REPOSITORY AND KNOWLEDGE BASE ON INFECTIOUS DISEASES FOR SEAFARERS.

The results of the DESSEV Erasmus+ project.

GERMAN DE MELO RODRIGUEZ
Universitat Politécnica de Catalunya, Barcelona, Spain
e-mail: german.de.melo@upc.edu
Orcid: 0000-0001-9212-6476

IZABELA BODUS-OLKOWSKA
Maritime University of Szczecin, Szczecin, Poland
e-mail: i.olkowska@pm.szczecin.pl
Orcid: 0000-0003-4366-0116

TOMAZ GREGORIC
Spinaker d.o.o., Portoroz, Slovenia
e-mail: tomaz.gregoric@spinaker.si

NATALIA WAWRZYNIAK
Maritime University of Szczecin, Szczecin, Poland
e-mail: n.wawrzyniak@pm.szczecin.pl
Orcid: 0000-0002-4429-7163

NATASZA BLEK
Institute of Outcomes Research, Maria Sklodowska-Curie Medical Academy in Warsaw, Warsaw, Poland
e-mail: natasza.blek@uczelniamedyczna.com.pl
Orcid: 0000-0002-3213-7330

KACPER DZIEDZIC
Institute of Outcomes Research, Maria Sklodowska-Curie Medical Academy in Warsaw, Warsaw, Poland
e-mail: dzikac5@gmail.com
Orcid: 0000-0001-8930-2221

REZA ZIARATI
Centre for Factories of the Future Ltd., Kenilworth, United Kingdom
e-mail: reza.ziarati@c4ff.co.uk
Orcid: 0000-0002-4227-1728

VANESSA MAKAR
Centre for Factories of the Future Ltd., Alingsas , Sweden
e-mail: vanessa@c4ff.se
ORCID:0009-0001-5113-4207

ARIS CHRONOPOULOS
IDEC SA, Pireas, Greece
e-mail: aris@idec.gr

JANNE LAHTINEN
Satakunta University of Applied Sciences, Rauma, Finland
e-mail: janne.lahtinen@samk.fi
Orcid: 0000-0003-1640-2656

HEIKKI KOIVISTO
Satakunta University of Applied Sciences, Rauma, Finland
e-mail: heikki.koivisto@samk.fi
Orcid: 0000-0002-8955-6421

ACKNOWLEDGMENTS

This article was created in cooperation with the partners of the DESSEV project implemented by the Maritime University of Szczecin, Poland, as part of the Erasmus+ program. Financed by the European Union. Contract number: 2022-1-PL01-KA220-VET-000087987. For more information please visit: www.desssevproject.eu.

Keywords

seafarers, infectious disease, repository, knowledge base, DESSEV project, Erasmus+

Abstract

The COVID-19 pandemic has brought about unprecedented challenges, including lockdowns, difficulties in verification, staff swaps, and restricted international travel. These factors underscore our global unpreparedness for such crises. While these challenges have significantly affected life on land, their impact on maritime activities has been even more pronounced.

Life at sea has been particularly strained due to the inability to disembark, prolonged voyages resulting from organisational replacements, limited access to hospitals during sea travel, and challenges in early illness diagnosis and effective isolation. These hurdles have collectively made navigating the pandemic exceptionally difficult for maritime communities.

Within the framework of the DESSEV project, funded by the European Union Erasmus+, a DEcision Support System (DSS) addressing epidemic threats on sea-going vessels has been developed. This initiative includes the establishment of a learning repository and knowledge base on infectious diseases. The repository encompasses WHO recommendations, IMO guidance, and procedures from select countries for managing individuals with symptoms or an imminent case of infection on board. Additionally, the repository includes detailed medical cases presented in the form of scientific articles, all of which are accessible free of charge on the project website, www.dessevproject.eu.

A second key objective of the project was the creation of a database on infectious diseases. This database comprises 22 infectious diseases described with 35 symptoms, grouped into 8 categories. The accumulated knowledge serves as the foundation for the development of IF...THEN... rules in the form of decision trees.

