to get skilled staff at exposed locations [6].

1.2. Safety in aquaculture

Safe production of Atlantic salmon is the key to ensure a healthy and sustainable expansion of the industry. The safety by definition is the "Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment" [7]. Achieving safety is predicated on reducing the risk "to a level that is as low as reasonably practicable, where the remaining risk is generally accepted" [8]. Safety has a broader interpretation in aquaculture industry due to dealing with living animals in an open marine environment to provide food to the end customers. Five dimensions of risk need to be considered and minimized to an acceptable level, which are the risk to personnel (i.e., personal injury and fatality), the environment (e.g., fish escape, pollution), the fish welfare (e.g., fish injury, mortality), the marine assets (e.g., fish farms, service vessels) and food safety (e.g., food poisoning of end customers) [9].

Abbreviation: AI, Artificial Intelligence; API, Application programming interface; BN, Bayesian network; FN, False negatives; FP, False positives; ICT, Information and communications technology; IMR, Inspection, Maintenance, and Repair; LOA, Length of the vessel; ML, Machine learning; NYTEK, Technical requirements for fish farming installations; ROC, Receiver operating characteristic curve; SVM, Support vector machine; TAN, Tree Augmented Naïve Bayes (TAN); TN, True negatives; TP, True positives; WEKA, Waikato Environment for Knowledge Analysis

* Corresponding author.
E-mail address: xue.yang@ntnu.no (X. Yang).

https://doi.org/10.1016/j.ress.2020.107208
Received 31 May 2019; Received in revised form 24 May 2020; Accepted 20 August 2020
Available online 23 August 2020
0951-8320/ © 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

1.3. High-risk aquaculture operations

Aquaculture operations (e.g., fish transportation, delivery of feed, feeding, net cleaning, delousing, and IMR (Inspection, Maintenance, and Repair), etc.) are critical to achieving safety objectives. Norwegian sea-based aquaculture is the second most dangerous profession after capture fisheries in terms of personal safety [10]. The industry is already operating at the edge of safety limits [24]. Operators working in aquaculture operations are exposed to harsh weather conditions such as high winds, stiff currents, and large waves that cause confined spaces unstable and moving. Crane operations can be complicated when the vessel is moving relative to net cages. The statistics of fish escape during 2010–2016 show that among sea-based aquaculture operations, most fish escapes happened during delousing operation, handling of the sinker tube, handling of the dead fish pump during net cleaning, and loading and unloading of fish [11]. The delousing operation is also critical to fish health and welfare. The Norwegian food safety authority has received over 400 incident reports during and after the delousing operation in one year indicating a severely compromised fish welfare [12].

1.4. Challenges in decision-making related to safe operations

The large and demanding aquaculture operations are mostly done by specialized service companies. The rapid development of the industry follows fast employment growth [13]. Concerns about inexperienced employee arise from both the industry and the authorities, especially in service companies. Today, one service company may provide service to over 100 fish farms in different locations along the coast of Norway. Some fish farms have more complicated topography than the others. The technical, operational and geographical differences among fish farms raise challenges for service companies. Operators must be able to make critical decisions to avoid personal injuries, system failures, fish escape and negative impact on the fish health and welfare before/during the operation in a relatively short time. The onsite personnel decides whether to start/delay/abort operations mainly based on weather conditions and their experience. What is perceived as bad weather is rather subjective depends on whether the personnel feels they can handle the operation in such weather. The risk is considered implicitly as “gut feelings”, which is described as “risk-as-feelings” [14]. There is no explicit operational limits and cut off criteria or for high-risk operations. This could be dangerous in some situations. Research has shown that in some adverse weather conditions, operators prioritize fish safety (e.g., reducing the risk of fish escape) over personnel safety [15]. The need for such operational limits, as well as a systematic and structured process for determining the limits, increases strongly.

1.5. Opportunities for a safer decision-making process

1.5.1. Multi-source data

The Norwegian aquaculture industry has taken the direction of using ICT-based solutions, such as monitoring and control tools in production. There are more operational data registered in the fish farms and services companies to keep records of the production and services provided. Along with the increase in using technology, the volume of open data is steadily increasing in the sector as well. Norwegian Meteorological Institute provides Application programming interface (API) to access historical weather and climate data. As such, obtaining weather data for different locations in the Norwegian coastal and marine areas has largely become feasible. Norwegian Food Safety Authority publishes sea lice reports from all fish farms weekly, and Norwegian Veterinary Institute publishes fish health data to reveal which fish farms have affected or are suspected of having pancreas disease and infectious salmon anemia.

All the above data from various sources contain valuable risk information that provides input to make safe operational decisions. A critical question, however, is to investigate how to identify the most critical information from multiple sources and how to integrate this information into risk models to derive operational limits. The risk-based operational limits should help the operator to interpret the presented information and comprehend the meaning of several deviations to highlight any indication of a dangerous situation.

1.5.2. Artificial intelligence and machine learning

One of the main paradigms of Artificial Intelligence (AI) is problem-solving in which an intelligent task to automate is interpreted as a series of problems to be solved [16]. In our safe decision-making scenario, the intelligent task(s) are problems that arise in determining major risk contributors and predicting whether an operation is risky and should be aborted.

Machine learning (ML) is a powerful technique for solving AI problems and has gained more traction over the past few years with the popularity of modern neural network and deep learning. Machine learning can be briefly explained as a reasoning strategy (inductive) that a program or AI agent harnesses to learn from past experiences or background knowledge to discover new, relevant information. It is normally employed a) in categorization in which it learns why examples are put together in a certain way, to later predict the category of an unseen example, and b) to learn a way of predicting the value of an unseen data attribute, given series of examples and some background information about those examples. In summary, ML is used to extract implicit patterns of data that cannot be easily found using experts and is concerned with ways to construct computer programs to improve automatically through experience. The main distinction between ML and statistical data modeling lies in their goals and strategies. Statistics is primarily concerned with model validity and accurate estimation of model parameters from which inferences are made. However, prediction of unseen examples that is the main goal of ML is less of a concern.

A wide range of fields try to harness ML techniques instead of relying solely on statistical models with strong or weak assumptions. However, within the aquaculture sector, modern data science methodologies, especially machine learning, have not been explored. The digitalization, open source data and operational data in today's aquaculture open the opportunity to employ ML techniques to build models to “unlock” the subjective experience of operators. By harnessing ML techniques, we would embrace the uncertainties brought by new technologies and more complex and unfamiliar operating environment.

1.6. Objective

The main objective of the paper is to explore an approach to define operational limits for aquaculture operations from a risk and safety perspective using modern data science. The defined operational limits aims to provide the aquaculture industry, especially the service companies, support to make safe operational planning decisions for both coastal and offshore fish farms.

2. Material and methodology

2.1. Methodology

As illustrated in Fig.1, the research starts with a literature review to identify accident scenarios and major risk factors that could impact the abort operation decision in typical aquaculture operations. An operational log from a service company is analysed and events are aggregated to extract relevant information. A selection of thirteen predictor attributes based on major risk contributors are identified and relevant data are collected from different sources. The data are integrated into one datasheet, cleaned, pre-processed, and transformed to categorical data to enable an effective machine learning process.

A statistical correlation analysis is carried out to identify attributes that significantly correlate with the decision. Attributes are evaluated by the use of machine learning techniques to find the most important predictors to avoid overfitting and improve prediction model performance. Subsets of attributes that are significant with respect to determining the decision are used to define operational limits, results of which are presented and discussed in the paper.

