New Approach to the Dissemination of Bivalve Consumption Safety Information

Manuel Beja da Costa
Instituto Superior Técnico, Universidade de Lisboa, Portugal

Background: Intoxication rates due to bivalve poisoning in Europe have prompted the European Commission to pass a regulation requiring fishery companies to include specific information on their commercial product labels. This allows consumers to evaluate bivalve safety for consumption. The use of a mobile application and OCR functionalities allows and easy access to this information. This study describes an algorithm for the for automated label-reading (LRA) of bivalves, evaluates whether labels include all the necessary information, and validates the algorithm's efficiency, concerning the correction of produced data.

Methods: Eighty-six labels of bivalves were randomly selected from multiple sale points in the Lisbon metropolitan area and photographed using regular smartphones. Each label was evaluated by a human analysis to determine its completeness regarding mandatory information. An automatic data extraction algorithm was developed, used on the collected labels and validated against the data extracted by the human analysis. All the implemented algorithms were included on a mobile application which was submitted to usability tests.

Results: The dataset analysis shows that 23.25% of the labels did not include all the information mandated by the European Commission, these labels lacked the harvesting zone. About 30% of the collected labels were originated in a foreign country. The LRA performs with an accuracy of 76.74%.

Conclusions: High accuracy of the developed LRA shows potential for providing instant confirmation of the bivalve consumption safety to consumers. Its integration into a mobile application will suffer from the incompleteness of included information, the variability of label structure and multiple countries of the products. Label readability would increase with the standardization of the label structure, and enforcement of European Commission regulations. Despite the fact that the LRA not being accurate enough for production phase, a form-based query mobile application has potential to improve the safety conditions of bivalve consumption.

***Index Terms*—Food safety, mobile health, bivalves, molluscs, commercial labels**

I. Introduction

Live bivalve molluscs are a popular ingredient in gastronomical traditions throughout Europe and have an important impact on the European economy[1]. These shellfish, however, are vulnerable to exposure to diverse marine biotoxins and bacteria that can pose health risks[2] to consumers.

Due to the impact of bivalve intoxication on public health, the European Commission passed regulations defining procedures that guarantee the quality and safety of these products for human consumption. For instance, Regulations 854/2004[3] and 853/2004[4] demand each member state to: define the geographical delimitations of the harvesting zones, classify these zones based on contamination levels and associated risks, regulate water quality test procedures and regulate the monitoring of the process from harvest to consumption. If the results of a water quality test suggest a risk to human health, the harvesting area must be shut down, and food business operators must ensure that any products recently harvested from that area are not traded, sold or consumed.

To allow the participants in the supply chain (consumers, commercial vendors, and food inspection authorities) to verify the origin and harvesting date of the product, fishery companies should follow a specific labeling regulation[3] that extends the general one [5].

National regulation entities are required to run toxicity tests on water and bivalve samples taken from each of the harvest zones under their jurisdiction. From these tests, bivalve consumption safety is determined, and this information is made accessible to the public through dedicated websites.

Although these are regularly updated and publicly available, access to their information can be facilitated and optimized by a mobile application. Such mobile application, included in the mHealth[6][7] class of applications, would improve food safety-related health through facilitating access to information. For example, the Food Safety Mobile Application [8] from the Centre for Food Safety of Hong Kong allows users to obtain and share general food safety information in a universal and straightforward way.

A similar mHealth application can be developed to track bivalve toxicity by allowing the user to verify the product's consumption instantaneously. This application would benefit both commercial entities, private consumers and inspection authorities, conveniently providing accurate information.

The work here described is twofold, 1) it evaluates if the labels include all the information necessary for the correct assessment of the consumption risks, and 2) to determine if an automatic label-reading algorithm can accurately process and decode the available information, in order to be implemented on a mobile application.

II. Methods

A set of 86 random labels was collected throughout the Lisbon metropolitan area in various fish markets and supermarkets between July 2017 and January 2018. The labels were found on the packaging of bivalves.

Two different smartphones, *iPhone 6S* and *Samsung J7*, were used to photograph the labels using the standard settings and without any particular lighting or photo preparation. The use of different devices intends to expose the algorithm to different image resolutions to assure its robustness to the resolution variability expected from potential users.