This article presents the tangible results achieved through the DESSEV project. It provides insights into how the project has contributed to addressing the unique challenges faced by crews of sea-going vessels during the pandemic, offering a valuable resource for maritime communities and public health professionals.

1 INTRODUCTION

Maritime transport is critical in this globalised world since 90% of the raw and processed materials consumed worldwide are transported by sea. The world merchant fleet consists of approximately 100,000 vessels transporting more than 11 billion tons of all types of goods. This makes shipping a strategic transport for all countries that cannot be stopped, not even in pandemic conditions, as it would further aggravate its effects and impoverish all countries.

The COVID-19 epidemic was declared by the World Health Organization (WHO) a public health emergency of international concern on January 20, 2020. The characterisation of 'pandemic' now means that the epidemic has spread to several countries and/ or continents of the entire world, and that affects a large number of people.

Against the backdrop of the global COVID-19 pandemic crisis, the maritime sector has once again turned its attention to the issue of epidemic risks on maritime vessels. The ship, as an isolated entity, can prove to be a potential source of epidemics. It is important to note that there is no specialised infectious disease unit or adequate medical staff on seagoing vessels. Several documents and studies provide some guidance on how to proceed on board a ship when an infectious disease is detected (Garcia, M., et al., 2017, Smith, J., & Jones, A., 2020, Patel, S., & Smith, K., 2016). However, the COVID-19 pandemic showed that these guidelines were insufficient in the case of a new unknown disease. Maritime transport, like other modes of transport, was affected by this COVID-19 pandemic, so the crew members of the world fleet were exposed to the contagion of the virus, as they docked with their ships in different ports on all continents. They did not have any medical professional on board who could determine if any of their crew members had contracted COVID-19 or another disease and the measures to be taken to cure it. The reasons given have led the members of the Maritime University of Szczecin, Poland (MUS), Polytechnic University of Catalonia (UPC), Satakunta University of Applied Sciences (SAMK), Medical University of Marii Skłodowskiej-Curie (MU MSC), SPINAKER, IDEC and the Center for Factories of the Future (C4FF), to carry out a project with ERASMUS+ funds. The objectives of the project are

- repository of data on epidemic situations;
- knowledge base in the form of rules (IF..., THEN...);
- decision support system on the risk of epidemic threats on a maritime vessel.

2 REPOSITORY OF INFECTIOUS DISEASES

The first aim of the DESSEV project was to develop a repository of knowledge on how to recognise infectious diseases, how they can be contracted and guidance on how to counter the spread or prevention of the disease. At the time of writing, the repository contains 102 items of literature and includes IMO (International Maritime Organization) recommendations, WHO (World Health Organization) recommendations, the guidelines of Centres for Disease Control and Prevention, specific case reports and a number of scientific medical articles, primarily related to the most recent COVID-19 outbreak.

All publications in the repository are available online on the project website: www.dessevproject.eu. Each document is categorised into two classes: instructions for vessel crews and medical. The first mentioned category contains a number of IMO guidelines, and recommendations from the WHO and other institutions related to the port service. The part of the repository relating to medicine is mainly concerned with travel medicine, referring to incidents of infectious diseases precisely on seagoing vessels, both passenger and cargo types. Many of the articles in the repository are, of course, about COVID-19, as the most recent case of global disease. It is thanks to the pandemic of this coronavirus that special attention has actually been paid to the principles of prevention before contracting the disease, education on how to prevent the spread of the disease and steps towards recovery. It also contains a number of scientific articles on cases of dealing with the presence of an infectious disease on board. Individual documents have also been prioritised by rank: very important, important, and less important. The very important class includes all IMO manuals and WHO recommendations, as well as medically relevant scientific articles describing the most serious diseases. The important class includes port regulations and advice from port authorities in the event of an infectious disease being detected on board a ship. The remaining category concerns popular science articles related to the description of individual cases of disease on board. Each document listed on the project website includes a title, the indication of the category and priority of importance, as well as an author`s identification and scope of the content. As the project is concerned with decision support for the crew management of a seagoing vessel in the event of the detection of an infectious disease outbreak on board, the vast majority of the material in the repository relates specifically to guidance and cases of detection and prevention of the spread of infectious diseases on board the vessel precisely (Fig.1).