The operational decisions analysed in this study was obtained from a service company which provides aquaculture operation service to fish farmers, such as delousing, net cleaning, and mooring and so on. Such an operational log is used to record working hours for the clients (163 fish farms). The recorded information includes the date of the operation, time to start operation, time to finish the operation, name of the fish farm, type of operation, vessels used for operation, whether the operation is aborted, and general comments. In some circumstances, the causes for aborted operation are commented, for example, “the wind is too strong”, “the visibility is poor”. Four meetings have been carried out with the service company and sea-based fish farms. The topics covered issues related to operation procedures, past accidents, possible accident scenarios, and risk factors to obtain domain knowledge.

Whether the operation is aborted is interpreted as our classification problem which have two classes: “Abort operation = YES” and “Abort operation = NO”. WEKA (Waikato Environment for Knowledge Analysis) open-source machine learning suite/workbench [17] developed by the University of Waikato is used as a toolkit. Weka enables users to harness a large number of machine learning algorithms. The algorithms are used to select attributes and evaluate and develop a prediction model based on the most relevant factors. Genie (Graphical Network Interface) software developed by the University of Pittsburgh [18] is also used to reproduce and present the prediction model to infer under different conditions. More details about the process are given in the following subsections.

2.2. Accident scenarios and major risk factors

In Norwegian sea-based aquaculture, the most common modes of fatalities related to aquaculture operations are loss of vessel, man overboard, and blow from an object [19]. From 2004 to 2015, the majority (67%) of the accidents happened onboard work vessels, including wellboats (i.e., a fishing vessel with a tank for the storage and transportation of live fish), and 21% happened in the fish farm. Severe injuries are mostly due to blowing from objects, falls, and entanglement. [20] point out in current practices in the industry that the hazards are generally identified; however, there is a lack of knowledge about risk factors during aquaculture operations.

Fish escape is regarded as the most severe risk to the environment and is a challenge for the Norwegian aquaculture industry. Farmed fish is regarded as a threat to the wild fish population due to transferring of disease, interbreed, and competing with wild stock [21]. The fish escape has been designated as one of three main categories of Norwegian fishery crime since 2011 [22]. The operators face fines if found responsible for large-scale escapes (i.e., escape of more than 10 000 fish [21]), in addition to reputation damage. The major causes of the fish escape are structural damage and holes in the net [23]. Collisions between non-operation related vessels and fish farms can also cause severe damages, not only to the cages (in the worst case it may result in loss of the fish farm) but may also lead to large-scale fish escape [24, 25].

In the research presented in this paper, the following five accident scenarios with corresponding risk factors are used as a basis for the work.

- Scenario 1: fish escape due to structural damage
- Scenario 2: fish escape due to holes in the net
- Scenario 3: fatality due to loss of vessel
- Scenario 4: fatality due to blow from objects
- Scenario 5: loss of fish farm due to ship collision

The following subsections describe important risk factors involved in the accident scenarios related to the weather, work vessel, the fish farm, operation, and crew. These factors are further used to derive attributes for machine learning research.

2.2.1. The weather

The weather condition is a predominant accident cause in aquaculture operations. The weather includes wind, waves, current, precipitation, daylight, and relative humidity. Interviews conducted by [23] show that what is perceived as bad weather depends on the individual farm site and its location in terms of typical wind direction and currents. Generally, as long as the personnel on the fish farm perceive that the weather is “acceptable”, the operation would not be called off.

2.2.1.1. Wind and waves. The upper parts of the vessel and the fish farm are mainly influenced by wind, while wind-generated wave and current loads mainly influence the lower parts. The wind load could account for 5–10% of the total forces on the mooring system of a farm [26]. Strong wind and high waves increase the capsizing risk of service vessels (scenario 3). It becomes difficult for the service vessels to berth to the floating collar and increase the chance for mooring lines of the net cage and the net come in contact with the vessels' hull and propellers. This increases the possibility to make holes on the net so that fish can escape (scenario 2). It is also observed that in rough weather, a strong wind gust can make it difficult for large service vessels to moor to the cages and that the operation has to be aborted eventually. Crane operations are particularly vulnerable to wind and waves. The strong wind increases the risk of the blow from objects in lifting operations (scenario 4). Moreover, the rough weather conditions make operations between vessels and fish cage challenging to perform, and, in turn, increase the risk related to operator slips/trips/falls while moving on the vessel deck or net cage. The strong wind and high waves also increase the likelihood of fish escape, especially for certain operations, such as handling the net and the sinker tube [23] (scenario 2).

The wave conditions depend on wind condition, topography, proximity to open sea, and bathymetry. There are two types of waves: wind-generated and swell-generated waves. In sheltered areas where the most fish farms are located, the ocean swells are not significant to operability, while the wind-generated waves are the major concern while planning for aquaculture operations.

2.2.1.2. Current. Current loads contribute to approximately 70% - 75% of the total forces on a typical mid-size fish cage in current conditions 0.5–1.0 m/s [26]. Besides, during aquaculture operations, when the service vessels, especially large service vessels moor at the weather side of the cage, the anchor load could increase significantly in high current velocity. There is limited numerical study of large service vessels operating at a fish farm in current, but a similar research on wellboat shows that the anchor load to the cage could increase by up to 90% in current velocity 1.0 - 1.5 m/s, because of the wellboat [27]. In such situation, the risk of escape due to structural damage will increase (scenario 1). The current is therefore considered a potential risk factor for aquaculture operations.

2.2.1.3. Visibility. Poor visibility has been identified as a risk factor in maritime accidents [28-31]. Heavy rain and fog can result in reduced visibility. The reduced visibility increases the demand for vessel maneuvering and positioning skills, which may increase the likelihood of vessels colliding with the fish farm (scenario 5). [23] reported darkness as a factor that led to human mistakes during operation, and night work has contributed to previous fish escapes and near accidents (scenario 2). With poor lighting, it is difficult to ensure that operations are performed properly. In some operations like net cleaning, the visibility becomes even poorer as water and biofouling are sprayed around.

2.2.2. The vessel

2.2.2.1. Size of the vessel and age of the vessel. Aquaculture operations are performed by different types of vessels. For cleaning operations, mooring, inspections, changing of nets, and transportation of workers between shore and the fish farm, the most used vessels are 14.9 m catamarans, 25 m catamarans, and 40 m monohulls. 60 m vessels are sometimes used for delousing operations and mooring operations. In general, the small and medium vessels are more vulnerable to waves than the large service vessels. The statistics show that the fleet of vessels of 24 m and smaller has the largest share of shipwreck and the highest fatality rate among the accidents from 2007 to 2016 [32]. The majorities of accidents registered by the Norwegian Maritime Authority in 2015 and 2016 involved small vessels under 15 m [32].

The vessel related accidents have not been analyzed statistically with focus on the aquaculture sector. In the maritime industry, the size of the vessel and age of the vessel have been identified as risk factors that influence the severity of the accidents [31]. Age of the vessel has been identified as an indicator of ship conditions [31], but old ships do not always suffer more accidents than young ships [33, 34]. The quality of the shipbuilding, combined with the maintenance of the ship during operation, influence the accident rates. [34] point out that an older ship over 20 years of age is more prone to total loss accident.

2.2.3. The fish farm

2.2.3.1. Location of the farm. The conventional fish farms are usually placed in areas that are sheltered from waves by islands or fjords. This implies that if the waves come from a direction such that islands are placed directly in front of the production facility; the operational conditions are better than if the waves come from an open sector. In harsh weather conditions, if the vessel fails to secure the mooring line to the cage, the vessel may drift quickly and collide with the islands or rocks nearby (scenario 3).