After the collection of labels, two different analysis were made, as described in the next sections.

A. Label - Human analysis

All 86 labels were studied for completeness of information, and a table was created detailing the following variables for each label:

- species;
- date of packaging;
- zone of harvest;
- harvesting method (aquaculture or wild harvest), and
- country of harvest.

Each of these variables must be included on the label according to the guide released by the European Union regarding consumer labels[9], but only species name, date of packaging and zone of harvest are necessary for an accurate assessment of consumption safety by the algorithm. An incomplete label is defined as lacking at least one of the variables required by the previously mentioned guide.

The table created is also used to validate the results obtained from the algorithm's readings and calculate percent accuracy.

B. Label - Algorithmic analysis

A label reading algorithm was developed to try to extract the relevant information from the bivalves labels.

After multiple experiments, the best architecture for this software system is presented in Figure 1.

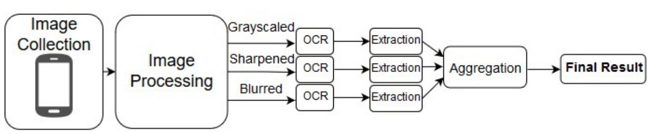

Fig. 1. Architecture of the label recognition algorithm.

After the collection of the labels (image collection step), the first stage of the algorithm is image processing. To optimize the next processing stages and reduce influences of camera capture quality and label faults, different image processing techniques were applied. The original photo was submitted to: conversion to greyscale; application of a 3x3 kernel sharpening filter[10] to the original image, and a generation of a binary image by applying histogram equalization followed by a 3x3 kernel blurring filter[10]. The sharpening filter improves the reading accuracy on poorly focused images, while the blurring filter's adaptive thresholding makes the algorithm robust to shadows.

This image processing step produces the outputs in the form of three distinct images that will be independently processed in the second stage of the algorithm by the optical character recognition (OCR) module. This OCR module was developed using the Tesseract Python[11] library that implements the Tesseract[12] algorithm. This library source code and necessary training data are publicly available in the open source community. To further optimize the produced output, Tesseract was configured to recognize only the Portuguese alphabet and to allow sparse text. The adjustment for Tesseract to read sparse text was necessary to increases accuracy when reading text with an unstandardized structure, as is the case with the bivalve labels sampled in this study.

The third stage performs the data extraction from the text returned from the OCR, by doing pattern matching and statistical analysis. Since the labels include information that is not essential for consumption safety assessment, some of the text that is read by the OCR module is irrelevant. Furthermore, some of the text recognized in the OCR stage may not exactly match the expected information because when Tesseract is incapable of recognizing the full word, it provides an approximation. These two factors require additional processing of the OCR results. An extra step on the algorithm that is capable of estimating and giving an approximation to the real data. The algorithm compares the text returned with the expected data formats (in the case of dates) and values (in the case of zones and species names). The algorithm has been programmed to compare and match the OCR returned text with the following:

- list with all harvesting zones;
- list with all species, in Portuguese and Latin;
- list with all possible countries of origin;
- date format with the ”/”, ”-” and ” ” separators;

By matching the returned text from each processed image with this expected text for each characteristic, and using proximity formulas, we can assign each processed image a fixed set of values for each characteristic: species, zone of harvest, harvesting method, date of packaging, and country of origin.

Now that each processed image has a value for each characteristic (associated with and accuracy), it is possible to merge the characteristics from the three images (grayscaled, sharpened and blurred) and extract the best one: the value with the highest accuracy is chosen as the final value and assigned to the original photo ”Combined Results”.

Obtaining the ”Combined” result for the date variable differs from the other variables since the comparison of numbers does not return a percent accuracy but instead, a binary response. A voting system was implemented on which, each image votes with a day and a month. The ones with the most votes are chosen as the ”Combined Result”. Because live bivalves have an expiration date of less than one week, it is not mandatory to inform the user of the year of harvest. Therefore, the algorithm is programmed to assume the current year as the year of harvest, which automatically gives the ”harvest year” reading an accuracy of 100% (except for the first ten days of January, where appropriate precautions must be applied).

For each photo, the extracted data is validated against the results of the human analysis to calculate the accuracy of the algorithm.