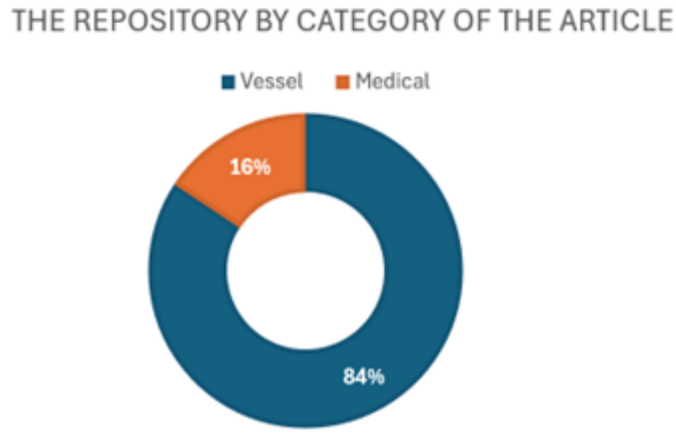

Fig. 1 [The repository by category to which the materials relate]

Figure 2 presents the percentage share of each author in the publications contained in the repository for vessel parts. It is worth noting that the Authors of scientific articles category is the largest. This means that the repository was not only based on IMO recommendations and WHO guidance but also acquired and made available in one place several articles presenting specific cases of shipboard infectious disease, as well as descriptions of response actions and prevention procedures.

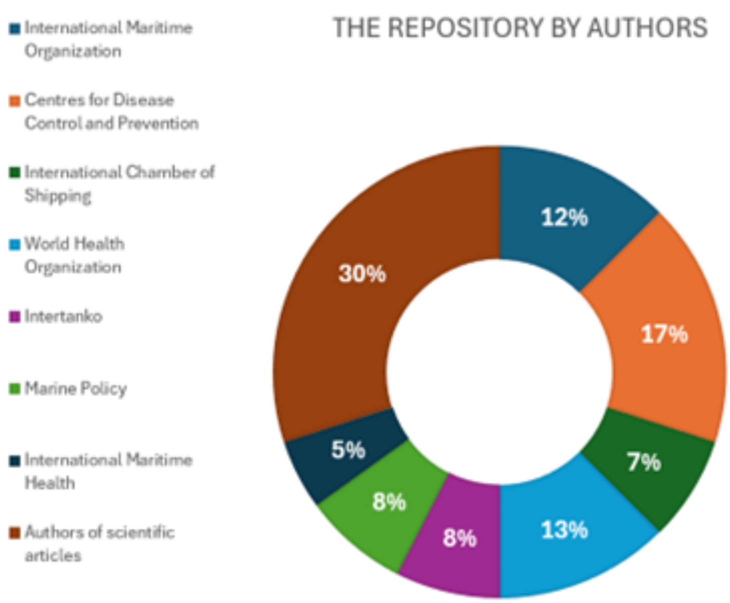

Fig. 2 [The repository by the authors]

3 KNOWLEDGE BASE OF SYMPTOMS AND INFECTIOUS DISEASES

Drawing upon current statistics and data from leading health organisations such as the World Health Organization (WHO, https://www.cdc.gov/nndss/index.html access by day 23 Feb 2024) through its Global Health Observatory (GHO), the Centers for Disease Control and Prevention (CDC, https://www.cdc.gov/nndss/index.html, access by day 23 Feb 2024) via the National Notifiable Diseases Surveillance System (NNDSS), and the European Centre for Disease Prevention and Control (ECDC, https://atlas.ecdc.europa.eu/public/index.aspx access by day 23 Feb 2023) through its Surveillance Atlas of Infectious Diseases, we have identified 19 diseases for concentrated analysis and monitoring within the scope of this project. These diseases have been selected based on their significant impact on global health, their prevalence, and the trends observed in recent data, thereby ensuring that our system is equipped to handle infectious diseases that pose threats to maritime operations.