2.2.3.2. Age of fish farm. Age of fish farm (i.e., calculated from the commissioning year) could be an indirect indicator of its technical condition. Also, in 2004, the technical standard NS 9415 "Marine fish farms - Requirements for design, dimensioning, production, installation and operation" was introduced as a response to a high number of fish escapes because of structural failures [35]. The NYTEK regulations (i.e., Technical requirements for fish farming installations) require that a fish farm shall comply with the standard which put the requirement to the physical design of all the main components in an installation, functionality after assembly, and how the installation shall be operated to prevent fish escape. The companies producing main components (i.e., net pens, floating collars, mooring lines, barges, etc.) need to have a producer's certification, and the products need to have a certificate, to ensure the equipment can sustain the operating environment. The effect of NYTEK and NS 9415 to reduce structural failures is significant, especially after the last revision of NS 9415 in 2009. Under the assumption that producer's certification could improve the quality of the fish farm components, whether the fish farms were built before 2009 or after 2009 is also considered under age of fish farm (see Table 1).

2.2.4. The operation

2.2.4.1. Type of operation. Most fish escapes happened during delousing operation, handling of the sinker tube, handling of the dead fish pump, net cleaning, and loading and unloading of fish [11]. These operations increase the likelihood of tearing holes in the net (scenario 2). The operations that involve the use of cranes expose the operators to higher risk and blow by objects (scenario 4) and have become the single most significant cause of fatalities in the last twenty years [19, 36]. These lifting operations have also been identified as contributing factor to fish escape [23].

2.2.5. The crew

Operational errors during delousing or maintenance have been identified as one of the causes to fish escapes [21]. Fish escape has happened, for example, due to the net being left open after sorting fish or changing of nets and tearing of the net when crowding the fish (scenario 2). A study carried out by [23] shows that the underlying causes to human errors are: interaction with technology, physical working environment, workload, work pressure, skills, training, experience, communication and safety management in general.

Human factors are also identified as a contributing factor to maritime accidents, along with the condition of the ship and other external factors, such as bad weather [37]. [31] summarize human conditions that contribute to accidents from different sources which include unprofessional behavior, decision failures, inappropriate planning, misuse of equipment, failures related to supervision, inadequate attention, communication and cooperation, distraction, confusion, fatigue, health, and education.

2.3. Data collection and data transformation

In our classification problem between the classes "Abort operation = YES", and "Abort operation = NO", we face the following challenges. First, the data obtained for the year 2018 from January to April is a continuous operational log that records both "Abort operation = YES" and "Abort operation = NO" classes. However, the data for the year 2016 and 2017 only contain "Abort operation = YES" class. This is because of the limitation of accessing to the database that we could not get the complete dataset. Second, if we classify the classes only based on continuous operational data for 2018, there is a low number of samples of "Abort operation" = YES (12%). Classification based on imbalanced datasets is rather challenging since one class (often the interesting one) does not have enough samples. As such, it would be difficult for learning algorithms to generalize and form a hypothesis around the minority class. There are a number of ways to tackle imbalanced datasets: under sampling the majority or oversampling the minority data samples. Oversampling technique is commonly performed by synthesizing data instances, the result of which should be considered with a grain of salt. However, in under sampling, no artificial data is produced [38]. Samples of "Abort operation = YES" from the operational log years 2016 and 2017 are merged with data from 2018 so that we have 724 samples with 390 class YES and 344 class NO.

2.3.1. Data collection and limitation

Thirteen attributes, described above, reflecting the risk factors and influencing the predicted outcome "Decision to abort operation" are used in this study. These attributes include *average wind speed, maximum wind speed of gust, significant wave height, type of operation, visibility, wind comes from open sector, precipitation, daylights, maximum relative humidity, vessel size, vessel age, age of fish farm*, and *island/Rock on collision path*. Current and crew factors have been identified as risk factors, but they are not included in the current analysis due to limited access to these data.

In today's practice, recording of wind measurement is not common on most of the sea-based fish farms. Hence, in this research, these data are retrieved from the closest official observation sites from Norwegian Meteorological Institution [39]. One limitation is that the observation sites may be located tens of kilometers away from the fish farm. The other limitation is that the availability of hourly-based wind data is rather low for most of the locations.

The wave measurement equipment is rarely installed due to expensive cost, which means that practical wave related data are challenging to obtain. The significant wave height, which indicates the wave condition is estimated by using the fetch method as described by [40]. The method uses wind data (i.e., average wind speed (6 h) and wind direction) in connection with the fetch length to estimate

Table 1
The attributes and categories for the analysis.

No.	Attributes	Categories	Description
1	Decision to abort operation	Yes No	Predicted attribute
2	Average wind speed	B0: [0.0, 0.2] Calm B1: [0.3, 1.5] Light air B2: [1.6–3.3] Light breeze B3: [3.4, 5.4] Gentle breeze B4: [5.5, 7.9] Moderate breeze B5: [8.0, 10.7] Fresh breeze B6: [10.8, 13.8] Strong breeze B7: [13.9, 17.1] High wind, moderate gale B8: [17.2, 20.7] Gale, fresh gale B9: [20.8, 24.4] Strong/Several gale B10: [24.5, 28.4] Storm B11: [28.5, 32.6] Violent storm B12: [32.6, →] Hurricane	Mean wind speed at 10 m above sea level on the date of the operation. NS 9415 recommends using data from the nearest or the two nearest weather stations when establishing the long-term statistics of wind speed. Definition of categories is based on Beaufort wind force scale, which is widely used worldwide and also adopted by Norwegian Meteorological Institution.
3	Maximum wind speed of gust	B0: [0.0, 0.2] Calm B1: [0.3, 1.5] Light air B2: [1.6–3.3] Light breeze B3: [3.4, 5.4] Gentle breeze B4: [5.5, 7.9] Moderate breeze B5: [8.0, 10.7] Fresh breeze B6: [10.8, 13.8] Strong breeze B7: [13.9, 17.1] High wind, moderate gale B8: [17.2, 20.7] Gale, fresh gale B9: [20.8, 24.4] Strong/Several gale B10: [24.5, 28.4] Storm B11: [28.5, 32.6] Violent storm B12: [32.6, →] Hurricane	Strongest wind gust on the operation day. Gust is "a flurry of wind" that is more powerful and can be significantly higher than the value of average wind speed. They are therefore, dangerous for operations such as lifting [44]. Definition of categories is based on the Beaufort wind force scale.
4	Significant wave height	S1: [0–0.1) Calm (rippled) S2: [0.1–0.5) Smooth (wavelets) S3: [0.5–1.25) Slight S4: [1.25–2.5) Moderate S5: [2.5–4.0) Rough S6: [4.0–6.0) Very rough S7: [6.0–9.0) High S8: [9.0–14.0) Very high S9: [14.0, →) Phenomenal	In aquaculture today, the wave conditions are mostly based on subjective observation, instead of measurement based. The subjective observations are not recorded in the operational logs. Two parameters represent wave conditions: significant wave height and wave period. Significant wave height is a statistical description of the wave phenomena, which is defined as the average wave height for the highest third of the waves in one registration [35]. According to a study conducted by SINTEF Ocean [45], the operational limits for service vessels are mainly set based on significant wave height. Definition of categories is based on the World Meteorological organization's codes for sea state.
5	Type of operation	Delousing General Mooring Net cleaning Net inspection Remotely Operated Vehicle (ROV) inspection Inspection of ring	The operations that are recorded in the operational logs.
6	Visibility	Good Bad	Visibility of the environment during marine operation. The categories are retrieved from descriptive information registered in the comment column of the operational logs.
7	Wind comes from open sector	Yes No	Wind direction is considered in connection with the fetch method, whether it comes from the open sector or not (Fig. 2).
8	Precipitation	Mainly dry: [0, 0.4) Light rain: [0.5, 2) Rain/Snow: [2, 20) Heavy: [20, →)	The accumulated amount of precipitation in 24 h. The heavy precipitation may lead to bad visibility and creates difficulties for the operation. Definition of categories is adopted from the Norwegian Meteorological Institute [46].
9	Daylights	Good: the task can be carried out thoroughly under daylight Medium: half of the task can be carried out in daylight Bad: the task will be carried out thoroughly in the darkness	Sunrise and sunset information of each location is retrieved from website [47].
10	Maximum relative humidity	R0: [0, 10) R1: [10, 20) R2: [20, 30) R3: [30, 40) R4: [40, 50) R5: [50, 60) R6: [60, 70) R7: [70, 80) R8: [80, 90) R9: [90, 100) R10: [100)	Maximum relative humidity in 24 h. The humidity may correspond to the possibility of fog forming, which would lead to bad visibility.
11			

