International Journal of Advanced Smart Convergence Vol.12 No.4 208-216 (2023)
http://dx.doi.org/10.7236/IJASC.2023.12.4.208

IJASC 23-4-26

Designing Dataset for Artificial Intelligence Learning for Cold Sea Fish Farming

[1]Sung-Hyun KIM, [2]Seongtak OH, [3]Sangwon LEE

[1]*Executive Researcher, Dept. of Data, National Information Society Agency, Korea*
kimcon@nia.or.kr
[2]*Research Fellow, Dept. of ICT Infrastructure & Platform, National Information Society Agency, Korea*
ohst@nia.or.kr
[3]*Professor, Dept. of Computer & Software Engineering, Wonkwang Univ., Korea*
sangwonlee@wku.ac.kr

Abstract

The purpose of our study is to design datasets for Artificial Intelligence learning for cold sea fish farming. Salmon is considered one of the most popular fish species among men and women of all ages, but most supplies depend on imports. Recently, salmon farming, which is rapidly emerging as a specialized industry in Gangwon-do, has attracted attention. Therefore, in order to successfully develop salmon farming, the need to systematically build data related to salmon and salmon farming and use it to develop aquaculture techniques is raised. Meanwhile, the catch of pollack continues to decrease. Efforts should be made to improve the major factors affecting pollack survival based on data, as well as increasing the discharge volume for resource recovery. To this end, it is necessary to systematically collect and analyze data related to pollack catch and ecology to prepare a sustainable resource management strategy. Image data was obtained using CCTV and underwater cameras to establish an intelligent aquaculture strategy for salmon and pollock, which are considered representative fish species in Gangwon-do. Using these data, we built learning data suitable for AI analysis and prediction. Such data construction can be used to develop models for predicting the growth of salmon and pollack, and to develop algorithms for AI services that can predict water temperature, one of the key variables that determine the survival rate of pollack. This in turn will enable intelligent aquaculture and resource management taking into account the ecological characteristics of fish species. These studies look forward to achievements on an important level for sustainable fisheries and fisheries resource management.

Keywords: *Annotation, Artificial Intelligence, Dataset, Labelling, Learning.*

1. Introduction

Aquaculture has replaced fishing in the future food industry by growing at an average annual rate of 4.8% over the last ten years, and the fisheries structure has reorganized to become more focused on aquaculture.

Manuscript Received: october. 21, 2023 / Revised: october. 27, 2023 / Accepted: November. 3, 2023
Corresponding Author: Sangwon LEE (sangwonlee@wku.ac.kr)
Tel: +82-63-850-6566
Author's affiliation: Professor, Dept. of Computer & Software Engineering, Wonkwang Univ., Korea

While the current aquaculture industry has primarily expanded in developing nations due to their comparative advantage in terms of labor costs and the natural environment, advanced nations with competitive labor markets are now spearheading the industry's growth through the application of cutting-edge technologies. However, because of the aging population and ongoing decline in the fishing population, the nation's aquaculture and fisheries industries have stagnated, putting the fisheries sector in dire straits. Furthermore, because aquaculture relies on natural disasters for its production, the industry's mortality rate is rising as a result of inadequate feed and disease control. Disease-related losses are projected to be between 25 and 30 percent, and the annual damage amount is estimated to be around 250 billion won. Furthermore, because domestic aquaculture uses less productive technologies than aquaculture in developed nations, it is becoming less competitive. These challenges are mostly being brought on by the rise in mortality from illnesses and natural disasters, the growing number of aquaculture technicians working abroad, and the absence of data linkage in developing technologies.

Therefore, the goal of this study is to use data on salmon and pollack to create a learning dataset that can be used for Artificial Intelligence (AI) [1, 2, 3, 4] analysis with Bigdata [5, 6, 7, 8]. By processing and labeling the data, this learning dataset can be used to increase the survival of salmon and pollack. By doing this, we hope to create large-scale learning data and lay the groundwork for the development of intelligent services, all the while advancing the growth of the AI sector and the aquaculture and salmon and pollack industries. By combining AI technology with gathered images, sensors, and management data, this study seeks to diagnose the current state of breeding management and generate healthy salmon and pollack. Additionally, by obtaining smart salmon/porkfish AI technology, we would like to support the Gangwon-do pollack saving project and the salmon cluster business. We anticipate that by doing this, the domestic fisheries and aquaculture sector will be able to adapt to changes in both the internal and external environment more successfully.