C. Mobile Application & Back end

In this section, the developed communication system to which the name ”IPMAbiv” was given is described. It is composed of:

- Back end - A server developed using the Python Web Framework Django.
- Front end - A mobile application developed using the TypeScript mobile framework, IONIC.

Initially, the mobile application requirements are described, as well as its functionalities. Then, both the back end and the front end of the communication system are described in depth. The mobile application developed to simulate a testing environment is also described in this chapter, due to the similarities it has with the main mobile application. Finally, the results of an inquiry done to 10 users of the mobile application are presented.

*1) **Mobile Application Requirements***

On this section, the reader can find what is required of the application. The requirements of each type of user as well as other requirements, considered of a more general sort. The application functionalities are too presented as they are very much correlated with the requirements.

*2) **Users***

This section presents the different types of users and how each one can use the application:

- Producers - This type of user is typically a stationary one, considering nurseries are fixed in one place. By using the application, producers can have easier access to updated information.
- Fishermen - This type of user can benefit from the application by consulting it before engaging in the harvest of bivalves. This way, fishermen can be sure they are not engaging in illegal harvesting.
- Consumers - Easier access to the most basic information regarding bivalves brings great benefit to this type of user, who tends to be less educated on the topic and, as such, commit a higher number of mistakes that can later lead to severe consequences. Another useful feature for this user is the possibility of scanning commercial labels. Scanning commercial labels ensures the user that the bivalves that are for sale were harvested under legal circumstances.
- Commercial Activities - Considering this kind of business is monitored by food regulating authorities such as Autoridade de Segurança Alimentar e Económica (ASAE), the condition of the food provided is of the utmost importance. The possibility of scanning the commercial labels before purchase ensures the restaurant that everything is in order and that no risks are being taken.
- Food Regulating Authorities - Using *IPMAbiv* to assess if the labels found in commercial establishments comply with the law.
- Administrator - *IPMA* personnel, qualified to manage the species and zone databases through the back end server.

*3) **Functionalities***

On this section, the functional requirements of the system and their purpose are presented. The two of them are considered to be useful for all types of users.

Form-Based Query When desiring to know the condition of a specific bivalve (this applies best in a scenario of purchase), the user has to provide the species, the date, and place of harvest of the bivalve to a form-based query and the system will provide the corresponding information.

Label Identification When dealing with bivalves that have already been labeled, the user can scan the label with the OCR system that is set up on the application, discarding the need to manually insert all the information asked by the query.

4) Architecture

The architecture of the system is as shown in Figure 2. It can be divided into a back end: the server; database; web server and administrator interface; and a front end, which consists of the mobile application *IPMAbiv*, explained in Section II-C6.

The server, which was hosted at the *Instituto de Engenharia de Sistemas e Computadores*(INESC) facilities, is responsible for setting connecting all parts of the system. Through the web server it connects to the *IPMAbiv* using typical HTTP methods, such as *GET* and *POST*.

Due to the simplicity of the data involved in this study and considering the fact experts will not manage the system, SQLite was the library chosen to implement the database. This way it will be easily maintained and transferable.

The administrator interface explained in depth on Section **??**, is also hosted in the server, as one of the endpoints generated by the web server. However, only users with credentials can log in into this section of the system.

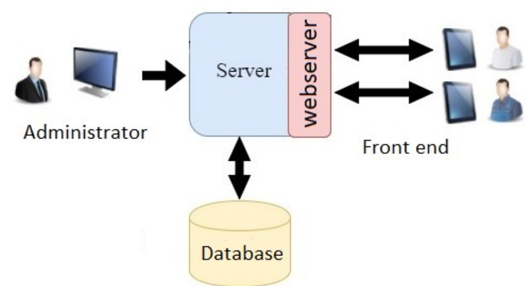

Fig. 2. The architecture of the developed system.

*5) **Back End***

The back end of the system is a *Django*[14] Project[14]. The *Django* framework was chosen due to the possibilities it offers regarding integration. /

*6) **Front End***

The front end of the system consists of a hybrid mobile application developed using the *IONIC* framework. It is compatible with both the *iOS* and *Android* operative systems.

*7) **Mobile Application***

The purpose of this mobile application is to inform its users on the bivalve state of consumption on a particular harvesting zone and a particular date. There are two distinct ways of querying the database for these results:(1) completing a form manually; (2) scanning a label.