Acquiring medical knowledge and identifying important and effective parameters regarding the diagnosis of infectious diseases occurred in two steps. In the first step, English language-based literature concerning selected 22 infectious diseases, presented in Table 1, diagnostic criteria for various infections, and guidelines for diagnosing infectious diseases were identified and included. Each disease was then described by several signs grouped into specific 8 categories (Table 2). Further literature was identified using the references of those already reviewed. Types of literature include international peer-reviewed articles, online reports, commentaries, editorials, electronic books and press releases from universities and research institutions, which include expert opinions. Grey literature published by the WHO, the US Centers for Disease Control and Prevention (CDC) and other local government publications and information outlets were also included. Research databases examined included PubMed, Google Scholar, Embase, Medline and Science Direct.

List of selected diseases:	
Chickenpox	Mumps
Chikungunya	Norovirus
Cholera	Pertussis
COVID-19	Rabies
Dengue	Rubella
Diphtheria	Tetanus
Ebola	Tuberculosis
Infectious mononucleosis	Typhoid and paratyphoid fever
Influenza	Hepatitis A
Malaria	Yellow fever
Meningococcal infection	Zika

Table 1 [List of selected diseases]

Groups of signs of infectious diseases			
1. General/systemic signs	continuous fever or fever with intervals less than 1 day intermittent fever every	**5. Hematological symptoms**	bleeding manifestations

	2-4 days lethargy sweating and/or chills head pain lack of appetite and/or weight loss		
2. Respiratory signs	chest pain cough phlegm shortness of breath sore throat runny nose	**6. Gastric symptoms:**	abdominal pain diarrhoea nausea vomiting
3. Musculoskeletal signs	back pain joint pain muscle pain lockjaw	**7. Dermatological or associated signs:**	neck swelling skin rash yellow skin and/or dark urine
4. Neurological signs	blurry vision cognitive difficulties difficulty swallowing dizziness emotional agitation neurological problems with sensation and movement seizures stiff neck and sensitivity to light	**8. Others signs:**	fear of water testicular pain eye redness

Table 2 [List of groups of signs]

In the second step, targeted interviews were conducted with medical experts to determine the crucial elements in the diagnosis of infectious diseases and their relationship to clinical decision-making parameters.

We have developed a table representing the frequencies of various disease symptoms. Each row in the table corresponds to a specific disease, and each column corresponds to a particular symptom. The intersection of a row and column contains the frequency or occurrence of a specific symptom for a particular disease.

Obtaining precise percentages for the occurrence of symptoms associated with various diseases proved challenging in evidence-based papers, reflecting a potential gap in detailed symptom frequency reporting within the existing literature.

4 PREDICTION ALGORITHM

The aim of the prediction algorithm is to give the possible infectious disease/s a patient may have based on her/his symptoms. The prediction algorithm has to be trained, and we trained it with the knowledge base of symptoms and infectious diseases.

The knowledge base of symptoms and infectious diseases was expressed in percentages, i.e. how many patients out of 100 would express a specific symptom when they were infected by a specific infectious disease. Based on these data, we randomly generated hundreds of artificial patients with specific symptoms but reached all of them the exact percentages from the knowledge base. E.g. if for a specific disease 25% of patients have symptom 1, 50% of them symptom 2 and 100% of them symptom 3, we could generate the following 5 patients from these data and the overall percentages would still fit the initial data (Table 3). We used this approach of randomly generating artificial patients to take into account that every human being is unique, so the symptoms appearing after infection are slightly different for each person.

Patient	Symptom 1	Symptom 2	Symptom 3
1	1	1	1
2	1		1
3		1	1
4			1
5		1	1
6			1
7		1	1
8			1
	25%	50%	100%

Table 3 [Randomly generated data from the knowledge base of symptoms and infectious diseases]

We finally used these randomly generated data to train our prediction algorithm. Our prediction algorithm uses three different AI models: random forest, decision tree and naive bayes. So, these models are used to directly predict a disease based on given symptoms.

The naive bayes model (Hand, 2001) is a classifier which assumes that the symptoms are conditionally independent, given the target disease. This assumption's strength (naivety) is what gives the classifier its name.

The decision tree model (Rokach, 2013) is a tree-like model of symptoms and their possible diseases, including chance event outcomes, resource costs, and utility. Each branch represents the outcome of the test (if a symptom is present or not), and each leaf node represents a disease. The paths from root to leaf represent classification rules are presented on Fig. 3.