2. Dataset for Artificial Intelligence

2.1. Concept of dataset

One of the most important tools for developing AI models and enhancing their functionality is an AI learning dataset [9], which is a collection of information used to train models to carry out a range of functions. The model uses this dataset to better understand and forecast the relationship between input and output. There are two primary components that make up the learning dataset. First, the information given to the model for learning and prediction purposes is known as the input data. Second, the expected output value for each input is the correct answer, or label, which aids in the model's ability to predict the right outcome. The supervised learning dataset is the most prevalent kind of learning dataset. The model is trained to understand the relationship between inputs and correct answers as well as to make precise predictions about new inputs because the inputs and correct answers in this dataset are paired. A supervised learning dataset with a numeric label for each image, for instance, is used to train a model to recognize handwritten numbers. Most of the time, unsupervised learning datasets yield incorrect responses. The way the model learns is by identifying patterns, structures, or rules on its own in the input data. The purpose of unsupervised learning algorithms like dimensional reduction and clustering is to find and comprehend patterns in data. The reinforcement learning dataset is another kind that agents use to interact with their surroundings and carry out specific tasks. By documenting agent behaviors and assigning incentives or penalties, this dataset enhances the model. The movements of the agent and the rewards it receives, for instance, constitute a learning dataset in reinforcement learning tasks where robots move through the environment and accomplish their objectives. The performance and generalization of the model are significantly impacted by the caliber and variety of the learning dataset.

Reputable datasets support models' performance across a range of scenarios. When gathering datasets, one should be aware of bias, imbalances, and noise. Furthermore, an adequate quantity of data is necessary, and the data needs to be appropriately labeled. The dataset's complexity and size have a direct impact on how the model is trained. Models can discover more patterns and general features with the aid of large datasets. Furthermore, the variety of data aids in the model's ability to adjust to various settings. Thus, careful planning, data collection, refinement, labeling, and ongoing assessment and modification during model training are necessary to create an efficient AI learning dataset. The model's real-world performance will increase with the quality of the dataset.

2.2. Status of data labeling

The human eye gathers and processes sensory data, which allows it to surprisingly naturally comprehend its surroundings. But unlike humans, machines - particularly artificial intelligence and machine learning models - do not have the same intuitive understanding of the world. This necessitates providing the data with structure and meaning, which is the essence of data annotation [10] or labeling [11]. The process of transforming data into a format that machine learning algorithms can understand by endowing it with meaningful information is known as data labeling. It is especially crucial to supervised learning because it makes it easier for the model to predict new inputs accurately by outlining the connection between input data and matching outputs or labels. Several notation types are used in the process of data labeling. For images, there is an announcement for the bounding box that indicates the object's location, an announcement for segmentation that indicates the object's pixel level, and an additional announcement for instance segmentation that indicates which object each pixel belongs to. Among the common annotation tasks in text data are named entity recognition (NER), emotion analysis, and document labels. Labeling data is frequently a passive process. Manual annotation is an expensive and time-consuming process, particularly when done on large amounts of data like movies and photos. Additionally, while labor-intensive, this calls for accuracy and skill; consistency and quality control are crucial.

Machine learning-based automatic annotation techniques have been evolving recently. Research is being conducted, in particular, to automatically perform annotation using transfer learning and deep learning techniques in the field of computer vision. However, there are still issues with generalization and accuracy with these methods. Ethical issues are taken into account during the data collection and processing phases when labeling data. Ethical discussions revolve around concerns pertaining to the handling of confidential data, taking into account cultural differences, and the obligations and roles of informants. Issues with prejudice and privacy need to be carefully considered. Enhancing machine learning and artificial intelligence model performance is largely dependent on data labeling. Rich and accurate annotations lessen bias and support diversity considerations by enabling the model to generalize and function in a range of settings. In the field of data labeling, there are still certain difficulties. Better ethical frameworks and guidelines will be required in addition to technological advancements that efficiently process large amounts of data and increase the accuracy of automatic notations. In the contemporary digital world, data labeling plays a crucial role in the interpretation and value of information. This bolsters the progress of AI and machine learning, and reliable and moral annotations enable models to carry out useful tasks in the real world. Future data-driven technology innovations will be propelled by data labeling, which will be achieved through ongoing technological innovation and ethical mindfulness.