Due to the slight variability observed in the nomenclature of zones and to ensure that a new zone/species is available on the mobile application as soon as it is inserted on the back end server, the mobile application will always query the server for these two lists (list of species, list of zones) upon initialization. This method is viable due to the low size of data involved in the process. However, this makes it impossible to use mobile application when there is no internet connection available.

An important issue is the fact that the server only registers the dates of updates. When a user queries the server for a given date, the server will respond with the results of the update that was released closest, yet prior, to that date. If the most recent update happened more than five days before the queried date, an error message is returned. This is done to protect the user from outdated information. The time interval of five days proved to be a good one since IPMA releases updated with a higher frequency.

Form Option
The form option consists of having the user manually insert all the necessary information to query the backend server regarding consumption safety. The necessary information is: Species, harvesting zone, and date of packaging.

After inserting this information, the user can click the ”Obter Resultados” button to query the server, as shown in Figure 3. The server response consists of a JSON file with a clear message stating if the bivalves are safe for consumption or not.

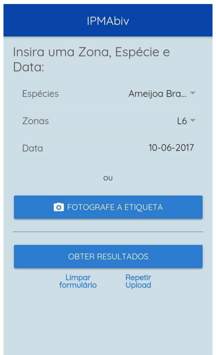

Fig. 3. The form method, ready to query the back end server.

The ”Positive Response” consists of a message stating that the bivalves in question were harvested legally. The ”Negative Response” advises the user not to consume the bivalves since they were not harvested legally.

To minimize human error, once a species/zone is selected, the zone/species field will only present the user the zones/species possible. This eliminates the possibility of querying the back end server for a combination ”specie, zone” that does not exist.

Label Reading Option
The label reading algorithm, described in depth on Section II-B, is responsible for scanning the text present on a label and returning the result (a species, a zone, and a date) to the user.

For the user to be able to confirm the results returned by the algorithm, the interface is also formatted as a form. This form is automatically filled with the results outputted by the algorithm. This allows the user to correct any misread information before querying the back end server, as shown in Figure (4).

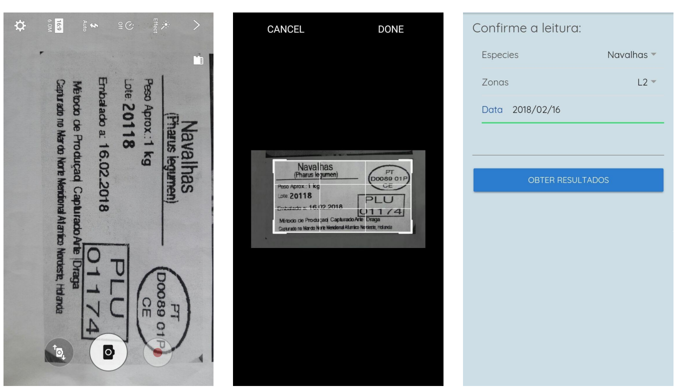

Fig. 4. The camera process (left); The cropping tool (centre); Confirmation step (right)

The user has two available options for the label reading algorithm:

1) Online label reading algorithm - The photo is uploaded to the back end server and given as the input to the algorithm.
2) Offline label reading algorithm - The algorithm is incorporated in the mobile application.

III. Results

The dataset is composed of 86 labels. These labels were relative to the following species: *Pharus Legumen*, *Ensis spp*, *Donax spp*, *Spisula solida*, Ruditapes philippinarum, *Ruditapes decussata*, *Cerastoderma edule*, *Crassostrea spp*, and *Mytilus spp*. These samples were originated from: Portugal (69.8%), Spain (25.6%), and the Netherlands (4.6%), more than 30.2% of the samples corresponded to bivalves harvested outside of Portugal.

A. Human Analysis

Table I presents the results of human analysis of the labels. 78.39% of the labels are complete, and provide all the information required by the European Union[13].

In all the incomplete labels, the missing information is the harvest zone. Table 1 further details this by presenting the number of labels with incomplete harvest zone and the percentage of labels from the FAO27 subzone, by country of origin.

TABLE I
Results shown by nationality.

