# COMPUTER-AIDED DISEASE DIAGNOSIS IN AQUACULTURE: CURRENT STATE AND PERSPECTIVES FOR THE FUTURE

JAYME G. A. BARBEDO*

Brazilian Agricultural Research Corporation (Embrapa)
Embrapa Agricultural Informatics
Av. André Tosello, 209 - Barão Geraldo - CP 6041 - 13083-886 - Campinas (SP) - Brazil
jayme.barbedo@embrapa.br

**ABSTRACT.** Automation of essential processes in agriculture is becoming widespread, especially when fast action is required. However, some processes that could greatly benefit from some degree of automation have such difficult characteristics, that even small improvements pose a great challenge. This is the case of fish disease diagnosis, a problem of great economic, social and ecological interest. Difficult problems like this often require a interdisciplinary approach to be tackled properly, as multifaceted issues can greatly benefit from the inclusion of different perspectives. In this context, this paper presents the most recent advances in research subjects such as expert systems applied to fish disease diagnosis, computer vision applied to aquaculture, and image-based disease diagnosis applied to agriculture, and discusses how those advances may be combined to support future developments towards more effective diagnosis tools. The paper finishes suggesting a possible solution to increase the degree of automation of fish disease diagnosis tools.

**KEYWORDS:** Fish diseases, Automation, Expert systems, Digital image processing, Aquaculture.

## 1 INTRODUCTION

The use of automation in agriculture is not new, and as the cost of technology plummets, its use becomes increasingly widespread. However, the technical challenges vary greatly depending on the problem to be tackled. While some simpler automation systems, like timed irrigation, have been solved long ago, others require a much higher level of sophistication in order to become feasible. This is the case of automated disease diagnosis, and particularly so for fish diseases.

Fast and reliable diagnosis of diseases is essential if the problem is to be handled properly, no matter if we are dealing with a person, an animal, or a plant. The best way to achieve that is by having a specialist readily available to assess the situation and to suggest some line of action. However, this is not always possible, especially in remote places. This is particularly true in the case of fisheries, as those are often located in places far from urban centers.

There are basically three approaches that can be used to overcome this problem. All three employ technological tools, but they possess different degrees of sophistication and automation, as shown below:

- Remote diagnosis system: in this type of approach, the person interested in obtaining a diagnosis sends information through some communication channel (usually internet or telephone), and such an information is analyzed by a human specialist, who sends a diagnosis (or a request for more information) back to that person. Because it has a low level of automation, this approach tends to be time demanding. On the other hand, the involvement of a human specialist improves the chances of a correct diagnosis. This option should be preferred when there are no reliable methods with a higher degree of automation, which still is the case for the vast majority of cases. This type of system was proposed by (LI et al., 2006).

- Expert systems: this kind of system tries to emulate the decision-making ability of a human expert (JACKSON, 1998). They employ a series of rules in order to reach a diagnosis. Human experts are essential during the development of the tool, but they usually do not have any involvement once the tool is put into use – in general, the user answers a number

*ARTIGO CONVIDADO: submetido em 29/01/2014; revisado em 17/02/2014. Artigo aceito sob recomendação do Editor-Chefe Prof. Dr. Ginalber L. O. Serra.

of questions about the symptoms of the disease, until the system has enough information to provide a diagnosis. This may be considered a semi-automatic diagnosis method, as the computer program depends on a number of human inputs in order to work properly. The higher degree of automation with respect to the previous approach means that it will be faster. Another advantage of this kind of system over the remote one is that it is usually cheaper to maintain. On the other hand, the quality of the diagnosis will be only as good as the inference engine and the knowledge base employed in the system, and design flaws may lead to error. Since the system depends on human input to reach a diagnosis, and such an input is usually provided by a non-expert, this may also give rise to incorrect outputs. Expert systems have been used for some time in fish disease diagnosis – a comprehensive list of this kind of system can be found in (ALAGAPPAN; KUMARAN, 2013). A discussion about this kind of tool, together with a brief description of some of the most relevant systems proposed so far, is presented in Section 2.

- Image-based automatic systems: this kind of system explores color and morphological characteristics of an image of the symptoms to infer the disease. Since there is no human interaction, this kind of approach tends to be very fast. Moreover, such a high degree of autonomy means that a system like this may be used to monitor vast areas without the need for humans to go to the field. This implies that a really reliable image-based system may be a much more advantageous option in comparison with the other two approaches. The problem is that achieving such a reliability is very challenging, being the main bottleneck for the dissemination of this kind of system. Despite the challenges involved, automatic diagnosis systems are becoming common in several areas of agriculture and medicine, but the difficulties involved in aquaculture have prevented any method with focus on fish to be proposed so far. To the author's knowledge, the only work that has some relation with this kind of approach was proposed by (PARK; OH; HAN, 2007), which has the objective of providing a diagnosis by identifying pathogens in microscopic images. A deeper discussion about image-based automatic diagnosis systems is presented in Section 4.

Some disease diagnosis problems have already achieved a high degree of automation. In the case of fish diseases, however, the successes have been limited to a handful of expert systems, most of them designed to deal with a single species and a limited set of diseases. This is due to the very difficult characteristics surrounding the problem, such as the large variety of species, the symptoms heterogeneity, the difficulties involved in gathering data in the field, the fact that fish live underwater, among others. The use of digital imaging is particularly challenging: according to (ZION, 2012), "the sensing technology has to overcome limited visibility, temporal and spatial variations in lighting, varying distances and relative orientations between cameras and objects, motion and density of the monitored targets, and even lack of physical stability." Those challenges suggest that an interdisciplinary approach may be needed in order to increase the degree of automation in diagnosis. It is important to emphasize that a fully automated system for disease diagnosis in fish would probably require very sophisticated artificial intelligence, capable of autonomously performing analysis based on images, biopsy samples, water samples, etc. Such a technology is far beyond what is available today.