(continued on next page)

Table 1 (*continued*)

No.	Attributes	Categories	Description
	Vessel size (Length of the vessel)	Small: 15 m Medium: 25 m Large: > 40 m	The vessel size is defined based on the length of the vessel (LOA), which are usually used in aquaculture operations.
12	Vessel age	Very old: [25,) Old: [20, 25) Medium-old: [15, 20) New-Medium: [10, 15) Relative new: [5, 10) New: [0, 5)	Age of the vessel at the time of the operation. The study of accidents about the age of the ships shows that most of the vessels are over 25 years old [33], which are defined as "Very old". In 2015, the Norwegian Maritime Authority enforced a new regulation for cargo vessels of 8–24 m length, which introduced new technical requirements for the aquaculture fleet below 24 m to improve the safety level of the service vessels. This forms the basis to define to vessels less than 5 years old as "New".
13	Age of fish farm	Old: [<1979) Old-Medium: [1979, 1989) Medium: [1989, 1999) Medium-New: [1999, 2009) New: [2009, →]	The age of the fish farm. To incorporate the effect of NS 9415, the fish farms that are established after 2009 is classified as "New". The others are grouped in a ten-year range.
14	Island/Rock on collision path	Good: The average fetch length > 500 m Bad: The average fetch length < = 500 m	The average fetch length from the *Fetch analysis* also indicates the vicinity of the site to islands or land and can be used to indicate whether there are islands in an average radius of 500 m. The resolution of the fetch search is 1 deg x 50 m, so islands smaller than 50 might not be detected.

significant wave height upon fish farms that are registered in the operational log on the day of the operation. The basic principle is that the longer the fetch length and the higher the wind speed, the more energetic the sea state will be, which means that the fish farm is exposed to a tougher operating environment. Fetch plot can be generated with indicated wind direction, which illustrates whether the wind comes from the open sector[1] or not (Fig. 2). The fetch analysis must be used with the understanding that there is uncertainties in the estimates of the wave parameters. The method assumes deep water and constant wind over a period of 6 h. Also, it does not include swell waves.

The operation and vessel related information is retrieved from the service company, and the fish farm related information (e.g., location, commission year) is obtained from the Norwegian Directorate of Fisheries [41] for further processing.

2.3.2. Data transformation

Data transformation is the process of aggregating, summarizing, and in general preparing the data into forms that can be used by machine learning algorithms [42]. In this project, original operational data consists of several events and actions logged during an operational day, and these events are all aggregated into one row. The final daily operational decision is presented in the log and highlights whether the operation was eventually aborted or not. Often, during a day, the operation starts and later aborts due to change of weather condition, and it would be possible to keep all interim decisions and their corresponding weather conditions as one data point. However, since we did not have access to hourly weather data that caused the possible changes of the decision during each day, we focus on the general weather condition of an operational day and final decision in this study. Therefore, we decided to only keep the final decision of each day along with corresponding weather conditions.

The original dataset consists of a mixed numeric (e.g., average wind speed) and categorical data (e.g., type of operation, outcome decision). The numeric data is discretised as this has been shown to be an effective measure to have improved performance of several Bayes net and logistic regression techniques [43]. The defined categories and descriptions of the attributes are listed in Table 1.

[1] When the fetch reaches 40 km, the search for island is terminated, and the sector is assumed to be open [40] Lader P, Kristiansen D, Alver M, Bjelland HV, Myrhaug D. Classification of Aquaculture Locations in Norway With Respect to Wind Wave Exposure. ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering: American Society of Mechanical Engineers; 2017. p. V006T05A5-VT05A5.

2.4. Classification

2.4.1. Classification methods

In an initial analysis, the performance of classification model with all 13 attributes using different classifiers are assessed. The classifiers include decision trees, Bayesian networks, and Support Vector Machines. The best success rate reaches 87.4% with ROC area 0.942, which means 87.4% of the cases can be appropriately classified based on the pre-identified attributes. This can be considered as a significant result in the domain of aquaculture operation, in which we don't have all the possible information at our disposal.

Random Forest (unlimited depth), Bayesian network using Tan search algorithm, and support vector machine (SVM) showed good predictive power with rather similar performances. Among the classifiers, the graphical Bayesian network (BN) model which captures the compositional structure of the relations offers an interesting perspective on interpreting the outcome decision. BN is a graphical representation of a joint probability distribution of a set of attributes. The node in the directed acyclic graph represents a variable, and a directed arc provides independence/dependence relationships between the nodes. A BN is both descriptive and predictive. One attractive feature of BN is the inference capability of the class given the observed values of the attributes. We may also infer the probability distribution for some attributes given the values or distributions for the remaining attributes. Therefore, the BN was chosen for further detailed analysis.

In the research, K2 with the number of parents restricted to a predefined maximum, Hill-climber, and Tree Augmented Naïve Bayes (TAN) are used as network training algorithms. K2 algorithm has a fixed ordering of attributes and processes each node in turn, and greedily considers adding edges from previously processed nodes to the current processing one to maximize the network's performance [48]. Hill-climber algorithm follows a hybrid search-and-score principle that first reconstructs a skeleton of a BN and then adding and deleting arcs with no fixed ordering of attributes to improve the network performance [49]. The algorithm for learning TAN classifiers first learns a tree structure over node set, then use mutual information tests conditioned on classification node. A link is added from the classification node to each attribute node [50]. The class then has no parents, and each attribute has the class attribute as the parent.

2.4.2. Accuracy of classifiers

The following classifier accuracy measures are used to evaluate the prediction models.

- Classification success rate: the percentage of test set tuples that are

Table 2
A confusion matrix for positive and negative tuples.

Predicted class			
		a (*Abort operation* = NO)	*b* (*Abort operation* = YES)
Actual class	*a* (*Abort operation* = NO)	True positive (TP)	False positive (FP)
	b (*Abort operation* = YES)	False negative(FN)	True negative (TN)

correctly classified.

- Confusion matrix (contingency table): a measure that shows how well the classifier can recognize tuples of different classes. In our case, *Abort operation* has two classes: YES or NO. *True positives (TP)* shows the number of cases that the classifier predicted NO, and they indeed are NO in the test dataset. *True negatives (TN)* shows the other class that they are predicted YES and actually are YES too. *False positives (FP)* are the cases that are predicted YES, but their actual class is NO. *False negatives (FN)* is the number of tuples that are predicted NO, but they actually are YES. High *True negative* rate is practically more important than high *True positive* rate, because the more *Abort operation* = YES are correctly classified, the better to ensure safe operations. The low *True positive rate*, which means high *False positives* rate that the operation was classified not need to be aborted, but aborted, is not safety critical. So, if two models have the same classification success rates, higher TN is preferred than the lower one. The results of the confusion matrix in Section 3 are presented based on contingency table (Table 2).
- ROC area (the area under Receiver operating characteristic curve): ROC curve is a plot that shows the True positive rate against the False positive rate at various threshold settings. The area under the ROC curve is a metric commonly used to evaluate the overall performance of the classification model. A perfect classifier will have an area of 1.0. Therefore, we can evaluate the models by comparing the areas under the different ROC curves.