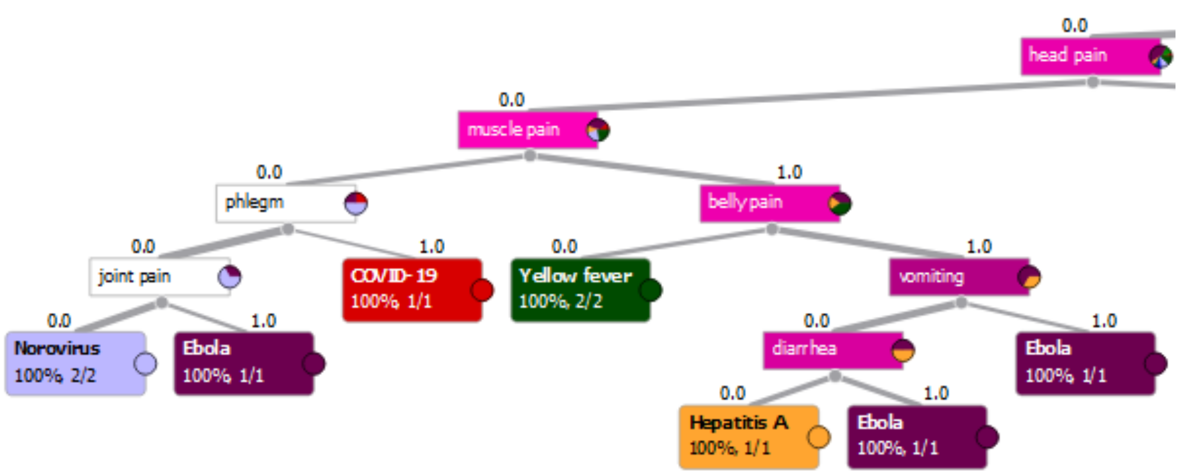

Fig. 3 [Small part of the decision tree model]

Random forest model (Breiman, 2001) is based on the decision tree model, but in the random forest model, a forest (a big number) of decision trees is generated considering only some symptoms for each decision tree. The output of the random forest is the disease selected by most decision trees.

The authors used Orange data mining software to build these models (Fig. 4).

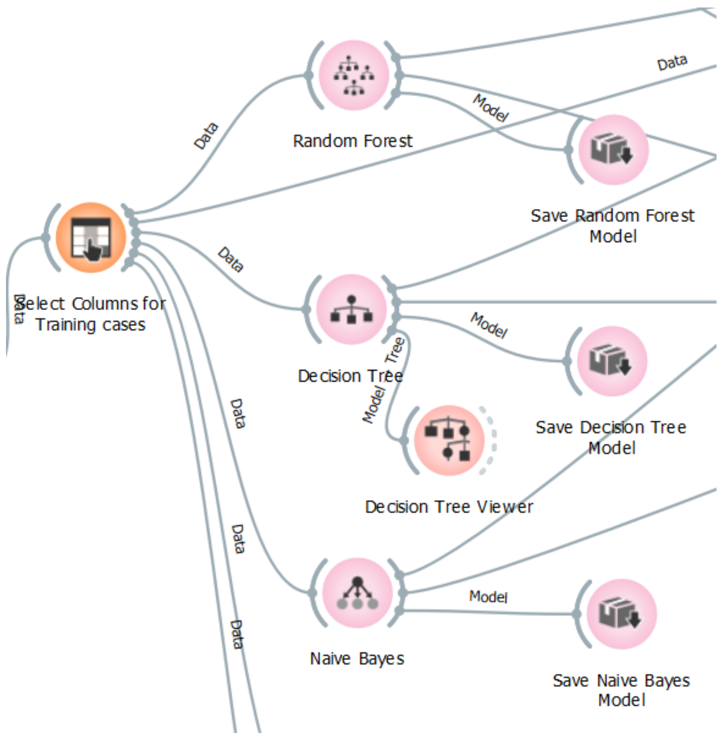

Fig. 4 [Partial visualisation of the project in Orange data mining software]

Orange data mining software is one of the software for machine learning and data mining. We used it as it is a free, widely used software from which the generated models can be exported and then used outside the software just by using an Orange python library (Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox in Python, Journal of Machine Learning Research 14(Aug): 2349−2353. http://jmlr.org/papers/volume14/demsar13a/demsar13a.pdf). This allows us to easily integrate the prediction algorithm into the website or phone app and make it publicly available.