In this context, the objective of this manuscript is to discuss some possible ways to gradually increase the automation of the process, by analyzing how the most current advances in areas such as expert systems for disease diagnosis in aquaculture, computer vision applied to various aquaculture problems, and image-based disease diagnosis in agriculture (especially plants), could be combined into a more effective diagnosis tool.

## 2 EXPERT SYSTEMS IN AQUACULTURE

Expert systems try to emulate the way a human expert makes a decision. They are normally composed of two modules: the knowledge base, which stores facts about the problem at hand, and the inference engine, which reasons about those facts and uses rules and other forms of logic to deduce new facts or highlight inconsistencies (HAYES-ROTH; WATERMAN; LENAT, 1983). According to (ZELDIS; PRESCOTT, 2000), there are five major problems faced by this kind of system:

- No disease exhibits all the signs described in the literature.

- There is a time progression for every disease, that is, the symptoms will differ depending on the stage of the disease.

- There is a lack of standardized terminology.

- More than one disease can be present at the same time.

- The human inputs are not always completely reliable.

Thus, the main challenge that this kind of system has to face is to take all those problems into consideration, and create mechanisms that can minimize their effects.

Several expert systems have been proposed so far to infer fish diseases. (ALAGAPPAN; KUMARAN, 2013) provide a list with a very brief description of 30 of those systems. This section provides a more in-depth analysis of some of the most relevant of such systems.

One of the earliest expert systems dedicated to fish disease diagnosis was proposed by (STEINEBACH; PEREIRA, 1989). Their system focused on identifying bacterial diseases in salmonids growing in sea water, and used rules based on factors like food (composition, storage time, amount), environmental factors (geography, water parameters, fish data, farming data), external signs (behavior and appearance of the fish) and internal signs (kidneys, heart, liver, internal fluids, etc.). Due to the uncertainty that is usually present in any diagnosis attempt, the system provides not only a list of possible diseases that may match the answers, but also their likelihood (one of six possibilities, from "impossible" to "sure").

The system proposed by (ZELDIS; PRESCOTT, 2000), the so-called Fish-Vet, combines fuzzy, rule-based and statistical elements to provide more reliable results. The diagnosis process is iterative: the user first chooses species, water type and all signs observed in the affected fish; then, a diagnosis is run and a list of potential candidates, with the respective likelihoods, is presented; the signs expected for each disease in the list is presented to the user, who must chose those actually observed; then, a new diagnosis is performed. According to the authors, after two or three iterations, the first disease in the list is most likely the correct one.

One of the first web-based systems, the so-called Fish-Expert, was proposed by (LI; FU; DUAN, 2002). According to the authors, their system has the following characteristics: can be easily accessed via web; it is able to mimic the real practice of fish disease diagnosis by matching combinations of symptoms, microscopic examinations and water quality; can identify 126 fish diseases amongst nine species of freshwater fish; farmers, technicians and experts can input and update the information contained in the system's database; uses a multimedia user interface for more effective communication; gathers users feedback for continuous improvement; provides general information support tools. The authors conclude that, as any expert system, this one is not perfect, and that it would be useful to couple it with a tele-diagnosis system, where a human expert could provide additional insights.

Eel disease diagnosis is the subject of SEDPA, an expert system proposed by (GUTIERREZ-ESTRADA et al., 2005). Their system has several modules, but only the inference engine is described. The inference engine has three main parts: an augmented transition network (ATN), which allows the inference engine to access the data stored in the domain databases, starting from the information in natural language entered by the user; a fuzzy logic controller, which allows a point to be described as a function of its membership in different sets; and a system of uncertainty management based on the Dempster–Shafer theory (DST), which deals with the issues of representation, propagation and combination of uncertain information. The authors reported an accuracy close to 70%, and a performance close to that achieved by human experts.

The system proposed by (ZHU; LI, 2008) uses a different principle from most of its predecessors, as it is built upon the concept of case-based reasoning. A CBR system extracts information from cases solved previously, instead of relying on general domain knowledge, as is the case of rule-based systems. One of the advantages of this kind of approach is that the system can be continuously updated by incorporating the information contained in the new cases. The system employs a two-step case indexing model, which was designed to allow a retrieval procedure capable of finding similar cases in a fast and reliable way.

In (XIAOSHUAN et al., 2009), the application of evolutionary prototyping model in the development of an intelligent decision support system for fish disease/health management is described. The word "evolutionary" is used here in the sense that the system was developed in four stages, and that each stage provided the subsidies needed to develop the next one. The first stage generated the first standalone version of the program, which included information about six traditional Chinese fish species, and incorporated the concept of forward reasoning. In the second stage, the web-based system was implemented and backward and hybrid reasoning were incorporated. The third stage included early warning knowledge and models, non-traditional and noble fish species, and incorporated the probabilistic knowledge representation. Finally, the fourth stage included a call-center based system and the evaluation of drug withdrawal periods.

The main novelty of the system proposed by (XU; ZHANG; TAN, 2013) is that it explores Short Message Services (SMS) to aid people that do not have access to computers or to the internet. The system itself is based on Bayesian networks, which has the challenge to extract the information given by the user through the SMS and provide a diagnosis, no matter how incomplete is the information contained in the message. As expected, the authors report some errors due to the impossibility of a deeper interaction with the users.

(NAN et al., 2009) proposed an indirect approach for the problem of fish disease detection. Instead of focusing on the fishes themselves, their early warning expert system uses information about the quality of the water to decide whether it should issue a disease alert or not. The system has four modules: an assessment module for pond water eutrophication, which analyzes if oxygen conditions are suitable; an assessment module for pollutant potential biological toxicity, which can affect fish health; a factor forecast module, which tries to dynamically forecast short-term changes in primary factors of the pond; and a fish disease/health early warning module, which is responsible for deciding if an alert should be issued. The knowledge base used in the system was built from farming log records, laboratory experiments, experts questionnaires, and literature review.