2.5. Relevance analysis

Relevance analysis, which includes correlation analysis and wrapper feature selection, are used to highlight attributes that are more relevant to the outcome (abort operation or not-abort operation decision) [42]. The results of the relevance analysis give a list of selections of attributes for further evaluation via different classifiers to find the best prediction model.

The correlation analysis is carried out between attributes and outcome, based on only 2018 dataset that has a representative of 391 aggregated events. Person's Chi-square test is commonly used to carry out the correlation analysis for categorical attributes. However, in the dataset, more than one cell of the contingency table has less than five observations, which means the data is considered skewed and Chi-square is not suitable anymore. Fisher's exact test, which is specifically designed for small samples (less than 1000), is used instead, as recommended by [51].

The following data processing are carried out for Fisher's exact test to avoid bias caused by missing values, as recommended by [52]:

a) Dataset without missing values.
b) All the events that have missing values are excluded so that a complete dataset can be available for the statistical analysis. 172 events are left for the analysis.
c) Dataset with missing values imputed.
d) All missing values are marked as level "NaN". All 391 events are included in the analysis.
e) Replace all missing values with means from the data for the statistical analysis. All 391 events are included in the analysis.

[53] pointed out that attributes that are significantly correlated with an outcome do not necessarily improve the prediction. This means significant attributes are not always good predictors, and the attributes with strong predictivity sometimes fail to be significant. In this study, the wrapper method using predictive machine learning algorithms is applied to evaluate attributes sets. One merit of wrapper method is that it supports detection of interaction between attributes. Another merit is that it searches for optimal attribute subset for the desired machine learning algorithm, which is Bayesian network in our study.

3. Results

The dataset (724 instances), which contains 390 class "Abort operation = YES" and 344 class "Abort operation = NO", is divided into two parts, with 2/3 allocated to a training set and the remaining 1/3 allocated to a test set. The training dataset is used to build the prediction model, while the test dataset is to estimate the accuracy of the classification model. Predictive capacity is assessed based on success rate, confusion matrix, and ROC area.

The BN generated using the selected features by the algorithms reported success rate 85.4% with ROC area 0.939. The Bayesian network TAN learning algorithm treats the classification node (i.e., abort operation) as the first node in the ordering to learn the structure, which means the classification node is treated as the parent of all other nodes. The Bayesian network is reproduced using GeNIe and the inferred conditional probabilities from data are entered into the nodes (Fig. 3). The conditional probabilities from each state are estimated from the data using the maximum likelihood method.

It is observed from the dataset that four attributes have the ability to override others when they reach a certain state. When *Average wind speed* reaches B7 [13.9, 17.1], or *maximum wind speed of gust* reaches B10 [24.5, 28.4], or *visibility* is Bad*, or precipitation* is heavy, the operations are aborted. This is also in line with the BN that has been reproduced. When above-mentioned four attributes are set up to limit states, the probability of abort operation is updated to be 97%, 90%, 84% and 82%. If we use the metaphor based on the traffic light, a single red light can call off the operation, the operational limits can be any of these overriding attributes reaching the above-mentioned limits. The challenge lies in situations when some of the overriding attributes are in a marginal situation which can be designated as orange. If two or three attributes have orange lights, what would be the best suggestion to the operators? The Bayesian network opens possibility to derive operational limits in such situations by updating belief based on entered evidences. In most circumstances, certain factors are known while planning for operations, such as type of operation, farm age, vessel size, and vessel age. The operational limits can then be represented by states of a combination of other attributes. For example, if we plan to do net cleaning operation tomorrow, using new small service vessels, farm age is new, what would be the operational limits in terms of weather conditions? From the updated BN based on evidences (Fig. 4), we can roughly interpret that, if the forecast for tomorrow indicates that the maximum wind speed of gust would be approximately or higher than B7 (i.e., [13.9, 17.1] m/s), average wind speed approximately or higher than B3 (i.e., [3.4, 5.4]), rain, there is a higher likelihood that the operation will be aborted. In such a manner, the BN provides the possibility to derive operational limits upon the operational contexts, such as condition of the vessel and the farm, and type of operation, to facilitate operational planning decision-making.

4. Discussion

The proposed approach is a risk-assessment guided development process using machine learning techniques to derive operational limits to support planning decisions related to high-risk aquaculture operations. The risk to personal safety, fish escape are considered in the process. The objective of the study is not to propose accurate operational limits, but an approach to define such limits. From the development process of the limits, the following topics for discussion arise, which are:

- Novelty of the approach
- Usefulness of risk-based operational limits
- Advantage and disadvantage of the approach
- Limitations

4.1. Novelty of the approach

Machine learning has been applied to safety and risk research in recent years mainly to explore the factors that contribute to accidents [26-29]. Within the aquaculture sector alone, the advantage of machine learning has not been explored and harnessed. In this study, we used a risk assessment guided approach and the possibilities that arise in open source data and operational data in aquaculture to explore subjective and experience-based decision-making process. The results show that based on carefully selected attributes and multi-source data, there is a potential to crack the code of "gut feelings" of the operators to explicitly express operational limits from a safety and risk perspective.

The process starts from identification of accident scenarios from accidents reports. Major risk factors that may contribute to the identified scenarios are further identified from relevant literature in aquaculture field and other industries that have similar operating environment. The attributes that can represent the risk factor are established based on domain knowledge and further feed into machine learning process. The results show that the initial selection of attributes guided by the accident scenarios and identified risk factors have a rather good predicative performance of the decision. The approach eliminates the number of attributes so that the computation cost can be reduced significantly.

The BN, Tan search algorithms was selected to build up the model based on most important attributes following wrapper feature selection method. The operational limits derived from BN enables inference under conditions of uncertainty so that operational limits can reflect known evidences and unknow status of the other attributes.

4.2. Usefulness of risk-based operational limits model

The research shows that the operational limits model has potential to improve situational awareness of operational contexts, which is critical for service companies who have to operate in more than 100 locations that have different technical, operational and geographical conditions. The model will facilitate understanding of how interactions between risk factors can influence the decision. The interactions are relatively hard to capture especially in an unfamiliar location with complicated typology. The risk-based operational limits should help the operator to interpret the presented information and comprehend the meaning of several deviations to highlight any indication of a dangerous situation. The model can be one of the responses to industrial challenge of not able to get sufficient skilled staff [7].

The model of operational limits derived from the proposed approach can also contribute to guiding data collection by specifying data need to be more accurately recorded. Acquisition of data is the most time-consuming and difficult task in this study due to challenges discussed in Section 2.3.1. The research reveals the importance of collecting and sharing critical data to ensure safe operations. For instance, the data quality can be improved in the future by recording weather forecast data while planning the operation, and wind measurement data prior to decisions.

4.3. Advantage and disadvantage of the approach

In a data science project, it is essential to gather data, aggregate and integrate it, clean the data and pre-process it, and ultimately select the features and design a predictive system. Thus many trials and errors should be performed while going through all the mentioned stages [54]. It is estimated that data cleaning and exploration constitutes as much as 80% of a data mining effort [55]. The risk assessment guided initial selection of attributes shows good results, and a good preparation of the attributes at the beginning of the analysis process can save time, avoid unnecessary trials and errors.

One finding of the study is that the feature selection process discarded several attributes without significantly influencing the predictive accuracy. This includes *wind comes from open sector, daylights, maximum relative humidity*, and *island/rock on collision path*. Following the law of parsimony, a simpler model that contains less number of features that can have higher predictive accuracy compared to more complex models is the drive behind feature selection. These features may be irrelevant in presence of the others or redundant from data processing perspective. Whether they can be regarded as non-safety critical is subject to discussion. The analysis is based on historical operational logs. What is "unlocked" are what the operators have been considering while making abort operation decisions. The derived model can illustrate how the decisions have been made, but not necessarily how the decisions should be made.