3. Data collection for cold sea fish farming

3.1. Procedure of data collection

Image data underwater is photographed at all times using high-definition Closed-Circuit Television (CCTV) and underwater cameras, and the main purpose of collection is to develop a salmon and pollack object recognition model, and to measure the length of salmon and pollack and to establish a growth prediction model. This data is performed in six tanks for 24 hours. In addition, sampling of image data is performed over three weeks by directly photographing organisms in the tank using a high-definition Digital Single Lens Reflex (DSLR) camera, and the main purpose is to measure the length of salmon and pollack and to build a growth prediction model. Sensor data is measured at all times for 24 hours in 6 tanks through Internet of Things (IoT) sensors, and the main purpose is to develop a growth prediction model for salmon and pollack and a water temperature prediction model. Finally, the management data collected by hand is collected directly by data item and is performed on average once/day. The main purpose is to collect data for the growth prediction model of salmon and pollack.

Data collection criteria consider time, resolution, location environment, and conditions. The underwater salmon image data and the underwater pollack image data are stored in image format, with a resolution of 958x720 respectively. This data is collected from a data collection tank in the form. In addition, the sampling salmon image data and the sampling pollack image data are stored in Joint Photographic (Experts) Group (JPG) format, with resolutions of 1440x811 and 6000x4000, respectively. This data consists of images taken within the farm and is collected as a sampling for specific purposes. The sensor data is stored in Comma Separated Value (CSV) format, and the resolution appears to have no specific value. This is information that is constantly measured in the data collection tank in the farm. Finally, the management data is stored in CSV format, and there is no specific resolution value. This data is collected primarily by hand within the form and is used for administrative purposes.

3.2. Considerations of data collection

The key to any process is the acquisition of the data. We decided to secure raw data through pre-sampling, and utilize tools supplied to fish houses for a long period of time and with high satisfaction. This tool was successfully utilized in last year’s data building business and has proven reliability. There are several key considerations in acquiring and managing data. All data (images, sensors, and management data) should be processed as time series-based data and should maintain a reliable record by minimizing daily missing data. In particular, image data will be used to develop length measurement models, so it is necessary to install a cradle that can infer the length in the water. Sensor data is kept accurate through constant inspection by deploying on-site resident managers, and management data is not collected in real time, so the on-site resident managers must collect data manually without omission. This careful approach will contribute to ensuring the reliability and usefulness of the data. An example of data collection schema for AI learning for cold sea fish farming is Table 1.

4. Data refinement for cold sea fish farming

4.1. Procedure of data refinement

We are systematically building data using our own collaboration system. Refining raw data is mainly divided into three main categories: image data from underwater and digital cameras, sensor data, and management data. We focus on applying refining methods to suit the characteristics of each data and improving data quality. In the case of image data, after converting to an image, the unsuitable image is deleted through visual inspection. This ensures consistency and accuracy of your data and prepares you for future use. Sensor data is refined using the outlier removal algorithm and missing value processing algorithm built into the integrated labeling system, and finally, in the case of management data, the system checks the management cycle for each management target and requests the person in charge of the form to re-enter the unentered data. Through this, we continue to push for efficient and reliable data construction. In the past, the frequency of blurred images has generally been high when refining converted images from images. However, this study attempted to minimize the number of blurred images by fixing the location, focus, and target of the underwater camera. In addition, high-pass filters were used to automatically improve object clarity when converting images into images.