Finally, the system proposed by (LOU; CHEN; YE, 2007) presents a hybrid approach that combines elements from both conventional case-based expert systems and image-based automatic systems. In the proposed system, the user submits images captured using digital cameras or microscopes. Then, a number of features are extracted and compared with the built-in image database; the reference images whose feature vectors are closer to the submitted image are retrieved and presented to the user, which then gives a feedback to the system by selecting those images that seem to correspond to the disease at hand. The results are then submitted to the case-based reasoning module, which compares the results with those of previously solved cases, using the K-nearest neighbors technique.

As can be seen, a variety of expert systems can be found in the literature. Some conclusions may be drawn about the types of approaches:

- Rule-based: the advantage of this kind of approach is that the knowledge accumulated by human experts can be relatively easily translated into the code using simple programming elements. Also, they can be easily understood by anyone with minimum knowledge on the subject.

- Case-based: this kind of approach is less intuitive in terms of the rules governing the decisions, but they have the ability of learning from previous experiences, so the system becomes better as more cases are processed.

- Hybrids: the shortcomings attached to any particular approach may often be overcome by combining them with other techniques with complementary characteristics. For example, rule and case-based approaches can be combined to provide a well defined set of rules that can be continually improved by previous experiences. Also, expert systems may be combined with digital image processing techniques, so visual cues may improve the inference process.

Another important remark is that relevant information may come from different sources, like visible symptoms, fish behavior, water variables, microscopic analysis, among others. One of the main challenges in developing an expert system is to find the most suitable way of combining such a heterogeneous set of data into a reliable diagnosis tool.

Finally, it can be observed that most systems are developed focusing on a very specific problem. This makes it possible to focus on the peculiarities of such a problem, potentially leading to more accurate results. On the other hand, generic systems, which do not focus on particular species and diseases, make it possible for the user to apply them immediately, without the need for a study to determine if such a tool is appropriate for that particular situation. However, this added flexibility often comes associated with diminished accuracy. This flexibility versus accuracy problem is common to many fields, and must always be taken into account.

## 3 COMPUTER VISION IN AQUACULTURE

Although there are no automatic methods for fish disease diagnosis based on digital images proposed so far, computer vision and digital image processing techniques have been successfully applied to a number of aquaculture problems. A comprehensive overview on the subject can be found in the article by (ZION, 2012). Because of that, this section will present only a brief overview of the advances achieved in some selected areas of aquaculture.

Counting objects is among the main applications of digital signal processing. In aquaculture, having a good estimate for the number of specimens at various growth stages is important for managing the fishery. Computer vision has been employed for counting eggs, larvae, fry and fish. Most counters are available as commercial packages, and some special settings are necessary for the counting to be possible: either the fish have to be sampled and contained in a space of known volume, or they are forced through some kind of pipe, conveyer or chute, where they are counted as they pass in front of the camera. According to (ZION, 2012), there is still a need for tools capable of counting larvae accurately, what is a great challenge considering that their numbers normally range from thousands to millions in a single tank or pond.

Measuring fish is also a problem that is being regularly tackled by means of computer vision. The information of interest is the weight, as this will have direct influence on the revenues. However, a direct calculation of the fish weight using digital images is not possible. Instead, other measurements are performed and inserted into equations carefully created to yield accurate estimates for the weight. The inputs to those equations include length (STRACHAN, 1993), side-view area (BALABAN et al., 2010), top-view area (HUFSCHMIED; FANKHAUSER; PUGOVKIN, 2011), and other size-related features (ODONE; TRUCCO; VERRI, 2001). Most methods require the fish to be at a certain position to work properly, which often means either killing, sedating, or guiding them through some passage, which may introduce bias. In order to overcome this limitation, some methods use the concepts of stereoscopic vision to add the dimension of depth to the measurements (BEDDOW; ROSS; MARCHANT, 1996; TILLETT; MCFARLANE; LINES, 2000; COSTA et al., 2009; TORISAWA et al., 2011).

Gender identification is also an important issue because, for many species, gender plays an important role in factors like growth rate, breeding strategies, and quality of the product (ZION, 2012). The most common approach in this case is to extract morphometric features (size and shape) as indicators of the gender (MERZ; MERZ, 2004),(ODONE; TRUCCO; VERRI, 2001). According to (ZION, 2012), the problem of gender identification has some satisfactory solutions, and the challenge now is to find a way to physically separate the fish into groups. The exact same challenge is present in species identification, which is described in the next paragraph.

Species identification is important when multiple species of fish are grown in the same environment. The vast majority of the proposals on the subject use morphometric and color features to discriminate between species. Some proposals use images of dead fish to perform the discrimination (STRACHAN; KELL, 1995; ZION; SHKLYAR; KARPLUS, 2000; WHITE; SVELLINGEN; STRACHAN, 2006), while others tackle the more challenging task of discriminating fish that are alive and mobile (CARDIN, 2000; CADIEUX; LALONDE; MICHAUD, 2000; AGUZZI et al., 2009).

Finally, automatic welfare monitoring has also received considerable attention, especially due to animal welfare awareness, and also due to potential production losses that may be caused by animal stress. Such a monitoring is usually done by observing changes in external behavior, such as abnormal movement patterns (XU et al., 2006; RODRIGUEZ et al., 2011; PINKIEWICZ; PURSER; WILLIAMS, 2011) and unusual feeding behavior (FOSTER; ITO; WARD, 1995; ISRAELI; KIMMEL, 1996). Due to the challenges faced by underwater surveillance, this kind of system works better in fisheries with some degree of control over variables like lighting and feeding position, and also in reservoirs with shallower water depths (ZION, 2012).