When tested on training data, the classification accuracy of all three models reaches nearly 100%, but the most important is the classification accuracy on test data. We collected a small number of real patients (1 per disease) and used their symptoms/disease combinations as the test data. It should be stressed that the test data was not used during the training process (while building these models). The classification accuracy of all three models when tested on test data was as follows:

- random forest: 100%
- decision tree: 57%
- naive bayes: 86%

In our prediction algorithm, the random forest model outperforms other models like in most similar applications in the medicine field (Sumwiza at all., 2023). These results confirm that the random forest model is the best model for disease prediction in the medicine field.).

5 CONCLUSION

This article introduced a tool established as a delivery of the DESSEV project. The online tool is available for seafarers with low thresholds, shortlisting the possible diagnosis based on the most likely causes of symptoms.

Furthermore, the tool guides medical care actions until professional medical care ashore is reached.

The introduced online tool leans into the knowledge base of symptoms and the prediction algorithm. This study has elaborated on the mechanism leading to the likely cause of symptoms. Despite the undeniable advantage the online tool provides to the onboard personnel, the knowledge base and the prediction algorithm have functional restrictions emerging in operational contexts.

First, the online users of the tool are not medical professionals. The online tool inevitably introduces seagoing personnel with unfamiliar medical terms and expressions. Moreover, some expressions, such as hard cough or high fever, are subjective. Relevant questions are: how heavy symptoms are required to meet the expressions "heavy" and high"? User-provided data remains an evaluation of the circumstances and, as such, subjective information. The online tool bases the diagnostic evaluation on the data provided by the user and the repository data bank. Therefore, it must be considered that all results provided by the tool are subjective to some extent.

Second, the online repository provides onboard personnel with timely and peer-reviewed information on infectious diseases. A practical shortcoming of the repository is the vocabulary and the associated medical expressions. Much of the repository's content may be challenging to interpret for non-medical personnel, potentially leading to misinterpretations. Two actions should always be considered. First, consultation of shore-based medical assistance should be commenced as soon as possible. Second, the repository content should guide the onboard diagnoses and the following treatment actions only when the interpretation of the online repository content is unambiguous.

Third, the practical implementation of the online tool requires co-existence with the shipping company's prevailing safety management system. It is generally understood that shipping companies operate their vessels under the International Safety Management Code facilitated by the International Maritime Organization. The safety management systems have existing procedures for handling onboard medical emergencies. The role of the online tool in supporting the accredited procedures and guidelines in the safety management system should be carefully planned. A controlled embedding of the online tool in the vessel onboard safety management system is conditional to the usability of the online tool with a low threshold.

Finally, it is noteworthy that the online tool is not meant to be a substitute for professional medical care. Under all circumstances, all effort must be seen to seek shoreside medical assistance. The online tool is a promising concept and a successful response to the DESSEV project list of deliverables. The online tool functionality requires verification through onboard use. Vessel crew feedback and user experience are valuable for further online tool development.

REFERENCES

[1] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
[2] Hand, David J., and Keming Yu. "Idiot's Bayes: Not So Stupid after All?" International Statistical Review / Revue Internationale de Statistique 69, no. 3 (2001): 385–98. https://doi.org/10.2307/1403452
[3] Kellen Sumwiza, Celestin Twizere, Gerard Rushingabigwi, Pierre Bakunzibake, Peace Bamurigire, Enhanced cardiovascular disease prediction model using random forest algorithm, Informatics in

Medicine Unlocked, Volume 41, 2023, 101316, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2023.101316
[4] Rokach, Lior and Oded Maimon. "Data Mining with Decision Trees - Theory and Applications. 2nd Edition." Series in Machine Perception and Artificial Intelligence (2013)