Country of Origin	Sample Size	Complete Label [%]	Missing Harvest Zone [%]	FAO27 Subzone
Portugal	60	88.34	11.66	13.95
Spain	22	65.39	34.61	2.32
Netherlands	4	0	100	0
Overall Amount	86	78.39	23.25	16.27

By comparing the labels from different origin countries, it is possible to observe that products harvested in foreign countries lack harvesting location more than national products.

It is also relevant to present the number of products harvested in the FAO27 subzone, because these water and products are not tested for quality by the regulatory entities, meaning the toxicity levels of these bivalves are not known.

B. Algorithm Results

Table II shows the results obtained from each OCR procedure (greyscale, sharpened, and blurred) and from the final data extraction and aggregation.

TABLE II
Accuracy of the reading algorithm.

Reading	Greyscale [%]	Sharpened [%]	Blurred [%]	”Combined” [%]
Species	80.24	74.41	**84.88**	88.37
Date:				
Components	**64.34**	55.04	56.97	69.38
Whole	**41.86**	30.23	32.56	50.00
Harvesting Zone	70.93	68.60	**74.41**	82.55
Type of Harvest	52.32	**60.46**	54.65	73.00
Country	72.00	**82.55**	72.00	84.88
Overall Accuracy (w/Components)	64.41	63.72	64.57	78.43
Overall Accuracy (w/Whole)	63.47	63.25	63.70	75.76

Each filter overcomes different problems and issues in the original image, in different conditions different filters are best that the other two. Proximity formulas correlate and aggregate the partial results into one final ”combined” result. This combined result has the highest percent accuracy across all five variables, as shown in Table 2. It shows that the “date” field has the lowest accuracy. The ”Date components” row expresses the reading accuracy of the three components of a date (day, month and year), whereas the ”Date (full)” row expresses the accuracy of the reading of a complete date (the three components must be correct).

A closer review of the data found that a lack of a standardized date format incurred errors in the algorithm: different characters, such as ”/”, ”.”, alternatively, ”-” are used as separators; years are either printed with 2 or 4 digits (17 vs. 2017).

The center pie chart of Figure 2 shows the number of labels from which the date was (in)accurately read.

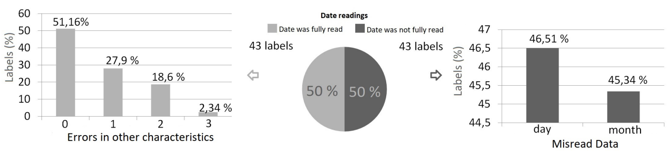

Fig. 5. Analysis of date recognition.

The left side of figure 2 regards the labels on which the date was accurately read. More than half of those readings had no errors on the remaining characteristics. These are promising results considering the variability found on the label structure.

The right side of figure 2 regards labels on which the date was not accurately read. It shows that there is no considerable difference between the percent error of days and months. These results support the idea that the issue in reading dates is on nonstandard formats used in labels.

1) Evaluation

As soon as it became possible, a prototype of *IPMAbiv* was given to a controlled group of 12 users. Six of them were male, and the remaining six female, all between the age of 20 and 50 years old. The users were asked to test the application and answer an inquiry. Whenever possible, the tests were conducted on the users’ mobile phone, which happened in approximately 50% of the cases.

The tests were performed in the following manner: Initially, the user would be given a bivalve commercial label. Then the user would be told to image he/she were at the supermarket and wished to know the consumption safety of the bivalves that corresponded to that label. The user would then initiate *IPMAbiv*. No instructions were given unless asked for, in order to assess the user-friendliness of *IPMAbiv* (in approximately 50% of the cases, no help was necessary). The Participant tested both functionalities, the form, and the automatic reading, and then answered the *Mobile Application Rating Scale* inquiry.

Mobile Application Rating Scale

The 12 testing subjects answered the Mobile Application Rating Scale (MARS) inquiry. The purpose of this inquiry was to understand how the developed mobile application would impact users by sharing with subjects that are considered to be potential users.