As can be seen, many areas of aquaculture have experienced considerable advances towards automation of their processes. However, the problem of disease diagnosis poses some particular challenges: the signs that characterize the diseases may be very faint, or may not even manifest visually. Also, the symptoms may vary greatly depending on the progress of the disease. Although the conditions for imaging plants are usually more friendly, the automatic disease diagnosis in plants actually share many of those challenges that plague its aquaculture counterpart. As such, analyzing how the problem has been tackled for plants may reveal some clues and inspire new solutions for the automation of fish disease diagnosis, which is exactly the objective of the next section.

## 4 IMAGE-BASED PLANT DISEASE DIAGNOSIS

Almost all automatic tools for disease diagnosis are founded on the principles of pattern recognition. In general, this kind of tool performs three main tasks: segmentation, which aims to separate the lesions from the rest of the image; feature extraction, which aims to extract as much information about the region of interest (usually lesions) as possible; classification, whose function is to combine the information present in the features into a reliable identification of the disease. Each one of those tasks is important. However, the part of segmentation already has some standard solutions, so it would not be useful to revisit those here. Also, feature extraction is highly dependent on the characteristics of the lesions and respective hosts, and since plants are so different from fish, most features are likely inappropriate. Additionally, despite the huge number of different features proposed in the literature, they are all variations of some core features that aim to extract information about color, morphology and texture of the lesions, so this is also well established.

The tools used for classification also are, in general, quite standard. However, many of those tools accept different settings, and they may be combined with other kinds of techniques in order to yield better results. Because of that, these were chosen as the reference in the presentation of the methods of image-based plant disease diagnosis. It is important to remark that this subject has already received a comprehensive review (BARBEDO, 2013). Thus, rather than presenting all the details of the techniques proposed in the literature, this section will concentrate on the main general aspects that may be extended to the problem of fish disease diagnosis. More details about the works cited in this section can be found in (BARBEDO, 2013).

### 4.1 Neural Networks

There are many types of neural networks. Arguably, the most commonly used is the Multilayer Perceptron (MLP), which is a feedforward neural network containing one or more layers of interconnected artificial neurons, each having a non-linear activation function associated (HAYKIN, 1998). The training of this kind of network is done using the so-called back-propagation algorithm. The main idea underlying the use of this kind of tool is that, if fed with an appropriate set of features, it will output the most likely classification, which is given by the label of the output neuron with the highest activation value.

Several systems using MLP neural networks (MLPNN) have been proposed so far. One of the early ones was proposed by (HETZRONI et al., 1994), and its most prominent distinctive characteristic was the combination of the neural network with statistical classifiers. A MLPNN with one hidden layer is fed with a number of texture features in the system proposed by (HUANG, 2007), whose main distinction is the sophisticated segmentation strategy. (SANYAL et al., 2007) proposed a system using two neural networks, one for color features and other for texture features, and then combines their outputs into the final disease classification. The system proposed by (Al Bashish; BRAIK; BANI-AHMAD, 2010) uses an MLP with a large number of hidden layers (ten) in order to combine color and texture features into a classification.

Radial basis function (RBF) neural networks are also commonly applied to the problem of classification. They are actually closely related to MLPs. The main difference is that RBF networks use radial basis activation functions for hidden neurons. In practice, this implies that MLPNN's activation functions process the inner product of the input vector and the synaptic weight vector of that neuron, while in RBFs the activation function processes the Euclidean norm between the input vector and the center of that neuron (ZHANG; GUPTA, 2000). MLPs have better generalization capabilities, but RBFs learn faster. (PYDIPATI; BURKS; LEE, 2005) used a RBF neural network to process 39 texture features, and compared the results with those obtained using a Mahalanobis minimum distance classifier.

Finally, (WANG et al., 2012) tested four different types of neural networks to process 50 color, shape and texture features: Multilayer Perceptron, Radial Basis Function, Generalized Regression, and Probabilistic, with no significant performance difference among them.

### 4.2 Support Vector Machines

This kind of classifier tries to find the hyperplane that best separates two classes in a given feature space (CORTES; VAPNIK, 1995). Although SVMs have been created to separate only two classes, they can be easily modified to deal with multiple classes.

The system proposed by (MEUNKAEWJINDA et al., 2008) uses a multiclass SVM to discriminate between three diseases; the authors also apply other machine learning techniques in different stages of the algorithm (prior to classification), including MLP neural networks, self-organizing maps, genetic algorithms and 2-class SVM. (YOUWEN; LI; YAN, 2008) proposed an SVM-based algorithm capable of identifying two diseases from color, shape and texture features; the authors claim that the SVM algorithm performs better than MLPNN-based algorithms. The system proposed by (YAO et al., 2009) follows those same principle, that is, they isolate the lesions, extract a number of color, shape and texture features, and then apply a SVM to perform the classification. (CAMARGO; SMITH, 2009) use the one-against-one method (HSU; LIN, 2002) to allow the SBM to deal with multiple diseases; they conclude that the combination of a SVM with texture features lead to the best disease discrimination. Finally, (JIAN; WEI, 2010) feed a variety of color, shape and texture features to a SVM using Radial Basis Function (RBF) as kernel, in order to discriminate between three different diseases.

### 4.3 Fuzzy Classifier

Fuzzy logic is a theory founded on approximate reasoning (NOVAK; PERFILIEVA; MOCKOR, 1999), that is, it is based on intermediate and relative classifications, rather than absolute ones. This is particularly useful in disease diagnosis:

- It is very common for different diseases to share similar symptoms, so there may be some degree of uncertainty when a diagnosis is performed uniquely based on those visible signs. With the use of fuzzy logic, it is possible to provide a probabilistic diagnosis, such as "there is a 30% probability that disease A is present, and 70% probability that disease B is present", so that the uncertainty is taken into account and quantified.