The performance of the derived operational limits is highly dependent on the quality of attributes interpreted from identified risk factors. There is uncertainty related to whether the most important risk factors have been identified, whether the attributes represent the risk factors sufficiently, and whether data for each attribute is available. For example, visibility can be influenced by precipitation, relative humidity, and daylights, however, sometimes, low visibility cannot be explained by these three attributes only. Sea fog usually occurs at a relative humidity near 100% (R10); however, fog can also form at a lower humidity, and sometimes even with a relative humidity of 100%, the fog fails to form . To ensure the most representative attributes, sufficient domain knowledge will be required. A line of research about risk and risk indicators can be the basis for better defining the attributes. Earlier work on safety-critical parameters, which are factors that have direct and significant influences on the risk involved in operation [56], can be further explored in further work.

4.4. Limitations

This approach assumes that the decisions (i.e., abort operation or not) were risk-based, which means the operators were considering the risk of doing operation while making the decisions. The scenarios are the ones that expose both the personnel and the fish to high risk. The scenarios are used to derive major risk factors to further find possible relevant attributes. However, the factors that may influence the severity of consequences are not covered in the study. In other words, the quantifications of losses is not considered/covered. This is a limitation of the approach.

In our case, the Bayesian network is selected as the most feasible model due to its high predictive performance, powerful representation of knowledge, and inference capability under conditions of uncertainty. There is inherent uncertainty in machine learning techniques,

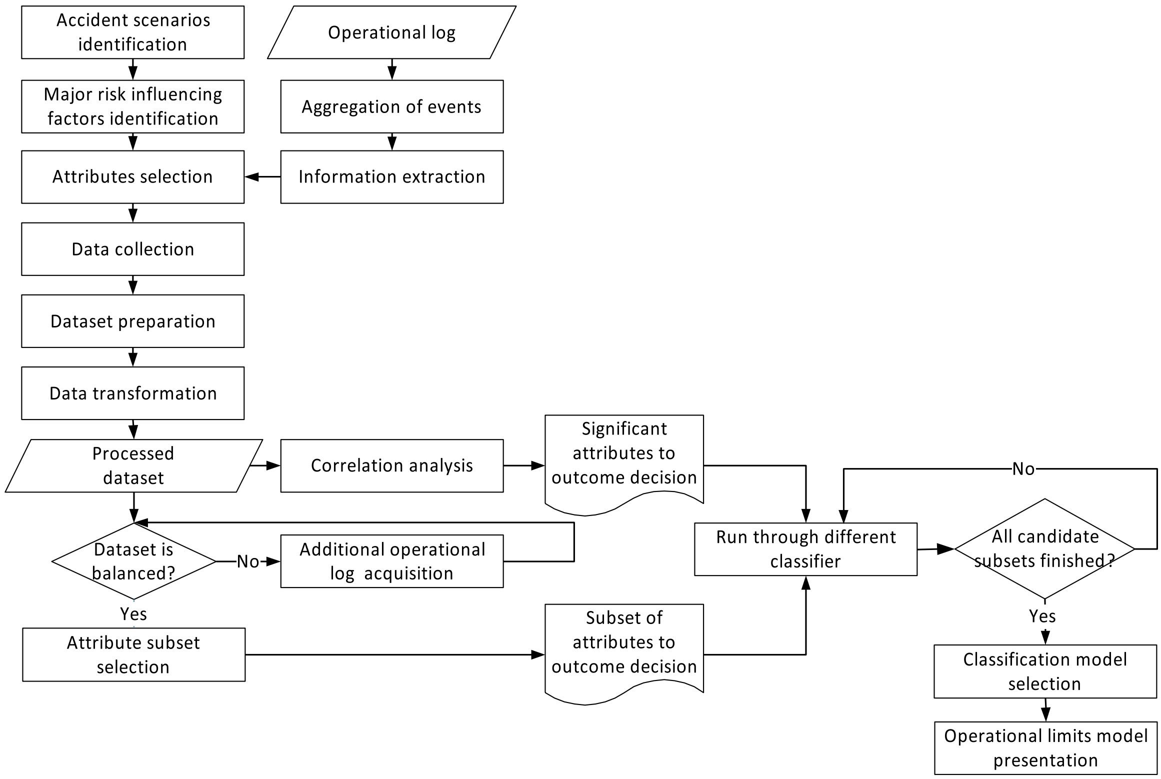

Fig. 1. Method used in the research presented in this paper.

especially in the attribute subset selection domain, due to the statistical nature of most of its algorithms. It is still a challenge that sometimes the attributes look irrelevant in isolation, but may become relevant in combination [54].

Besides the uncertainty rooted in the machine learning techniques, there are also uncertainties with respect to handling missing values. As typical, imputing missing values usually gives more accurate models

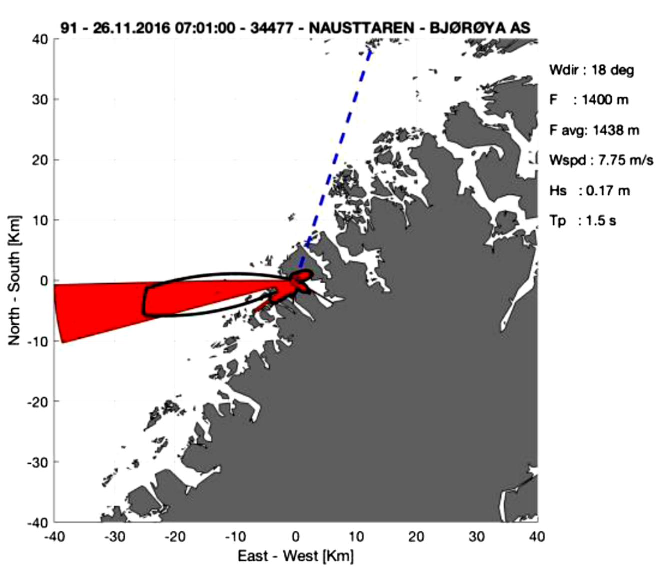

Fig. 2. Example of fetch plot that shows the wind direction and open sector [50].

than dropping the column entirely [42]. In this research, part of the missing values is filled by the most probable values manually. For example, some missing "gust"-values are replaced by the indicated wind speed in the comments´ column if there is any. Some algorithms (e.g., Bayes net classifier) replace all missing values in the dataset automatically the modes and means from the training data [57]. The uncertainties of the replaced value may introduce bias into the final results.

The on-site validation of the operational limits model is not covered in the current study yet. To be used by the service company, the model has to be further learned by larger datasets, and digitalization of the data collection and transformation will be necessary. Such validation will be implemented at later stage of the project with close collaboration with the service company.

5. Conclusion

The objective of the research is to explore the possibilities that arise from multi-source data to propose an approach to define operational limits, as an input to support safe operational decision-making. A risk assessment guided development process using machine learning techniques is proposed, and the resulted operational limits model contributes to a better understanding of operational contexts. The digitalization in aquaculture industry and advances in data science open up the possibility to turn implicit experiences to explicit knowledge. The operational limits and the proposed approach to define such limits, will be validated on field operations in the same service company. Based on the research outcome and test results, a guideline for data collection will be prepared to improve data quality, especially for site-specific data. The potential of digitalizing the results to ICT system will also be discussed in the further work.