The inquiry is divided into six parts:

1) **Classification**
In this section, the Participant is requested to classify the application regarding **Focus** and **Main Purpose**. Each participant could select more than one of each:
Focus

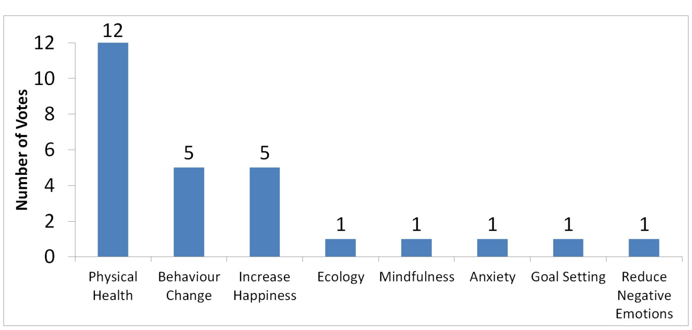

Fig. 6. Results of the ”Application Focus” section of the *MARS* inquiry.

Main Purpose

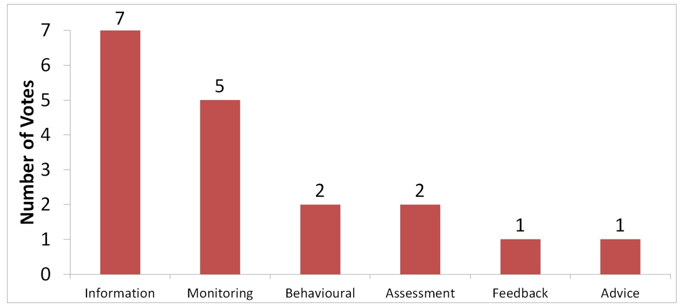

Fig. 7. Results of the ”Main Purpose” section of the *MARS* inquiry.

2) **Quality Rating**
In this section, the Participant is asked to answer questions regarding the quality of the application. It is divided into four different parts: Engagement, functionality, aesthetics, and information. Each possible answer was between 1 (Completely disagree) and 5 (Completely agree). The average of the result of each part is called ”Application Quality Mean Score”. Results are shown on Figure **??**.
To have a more detailed view of the results of each section of the *MARS* inquiry, Figure 8 shows results by section and gender.

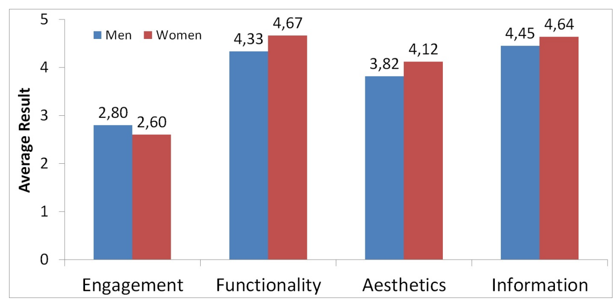

Fig. 8. Results by gender from the MARS inquiry.

3) **Subjective Quality**
On this section the participants were asked:
   - ”Would you recommend this app to people who might benefit from it?” Rate from 1 to 5 (1 - Not at all; 5 - Definitely) - 10 Participants selected ”5”, 2 Participants chose ”4” averaging 4.84.
   - ”How many time would you use this app in the next 12 months ()if it was relevant to you?” - 7 Participants selected ”3-10”, 2 Participants selected ”10-50” and 3 Participants selected ” >50”.
   - ”Would you pay for this app?” - 7 Participants selected ”No”, 4 Participants selected ”Maybe” and 1 Participant selected ”Yes”.
   - ”What is your overall star rating of the app?” - 11 Participants selected ”4 stars”, 1 Participant selected ”3 stars”, averaging 3.92 stars.

IV. Conclusion

The label-reading algorithm described in this study, if incorporated in a mobile application, could represent a new approach to accessing and sharing information regarding toxicity levels and consumer safety. This study identifies three primary sources of incompatibility with this approach. First, there is a significant percentage of incomplete labels that do not include the harvest zone. This piece of information is necessary to evaluate consumption safety and whether the bivalves were harvested legally.

Secondly, over 30% of the sampled labels come from bivalves that were harvested outside of Portugal. Since national regulating entities are not currently sharing their information, even if these foreign labels were complete and included their harvest zone, a mobile application that is not linked to all national regulating entities in Europe would not be able to assess foreign products' consumption safety accurately. The creation of a platform that unifies the information from all regulating entities in Europe and that provides support to users in a variety of languages would be a meaningful solution. This way, any user would have access to updated information regarding the toxicity levels of all harvesting zones in Europe, either directly through the platform or through a mobile application (with a label-reading algorithm) which would be linked to the platform.