- More than one disease may be present in a given sample, so a relative classification indicating the relative degree of severity for those diseases, using a similar approach as the example presented in the previous item, may be very useful.

The use of fuzzy logic is more common in specialist systems, but there have been some proposals focused on automatic disease diagnosis. (HAIRUDDIN; TAHIR; BAKI, 2011) proposed a method that extracts a number of color and texture features, and submits them to a fuzzy classifier that provides a relative classification considering four types of nutritional deficiencies. The system proposed by (XU et al., 2011) also deals with nutritional deficiencies, uses a genetic algorithm to select and recombine a number of color features, and applies a fuzzy K-nearest neighbor classifier, which is responsible for the final identification.

### 4.4 Rule-Based

Sometimes, creating simple rules based on the experience of experts instead of employing complicated computational intelligence tools may lead to better results. The downside of this kind of approach is that it will only work for the problem for which it was designed, that is, its generalization capabilities are very weak. Those rules can be organized as a chain (KURNIAWATI et al., 2009), or they can be created separately for each feature, whose results are compiled in a voting system (ZHANG; MENG, 2011).

### 4.5 Others

Some systems proposed in the literature adopt approaches that are not largely used, but may be very effective for some problems.

Some methods rely solely on color information. In general, those systems compare the colors found in the leaf under analysis with a previously built dictionary containing color information typical of healthy and diseased tissue. This approach was adopted by (WIWART et al., 2009) and (PUGOY; MARIANO, 2011).

In the system proposed by (PHADIKAR; SIL, 2008), pixel values (gray scale) of selected areas of the image are directly fed to a self organizing map (SOM), which performs the final classification.

(PYDIPATI; BURKS; LEE, 2006) proposed a method in which 39 features are extracted, a redundancy procedure is applied in order to keep on the relevant features, and discriminant analysis is applied to reveal the final classification.

## 5 COMPUTER-AIDED FISH DISEASE DIAGNOSIS: CURRENT PROSPECT

With the development of technology, the automation of several activities and processes in the daily life is now within reach. However, the degree of automation that can be achieved at any given moment depends not only on the current technological advances, but also on the characteristics of the problem at hand. Problems that are well defined and have little variation, such as car assembling, are much more likely to be fully automated than problems that are ill-posed and have lots of variation, which clearly is the case of fish disease diagnosis. Given the current technological background, a fully automated method of diagnosis is unfeasible even for plants, which have more tractable characteristics than fish.

Currently, the reference in computer-aided diagnosis of fish diseases are the expert systems. They introduce some degree of automation, as the role that would be performed by a human specialist is emulated by a computer program. However, these systems also have a strong manual component, as the users have to answer a number of questions about the problem at hand. This may lead to three main problems: depending on the number of questions and rules, the process may be time consuming; the subjectivity involved in the process may lead to error; even small inconsistencies in the rules may lead to wrong diagnosis. This last problem becomes the more serious as the system becomes more comprehensive, because additional generality makes it more difficult for the rules to correctly capture and emulate the classification process. One possible solution for those problems is to introduce some kind of automatic method capable of refining the problem by eliminating unlikely possibilities before the expert system is activated. This may greatly improve the performance of the system as a whole, since it can now apply only the questions that fit the universe of possibilities determined by the automatic part.

In this context, a seemingly feasible line of action would be taking, as starting point, a technology that is reasonably established and has some degree of automation, like expert systems, and then slowly introduce more autonomous tools. Arguably, the best candidate to lead this shift towards more automated systems would be the computer vision.

Previous sections have mentioned several different techniques for pattern and object recognition. The vast majority of the proposals adopt very conventional approaches in terms of segmentation, feature extraction, identification, and ensemble of those parts. Also, comparing the results reported by the authors, no machine learning tool has a clear advantage over the others. This relative stagnation and lack of breakthroughs may be one of the reasons why no image-based tool for fish disease diagnosis has been proposed so far. So, is it really worth pursuing a solution which has computer vision as one of its main pillars? Despite the challenges involved, the potential benefits of a highly automated system certainly suggest so. A possible way to achieve this objective is suggested in the following, under the form of a "computer vision-expert system" hybrid solution.

## 6 INCREASING AUTOMATION: A HYBRID SOLUTION

As commented before, the hybrid solution suggested here would use computer vision to limit the universe of possibilities to be considered by the expert system. Also discussed before is the fact that expert systems are relatively well established for fish disease diagnosis, so their design details have been already extensively explored in the literature. Because of that, this section will focus on the computer vision part of the solution. In particular, emphasis is given to the conditions that should be met in order to make the system robust and useful.

### 6.1 Image Database

A major bottleneck present in the development of almost all computer vision-based techniques is the lack of a suitable image database. This situation is motivated by the huge challenges involved in the task of building a good database:

- In general, the images are collected in the field. The samples to be imaged may be located in remote areas, and the team visiting those areas have to carry some device capable of capturing images (digital camera, scanner, cell phone, etc.). In the case of fish, the problem is even greater. First, diseased animals have to be detected, which may be very difficult depending on the water turbidity. Then, those animals need to be removed from the water in order to be properly imaged. Considering that a proper image database has to have from hundreds to thousands of samples, this can be a very strenuous and time demanding task. Underwater cameras might be used to capture the images, however the distortions caused by the water, as well its turbidity, may render the images useless.

- With a few exceptions, a computer vision-based technique will only be able to deal properly with those cases for which it was trained. In the specific case of diseases, such a database should contain all variations of the symptoms, including: different stages of development, different degrees of severity, color and morphology variations caused by external factors, and even superposition of symptoms caused by different diseases. Again, this problem is more challenging in the case of fish, because visually detecting those variations in mobile animals that are under water is almost impossible, thus it may be necessary to remove several animals, in a trial and error effort.