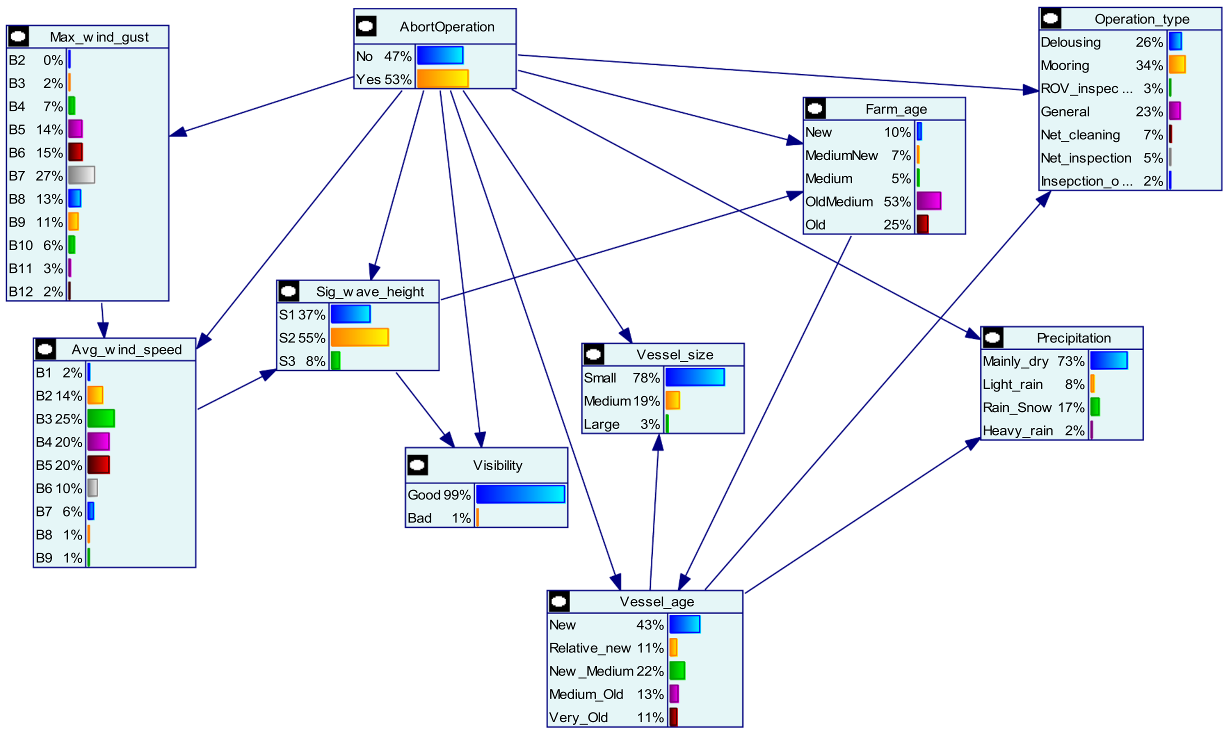

Fig. 3. Bayesian network built by selected attributes using Tan algorithm.

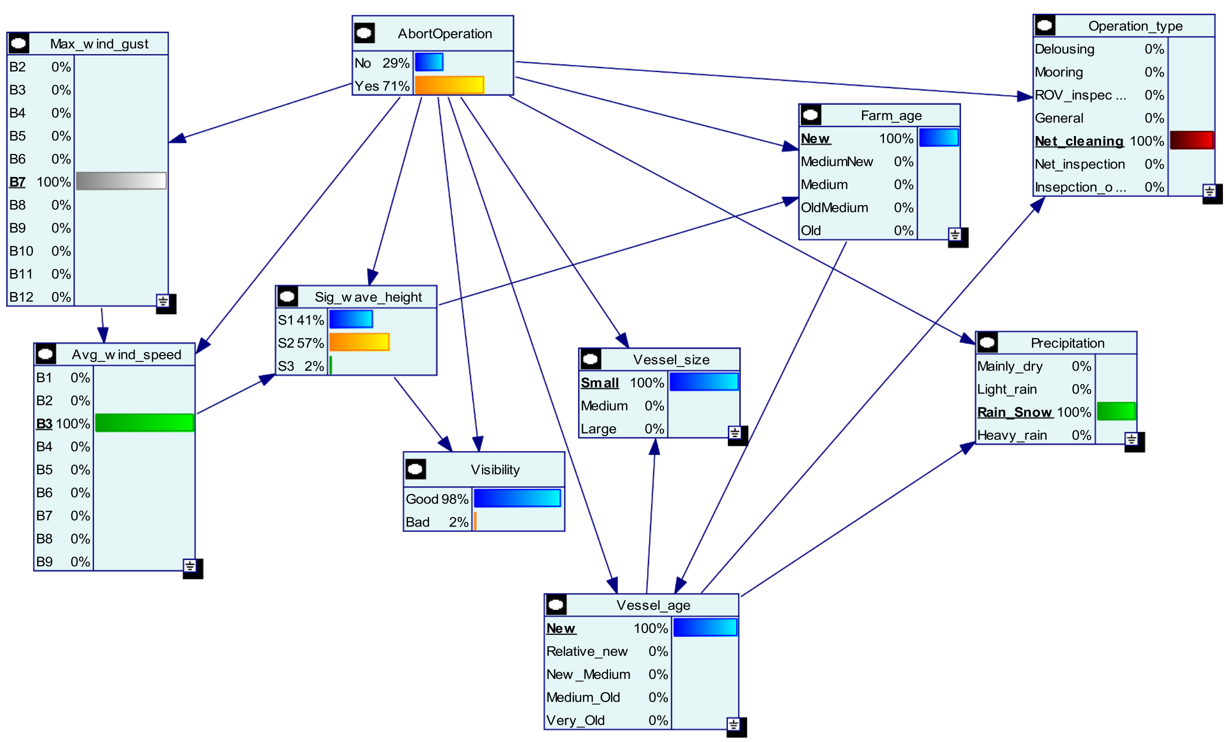

Fig. 4. Updated Bayesian network by setting up evidences.

CRediT authorship contribution statement

Xue Yang: Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft. **Ramin Ramezani:** Conceptualization, Methodology, Writing - review & editing. **Ingrid Bouwer Utne:** Conceptualization, Writing - review & editing, Funding acquisition. **Ali Mosleh:** Writing - review & editing. **Pål Furset Lader:** Data curation.

Declaration of Competing Interest

None.

Acknowledgments

This work has been carried out as part of the Reducing Risk in Aquaculture project. The Norwegian Research Council is acknowledged as the main sponsor (project number 254913). We would like to thank the service company for providing operational data and insightful feedbacks. We also thank Bjørn Magnus Mathisen and Dr. Jeevith Hegde for their assistance with data acquisition.