Similarly, over 16% of the sampled labels correspond to bivalves harvested from the FAO27 subzone, which is outside of any national regulating entity's jurisdiction. The proposed mobile application cannot make an accurate assessment regarding the consumption safety of a product that has been harvested from the FAO27 subzone, as it is the responsibility of the fishery company to perform toxicity tests before selling.

If the labels on bivalve packaging were 100% complete, the label-reading algorithm described in this study performs with an overall accuracy of 76%. This indicates that the label-reading algorithm, if incorporated into a mobile application, is a valid approach to accessing updated information regarding bivalve harvesting and consumer safety. Multiple factors impede the algorithm's accuracy, including an unstandardized label structure, poor quality of label printing, poor label conditions, and/or reduced photo illumination and focus. Standardization of label structure would allow for a more accurate and efficient Machine Learning Algorithm (with the use of Convolutional Neural Networks). A non-variant label structure is imperative for optimal implementation and functioning of this more advanced algorithm. Ideally, the European Commission would require labels to be printed according to a standardized structure and the fields mandated by Regulation XX would be separated from any non required text.

The developed mobile application was well received by the test users, scoring particularly high, according to the *MARS* inquiry, on ”Functionality” and ”Information”. This indicates that the purpose to which the application was developed was achieved. Nevertheless, it scored lowest on the ”Engagement” section of the inquiry. This is important data for future developments. In conclusion, the proposed approach to sharing information on bivalve consumption safety can only work with the full support and of the European Commission through appropriate policy changes and adequate regulation. However, using a mobile application as a form-based querying tool to obtain updated information on bivalve consumption safety is an option with great potential.

Acknowledgments

We would like to thank Prof. João Paulo Costeira for his support and counseling regarding the image processing aspects of the label reading algorithm. We would also like to thank Dr. Mafalda Carapuço and Mr. Rui Oliveira from IPMA for references for the relevant legislation on food and bivalve safety.

References

[1] Centre for the Promotion of Imports from Developing Countries. CBI Product Factsheet: Bivalves in Europe. CBI Market Intelligence 2015

[2] Visciano P, Schirone M, Berti M, Milandri A, Tofalo R, Suzzi G. Marine Biotoxins: Occurrence, Toxicity, Regulatory Limits and Reference Methods. Front Microbiol 2016; 7; 1051.

[3] Regulation (EC) NO 854/2004 of the European Parliament and of the Council of 29 April 2004 (2004)

[4] Regulation (EC) No 853/2004 of the European Parliament and of the Council of 29 April 2004 (2004)

[5] Regulation (EC) No 1379/2013 of the European Parliament and of the Council, Annex I (2003) of 11 December 2013 (2013)

[6] World Health Organization (WHO). Classification of Digital Health Interventions v1.0. WHO, 2018.

[7] World Health Organization (WHO). mHealth: New horizons for health through mobile technologies: second global survey on eHealth. WHO, 2011.

[8] Centre for Food Safety, The Government of The Hong Kong Special Administrative Region, Hong Kong. Available from: http://www.cfs.gov.hk/english/food_safety_app/Food_Safety_App.html [Accessed 15 April 2018]

[9] European Commission. A pocket guide to the EU's new fish and aquaculture consumer labels. Publications Office of the European Union, 2014

[10] Szeliski R. Computer Vision: Algorithms and Applications. Springer-Verlag London Limited 2011

[11] Smith R, Podobny Z. Tesseract OCR. Available from: https://github.com/tesseract-ocr/tesseract [Accessed 15th April 2018]

[12] Smith R, Google Inc. An Overview of the Tesseract OCR Engine. In: IEEE Computer Society Proceedings of the 9th IEEE International Conference on Document Analysis and Recognition ICDAR; 23-26 2007 Sep 23-26; Curitiba, Paraná, Brazil, IEEE, 2007:629—633.

[13] European Commission. A pocket guide to the EU's new fish and aquaculture consumer labels. Publications Office of the European Union, 2014

[14] The Django Project[Online]. Available from: ”https://www.djangoproject.com/”