- The conditions under which the images are captured may have a great impact on their quality. Lighting conditions are of particular importance. This factor is influenced by weather conditions (sunny, overcast), by whether the image is captured indoors or outdoors, by the angle of incidence of the light, and by the reflectance of the fish and background. The angle of capture is also an important factor to be considered. Because of that, it is highly advisable to define some minimum protocols that should be followed during the process of image capture.

- It is desirable that the images used in the development of a computer vision technique be as high quality as possible. However, if such a technique is to be used in practice, it may have to deal with low quality images regularly. Therefore, the database should include high quality images for the first stages of development, and low and medium quality images for testing robustness and for pointing out the aspects that need improvement.

### 6.2 Uncertainty Treatment

Uncertainty is a perennial factor in disease diagnosis, no matter if this is done manually or automatically. Thus, if even a human expert cannot be expected to give an unequivocal diagnosis in 100% of the cases, it would not be realistic to think that a machine would be able to do so with the limited amount of information present in an image. One way of dealing with this uncertainty is by employing the principles underlying the fuzzy logic. Under this philosophy, the computer-vision based part of the system would analyze the symptoms, and assign a probability to all diseases that could possibly be associated to those symptoms. This information could then be fed to the expert system, which would immediately remove all questions and rules attached to diseases with very low probabilities of occurrence. As a result, the expert system is much less likely to misinterpret the user's responses and go in the wrong direction.

### 6.3 Module Integration

As commented in Section 6.2, the results provided by computer vision-based part of the program can be used as inputs for the expert system, which then would trim the rule branches accordingly. However, the integration between these two parts is not that trivial. The rules of an expert system may be quite intricate, so eliminating all the rules that have something to do with the diseases deemed unlikely by the computer vision-based part of the algorithm may discard venues that could actually lead to the correct answer. In order to avoid this kind of situation, the best option is to create the rules of the system already having in mind the kind of information that will be provided by the automatic module, so that rule selection becomes more natural.

### 6.4 Discussion

It is very important to emphasize that any system for disease diagnosis, independently of the degree of automation, will provide a diagnosis that will be as good as the information it receives. More importantly, even if the information fed to the system is absolutely correct and unequivocal, it may not be enough to provide a diagnosis with 100% certainty. Instead, the output of the system must always be viewed as an educated guess; in fact, this is often the case even with human experts, because in many cases a final diagnosis can only be reached after careful clinical analysis. This does not diminish the positive impacts of tools like these, because a good disease diagnosis system will provide correct answers most of the time. As a result, although those diagnoses cannot be taken as definitive, at the very least they allow people to make arrangements in advance, so when a definitive diagnosis is available, they are already prepared to immediately begin the necessary actions. Those diagnoses may also guide precautionary actions, like isolating animals that may be contagious. In conclusion, time is of the essence if someone wants to control the propagation of a disease and cut losses, so a tool capable of speeding up time responses is of invaluable importance for any producer.

## 7 CONCLUSIONS

This paper presented an overview about the use of automation in disease diagnosis and, particularly, in aquaculture. It was shown that the challenges involved in using computer vision to detect and identify diseases in fish are difficult to overcome, because the conditions involved in the capture of the images and in the development of the methods are considerably more complex than in other areas of agriculture. Because of that, the use of computer vision in aquaculture, and particularly in disease diagnosis, has evolved slowly. Despite the difficulties, it was shown that automation levels may be increased by combining techniques that have experienced some success, like expert systems, with more autonomous tools, especially those based in digital image processing and computer vision.

Efforts are currently underway to put the ideas raised throughout the paper into practice. In this project, fish pathologists and information technology experts will join forces to overcome the barriers that have hampered the development of better fish disease diagnosis tools. The species to be considered in these initial efforts will be the Tambaqui ($Colossoma\ macropomum$), but other species may be included as the research evolves.

**ACKNOWLEDGEMENTS**

The author would like to thank São Paulo Research Foundation (FAPESP) for funding, under grant proc. $2013/06884-8$.

**REFERENCES**

AGUZZI, J. et al. A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna. *Sensors*, v. 9, n. 11, p. 8438–8455, November 2009.

Al Bashish, D.; BRAIK, M.; BANI-AHMAD, S. A framework for detection and classification of plant leaf and stem diseases. In: *2010 International Conference on Signal and Image Processing*. [S.l.: s.n.], 2010. p. 113–118.

ALAGAPPAN, M.; KUMARAN, M. Application of expert systems in fisheries sector – A review. *Research Journal of Animal, Veterinary and Fishery Sciences*, v. 1, n. 8, p. 19–30, September 2013.

BALABAN, M. et al. Prediction of the weight of alaskan pollock using image analysis. *Journal of Food Science*, v. 75, n. 8, p. E552–E556, October 2010.

BARBEDO, J. Digital image processing techniques for detecting, quantifying and classifying plant diseases. *SpringerPlus*, v. 2, n. 660, December 2013.

BEDDOW, T.; ROSS, L.; MARCHANT, J. Predicting salmon biomass remotely using a digital stereo-imaging technique. *Aquaculture*, v. 146, n. 3-4, p. 189–203, November 1996.

CADIEUX, S.; LALONDE, F.; MICHAUD, F. Intelligent system for automated fish sorting and counting. In: *Proc. International Conference on Intelligent Robots and Systems (IROS)*. [S.l.: s.n.], 2000. p. 1279–1284.

CAMARGO, A.; SMITH, J. S. Image pattern classification for the identification of disease causing agents in plants. *Computers and Electronics in Agriculture*, v. 66, n. 2, p. 121–125, maio 2009.

CARDIN, S. Advances in morphometric identification of fishery stocks. *Reviews in Fish Biology and Fisheries*, v. 10, n. 1, p. 91–112, March 2000.

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, September 1995.