References

[1] Olafsen T, Winther U, Olsen Y, Skjermo J. Value created from productive oceans in 2050. trondheim. Norway: SINTEF Ocean; 2012.
[2] Statistics Norway. Aquaculture, 2018, final figures. https://www.ssb.no/en/jord-skog-jakt-og-fiskeri/statistikker/fiskeoppdrett. 05.04.2020.
[3] Holmer M. Environmental issues of fish farming in offshore waters: perspectives, concerns and research needs. Aquac Environ Interact 2010;1:57–70.
[4] Bjelland HV, Føre M, Lader P, Kristiansen D, Holmen IM, Fredheim A, et al. Exposed aquaculture in Norway. OCEANS'15 MTS/IEEE Washington: IEEE; 2015. p. 1–10.
[5] Gardar J. Storms wreak havoc on marine harvest test cages. http://www.intrafish.com/news/646806/storms-wreak-havoc-on-marine-harvest-test-cages. 30.05.2017.
[6] Marine Harvest. Experience and challenges. SFI Exposed Days 2017.
[7] Department of Defense. MIL-STD-882D, standard practice for system safety. Washington, DC: U.S.Department of Defense; 2000.
[8] Rausand M. Risk assessment: theory, methods, and applications. 1 editon Hoboken, New Jersey: Wiley; 2011.
[9] Yang X, Utne IB, Holmen IM. Methodology for hazard identification in aquaculture operations (MHIAO). Saf Sci 2020;121:430–50.
[10] Aasjord H, Geving IH. Accidents in norwegian fisheries and some other comparable norwegian industries. IFISH 4 - The 4th international fishing industry safety and health conference. 2009.
[11] Føre HM, Thorvaldsen T. Causes for farmed salmon and trout escape during period 2010-2016. Trondheim: SINTEF Ocean; 2017.
[12] Norwegian Food Safety Authority. Annual report 2016. Norwegian Food Safety Authority; 2017.
[13] EY. The norwegian aquaculture analysis 2017. Ernst & Young AS; 2018.
[14] Finucane ML, Holup JL. Risk as value: combining affect and analysis in risk judgments. J Risk Res 2006;9:141–64.
[15] KV Størkersen, first Fish. Sharp end decision-making at Norwegian fish farms. Saf Sci 2012;50:2028–34.
[16] Ramezani R.An artificial intelligence framework for investigative reasoning (Doctoral dissertation, Imperial College London).
[17] Witten I.H., Frank E., Hall M.A., Pal C.J. Data mining: practical machine learning tools and techniques 2016.
[18] BayesFusion. GeNIe. https://www.bayesfusion.com/.
[19] Holen SM, Utne IB, Holmen IM, Aasjord H. Occupational safety in aquaculture – Part 2: fatalities in Norway 1982–2015. Mar Policy 2018;96:193–9.
[20] Holmen IM, Utne IB, Haugen S. Risk assessments in the Norwegian aquaculture industry: status and improved practice. Aquacult Eng 2018;83:65–75.
[21] Ø Jensen, T Dempster, Thorstad E, Uglem I, Fredheim A. Escapes of fishes from Norwegian sea-cage aquaculture: causes, consequences and prevention. Aquac Environ Interact 2010;1:71–83.
[22] Ministry of Fisheries and Coastal Affairs. Report from the working group established by the Ministry of Fisheries and Coastal Affairs (In Norwegian). Minist Fisher Coastal Affiars 2012.
[23] Thorvaldsen T, Holmen IM, Moe HK. The escape of fish from Norwegian fish farms: causes, risks and the influence of organisational aspects. Mar Policy 2015;55:33–8.
[24] NRK. Drove the boat into the fish farm (In Norwegian). https://www.nrk.no/hordaland/kjorte-baten-inn-i-oppdrettsanlegg-1.12366000. 10.20.2017.
[25] Reuters. Up to 80,000 trout escape after cargo ship crashes into fish farm in denmark. http://gcaptain.com/up-to-80000-trout-escape-after-cargo-ship-crashes-into-fish-farm/.
[26] Li P.A theoretical and experimental study of wave-induced hydroelastic response of a circular floating collar. 2017.
[27] Shen Y, Greco M, Faltinsen OM. Numerical study of a well boat operating at a fish farm in current. J Fluids Struct 2019;84:77–96.
[28] Goerlandt F, Goite H, Valdez Banda OA, Höglund A, Ahonen-Rainio P, Lensu M. An analysis of wintertime navigational accidents in the Northern Baltic Sea. Saf Sci 2017;92:66–84.
[29] Wróbel K, Montewka J, Kujala P. Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliab Eng Syst Saf 2017;165:155–69.
[30] Chauvin C, Lardjane S, Morel G, Clostermann J-P, Langard B. Human and organisational factors in maritime accidents: analysis of collisions at sea using the HFACS. Accident Anal Prevent 2013;59:26–37.
[31] Bye RJ, Aalberg AL. Maritime navigation accidents and risk indicators: an exploratory statistical analysis using AIS data and accident reports. Reliab Eng Syst Saf 2018;176:174–86.
[32] Norwegian Maritime Authority. Focus on risks 2018. 2017.
[33] Ventikos NP, Stavrou DI. Andritsopoulos AJJoME, technology. Studying the marine accidents of the Aegean Sea: critical review. Anal Results 2017;16:103–13.
[34] Eleftheria E, Apostolos P, Markos V. Statistical analysis of ship accidents and review of safety level. Saf Sci 2016;85:282–92.
[35] Standard Norway. NS 9415 Marine fish farms - Requirements for site survey, risk analyses, design, dimensioning, production, installation and operation. Pronorm AS Postboks 252, 1322 Lysaker, Norway. http://www.standard.no.: Standards Norway; 2009.
[36] Holen SM, Utne IB, Holmen IM, Aasjord H. Occupational safety in aquaculture – Part 1: injuries in Norway. Mar Policy 2018;96:184–92.
[37] Lema E, Papaioannou D, Vlachos GP. Investigation of coinciding shipping accident factors with the use of partitional clustering methods. Proceedings of the 7th international conference on pervasive technologies related to assistive environments. ACM; 2014. p. 42.
[38] Krawczyk B. Learning from imbalanced data: open challenges and future directions. Progr Artif Intell 2016;5:221–32.
[39] Norwegian Meteorological Institution. Frost API. https://frost.met.no/index2.html.
[40] Lader P, Kristiansen D, Alver M, Bjelland HV, Myrhaug D. Classification of aquaculture locations in Norway with respect to wind wave exposure. ASME 2017 36th international conference on ocean, offshore and Arctic engineering. American Society of Mechanical Engineers; 2017. V006T05A5-VT05A5.
[41] Norwegian Directorate of Fisheries. Aquaculture. https://www.fiskeridir.no/Akvakultur. 16.06.2017.
[42] Han J, Kamber M, Pei J. Data mining: concepts and techniques. 3 ed Elsevier Science; 2011.
[43] Ekbal A. Improvement of prediction accuracy using discretization and voting classifier. 18th International Conference on Pattern Recognition (ICPR'06). IEEE; 2006. p. 695–8.
[44] Standard Norway. NS-EN 13000:2010 + A1:2014 cranes - Mobile cranes. European committee for standardization; 2010.
[45] Heide MA. Service vessels 2010 (In Norwegian). SINTEF Ocean; 2010.
[46] NMI. Personal communication. Norwegian Meteorological Institute; 2018.
[47] Time and date. Sunrise and sunset times – Norway. https://www.timeanddate.com/sun/norway. 09/09/2018.
[48] Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Mach Learn 1992;9:309–47.
[49] Buntine W. A guide to the literature on learning probabilistic networks from data. IEEE Trans Knowl Data Eng 1996;8:195–210.
[50] Friedman N., Geiger D., Goldszmidt MJMl. Bayesian network classifiers. 1997;29:131–63.
[51] Bellolio MF, Serrano LA, Stead LG. Understanding statistical tests in the medical literature: which test should I use? Int J Emerg Med 2008;1:197–9.
[52] Little RJA, Rubin DB. Statistical analysis with missing data. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2002.
[53] Lo A., Chernoff H., Zheng T., Lo S-HJPotNAoS. Why significant variables aren't automatically good predictors. 2015;112:13892–7.
[54] Domingos P. A few useful things to know about machine learning. Commun ACM 2012;55:78–87.
[55] Dasu T, Johnson T. Exploratory data mining and data cleaning. Hoboken, N.J: Wiley-Interscience; 2003.
[56] Yang X, Utne IB, Holen SM, Holmen IM. Implications from major accident causation theories to activity-related risk analysis – an application to the Norwegian Atlantic salmon farming industry. In: Haugen S, Barros A, Gulijk Cv, Kongsvik T, Vinnem J, editors. Proceedings of the 28th international European safety and reliability conference (ESREL 2018). editorsNORWAY: CRC Press; 2018.
[57] Bouckaert R.R. Bayesian network classifiers in weka. 2004.