COSTA, C. et al. A dual camera system for counting and sizing northern bluefin tuna ($Thunnus\ thynnus$; Linnaeus, 1758) stock, during transfer to aquaculture cages, with a semi automatic artificial neural network tool. *Aquaculture*, v. 291, n. 3-4, p. 161–167, June 2009.

FOSTER, M.; ITO, R. P. M.; WARD, R. Detection and counting of uneaten food pellets in a sea cage using image-analysis. *Aquacultural Engineering*, v. 14, n. 3, p. 251–269, 1995.

GUTIERREZ-ESTRADA, J. et al. SEDPA, an expert system for disease diagnosis in eel rearing systems.

*Aquacultural Engineering*, v. 33, n. 2, p. 110–125, August 2005.

HAIRUDDIN, M. A.; TAHIR, N. M.; BAKI, S. R. S. Overview of image processing approach for nutrient deficiencies detection in Elaeis Guineensis. In: *2011 IEEE International Conference on System Engineering and Technology*. [S.l.: s.n.], 2011. p. 116–120.

HAYES-ROTH, F.; WATERMAN, D.; LENAT, D. *Building Expert Systems*. [S.l.]: Addison-Wesley, 1983.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. [S.l.]: Prentice Hall, 1998.

HETZRONI, A. et al. Machine vision monitoring of plant health. *Advances in space research the official journal of the Committee on Space Research COSPAR*, v. 14, n. 11, p. 203–212, 1994.

HSU, C.-W.; LIN, C.-J. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, v. 13, p. 415–425, 2002.

HUANG, K.-Y. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Computers and Electronics in Agriculture*, v. 57, n. 1, p. 3–11, maio 2007.

HUFSCHMIED, P.; FANKHAUSER, T.; PUGOVKIN, D. Automatic stress-free sorting of sturgeons inside culture tanks using image processing. *Journal of Applied Ichthyology*, v. 27, n. 2, p. 622–626, April 2011.

ISRAELI, D.; KIMMEL, E. Monitoring the behavior of hypoxia-stressed $Carassius\ auratus$ using computer vision. *Aquacultural Engineering*, v. 15, n. 6, p. 423–440, Novemberl 1996.

JACKSON, P. *Introduction To Expert Systems*. [S.l.]: Addison-Wesley, 1998.

JIAN, Z.; WEI, Z. Support vector machine for recognition of cucumber leaf diseases. In: *2010 2nd International Conference on Advanced Computer Control*. [S.l.: s.n.], 2010. p. 264–266.

KURNIAWATI, N. N. et al. Investigation on image processing techniques for diagnosing paddy diseases. In: *2009 International Conference of Soft Computing and Pattern Recognition*. [S.l.: s.n.], 2009. p. 272–277.

LI, D.; FU, Z.; DUAN, Y. Fish-Expert: a web-based expert system for fish disease diagnosis. *Expert Systems with Applications*, v. 23, n. 3, p. 311–320, October 2002.

LI, D. et al. Toward developing a tele-diagnosis system on fish disease. In: BRAMER, M. (Ed.). *Artificial Intelligence in Theory and Practice*. [S.l.]: Springer US, 2006, (IFIP International Federation for Information Processing, v. 217). p. 445–454.

LOU, D.; CHEN, M.; YE, J. Study on a fish disease case reasoning system based on image retrieval. *New Zealand Journal of Agricultural Research*, v. 50, n. 5, p. 887–893, 2007.

MERZ, J.; MERZ, W. Morphological features used to identify chinook salmon sex during fish passage. *The Southwestern Naturalist*, v. 49, n. 2, p. 197–202, June 2004.

MEUNKAEWJINDA, A. et al. Grape leaf disease detection from color imagery using hybrid intelligent system. In: *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*. Krabi: [s.n.], 2008. p. 513–516.

NAN, L. et al. Developing a knowledge-based early warning system for fish disease/health via water quality management. *Expert Systems with Applications*, v. 36, n. 3, p. 6500–6511, April 2009.

NOVAK, V.; PERFILIEVA, I.; MOCKOR, J. *Mathematical Principles of Fuzzy Logic*. [S.l.]: Springer, 1999.

ODONE, F.; TRUCCO, E.; VERRI, A. A trainable system for grading fish from images. *Applied Artificial Intelligence*, v. 15, n. 8, p. 735–745, 2001.

PARK, J.-S.; OH, M.-J.; HAN, S. Fish disease diagnosis system based on image processing of pathogens' microscopic images. In: *Proc. Frontiers in the Convergence of Bioscience and Information Technologies*. [S.l.: s.n.], 2007. p. 878–883.

PHADIKAR, S.; SIL, J. Rice disease identification using pattern recognition techniques. In: *2008 11th International Conference on Computer and Information Technology*. [S.l.: s.n.], 2008. p. 420–423.

PINKIEWICZ, T.; PURSER, G.; WILLIAMS, R. A computer vision system to analyse the swimming behaviour of farmed fish in commercial aquaculture facilities: a case study using cage-held Atlantic salmon. *Aquacultural Engineering*, v. 45, n. 1, p. 20–27, July 2011.

PUGOY, R.; MARIANO, V. Automated rice leaf disease detection using color image analysis. In: *3rd International Conference on Digital Image Processing*. [S.l.: s.n.], 2011. v. 8009, p. F1 – F7.

PYDIPATI, R.; BURKS, T.; LEE, W. Statistical and neural network classifiers for citrus disease detection using machine vision. *Transactions of the ASAE*, v. 48, n. 5, p. 2007–2014, 2005.

PYDIPATI, R.; BURKS, T.; LEE, W. Identification of citrus disease using color texture features and discriminant analysis. *Computers and Electronics in Agriculture*, v. 52, n. 1-2, p. 49–59, June 2006.

RODRIGUEZ, A. et al. Optical fish trajectory measurement in fishways through computer vision and artificial neural networks. *Journal of Computing in Civil Engineering*, v. 25, n. 4, p. 291–301, July/August 2011.

SANYAL, P. et al. Color texture analysis of rice leaves diagnosing deficiency in the balance of mineral levels towards improvement of crop productivity. In: *10th International Conference on Information Technology (ICIT 2007)*. [S.l.: s.n.], 2007. p. 85–90.

STEINEBACH, C.; PEREIRA, C. SALMEX: An intelligent knowledge based system for the diagnosis of salmonid fish diseases. In: *Proc. 2nd Scandinavian Conference on Artificial Intelligence*. [S.l.: s.n.], 1989. p. 12–15.

STRACHAN, N. Length measurement of fish by computer vision. *Computers and Electronics in Agriculture*, v. 8, n. 2, p. 93–104, March 1993.

STRACHAN, N.; KELL, L. A potential method for the differentiation between haddock fish stocks by computer vision using canonical discriminant analysis. *ICES Journal of Marine Sciences*, v. 52, n. 1, p. 145–149, February 1995.

TILLETT, R.; MCFARLANE, N.; LINES, J. Estimating dimensions of free-swimming fish using 3D point distribution models. *Computer Vision and Image Understanding*, v. 79, n. 1, p. 123–141, July 2000.

TORISAWA, S. et al. A digital stereovideo camera system for three-dimensional monitoring of free-swimming Pacific bluefin tuna, $Thunnus\ orientalis$, cultured in a net cage. *Aquatic Living Resources*, v. 24, n. 2, p. 107–112, April 2011.

WANG, H. et al. Application of neural networks to image recognition of plant diseases. In: *Proceedings of the 2012 International Conference on Systems and Informatics (ICSAI)*. [S.l.: s.n.], 2012. p. 2159–2164.

WHITE, D.; SVELLINGEN, C.; STRACHAN, N. Automated measurement of species and length of fish by computer vision. *Fisheries Research*, v. 80, n. 2-3, p. 203–210, September 2006.

WIWART, M. et al. Early diagnostics of macronutrient deficiencies in three legume species by color image analysis. *Computers and Electronics in Agriculture*, v. 65, n. 1, p. 125–132, January 2009.

XIAOSHUAN, Z. et al. Applying evolutionary prototyping model in developing FIDSS: An intelligent decision support system for fish disease/health management. *Expert Systems with Applications*, v. 36, n. 2, p. 3901–3913, March 2009.

XU, G. et al. Use of leaf color images to identify nitrogen and potassium deficient tomatoes. *Pattern Recognition Letters*, v. 32, n. 11, p. 1584–1590, ago. 2011.

XU, J. et al. Behavioral responses of tilapia (*Oreochromis niloticus*) to acute fluctuations in dissolved oxygen levels as monitored by computer vision. *Aquacultural Engineering*, v. 35, n. 3, p. 207–217, October 2006.

XU, M.; ZHANG, J.; TAN, X. Intelligent fish disease diagnostic system based on SMS platform. In: *Proc. Third Intern. Conf. on Intelligent System Design and Engoneering Applications*. [S.l.: s.n.], 2013. p. 897–900.

YAO, Q. et al. Application of support vector machine for detecting rice diseases using shape and color texture features. In: *2009 International Conference on Engineering Computation*. [S.l.: s.n.], 2009. p. 79–83.

YOUWEN, T.; LI, T.; YAN, N. The recognition of cucumber disease based on image processing and support vector machine. In: *2008 Congress on Image and Signal Processing*. [S.l.: s.n.], 2008. p. 262–267.

ZELDIS, D.; PRESCOTT, S. Fish disease diagnosis program – problems and some solutions. *Aquacultural Engineering*, v. 23, n. 1, p. 3–11, January 2000.

ZHANG, M.; MENG, Q. Automatic citrus canker detection from leaf images captured in field. *Pattern Recognition Letters*, v. 32, n. 15, p. 2036–2046, nov. 2011.

ZHANG, Q.; GUPTA, K. *Neural Networks for RF and Microwave Design*. [S.l.]: Artech House, 2000.

ZHU, W.; LI, D. A CBR system for fish disease diagnosis. In: *Proc. International Conference on Computer and Computing Technologies in Agriculture*. [S.l.: s.n.], 2008. p. 1453–1457.

ZION, B. The use of computer vision technologies in aquaculture – a review. *Computers and Electronics in Agriculture*, v. 80, p. 125–132, January 2012.

ZION, B.; SHKLYAR, A.; KARPLUS, I. In-vivo fish sorting by computer vision. *Aquacultural Engineering*, v. 22, n. 3, p. 125–132, June 2000.

## AUTHORS BIOGRAPHIES

**Jayme G. A. Barbedo** (Corresponding Author) received the B.S. degree in electrical engineering from the Federal University of Mato Grosso do Sul, Campo Grande, Brazil, in 1998, and the M.Sc. and Ph.D. degrees from the State University of Campinas, Campinas, Brazil, in 2001 and 2004, working on objective assessment of speech and audio.

From 2004 to 2005, he worked with the Source Signals Encoding Group of the Digital Television Division, CPqD Telecom and IT Solutions, Campinas. From 2006 to 2007, he was with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, as a Postdoctoral Fellow. In 2008, he was awarded a fellowship from the Foreign Affairs and International Trade Canada (DFAIT) to work as a Postdoctoral Fellow at the Department of Computer Science of the University of Victoria, Victoria, BC, Canada, where he conducted studies in the areas of instrument identification and audio source separation. Since 2011, he has been a researcher at the Brazilian Agricultural Research Corporation (Embrapa), as part of the Agricultural Informatics research unit. He has published several papers in some of the most important journals and conferences of the area of digital audio and digital image processing. His research interests also include audio and video encoding applied to digital television broadcasting, automatic music transcription, super-resolution spectral estimation, computer vision applied to agricultural problems, among others.