Table 1. Data schema for data collection

Class	Class	Data Item	Data type	Acquired quantity	Data format	Acquisition cycle	Method to obtain
image	salmon image	underwater salmon image	image	80,000 copies	JPG	1 copy / 10 sec	underwater camera, CCTV
		sampling salmon image	image	800 copies	JPG	60 copy / 1 time	high-definition camera
	pollack image	underwater pollack image	image	40,000 copies	JPG	1 copy / 10 sec	underwater camera, CCTV
		sampling pollack image	image	400 copies	JPG	60 copy / 1 time	underwater camera
sensor	salmon sensor	DO	text	600,000 sets	CSV	1 set / 1 min	IoT sensor
		water temperature	text		CSV	1 set / 1 min	IoT sensor
		PH	text		CSV	1 set / 1 min	IoT sensor
		CO2	text		CSV	1 set / 1 min	IoT sensor
		ORP	text		CSV	1 set / 1 min	IoT sensor
		flow	text		CSV	1 set / 1 min	IoT sensor
		illumination	text		CSV	1 set / 1 min	IoT sensor
	pollack sensor	DO	text	300,000 sets	CSV	1 set / 1 min	IoT sensor
		water temperature	text		CSV	1 set / 1 min	IoT sensor
		inflow water temperature	text		CSV	1 set / 1 min	IoT sensor
		temperature	text		CSV	1 set / 1 min	IoT sensor
		flow	text		CSV	1 set / 1 min	IoT sensor
		illumination	text		CSV	1 set / 1 min	IoT sensor
check	salmon check	ammonia nitrogen	text	200 sets	CSV	1 time / 1 day	spectrophotometer
		nitrite	text		CSV	1 time / 1 day	spectrophotometer
		nitrate	text		CSV	1 time / 1 day	spectrophotometer
		alkalinity	text		CSV	1 time / 1 day	spectrophotometer
		total bacterial count	text		CSV	1 time / 1 day	portable measurer
		total gas pressure	text		CSV	1 time / 1 day	portable measurer
		sampling actual	text	800 sets	CSV	120 copy /1 time	portable measurer
		food supply check	text	200 sets	CSV	everyday	actual/manual
		biological	text	1 docs	CSV	1 doc / 1 year	actual/manual
		daily death population	text	200 sets	CSV	1 set /1 day	actual/manual
	pollack check	total gas pressure	text	200 sets	CSV	1 set /1 day	portable measurer
		Sea temperature	text		CSV	1 time / 1 day	portable measurer
		sampling actual	text	400 sets	CSV	60 copy /1 time	actual/manual
		food supply check	text	200 set	CSV	everyday	actual/manual
		biological	text	1 docs	CSV	1 doc / 1 year	actual/manual
		daily death population	text	200 sets	CSV	1 time /1 day	actual/manual

The raw data refining tool consists of several modules. The tool includes a transducer that converts images into images, a tool that screens images by visual inspection, a module for correcting missing values and removing outliers from time series sensor data, and a module that asks the form manager to re-enter if management data is not entered. The image data converts underwater camera images acquired in units of time into images once every 10 seconds and stores them in data storage. Sampling biometric image data taken manually is stored the same. Image data refining uses a dedicated tool utilizing the consortium's Windows application, which includes a missing calibration module and an outlier removal module. The missing value correction module uses a loss processing algorithm to address data loss caused by equipment or network errors. The outlier removal module maintains accurate data by removing measurements that cannot come from reality. In addition, the image data purification inspection tool uses a purification application dedicated to image data to deliver the primary refined data to the data inspector for purification inspection.

Table 2. Data schema for data refinement

Class	Class	Data item	Data type	Refined quantity	Data format
image	salmon image	underwater salmon image	image	80,000 copies	JPG
		sampling salmon image	image	801 copies	JPG
	pollack image	underwater pollack image	image	40,000 copies	JPG
		sampling pollack image	image	404 copies	JPG
sensor	salmon sensor	DO	text	663,220 sets	CSV
		water temperature	text		CSV
		PH	text		CSV
		CO2	text		CSV
		ORP	text		CSV
		flow	text		CSV
		illumination	text		CSV
	pollack sensor	DO	text	331,611 sets	CSV
		water temperature	text		CSV
		inflow water temperature	text		CSV
		temperature	text		CSV
		flow	text		CSV
		illumination	text		CSV
check	salmon check	ammonia nitrogen	text	536 sets	CSV
		nitrite	text		CSV
		nitrate	text		CSV
		alkalinity	text		CSV
		total bacterial count	text		CSV
		total gas pressure	text		CSV
		sampling actual	text	801 sets	CSV
		food supply check	text	264 sets	CSV
		biological	text	1 document	CSV
		daily death population	text	552 sets	CSV
	pollack check	total gas pressure	text	268 sets	CSV
		Sea temperature	text		CSV
		sampling actual	text	404 sets	CSV
		food supply check	text	536 sets	CSV
		biological	text	1 document	CSV
		daily death population	text	340 sets	CSV

4.2. Considerations of data refinement

Refining of raw data from the field should not be missing at all. In addition, data purification should be consistently accurate. Purification work should be carried out consistently with clear standards. In other words, ambiguity or outliers should be communicated with the data quality manager and work should be clarified. Refining work always requires careful attention, so the refining manual must be observed. Repetitive tasks shall be faithfully performed without omission, and daily business reports shall be executed without omission. Efforts should always be made to ensure that communication between workers is maintained smoothly. An example of data refinement schema for AI learning for cold sea fish farming is Table 2.

5. Data labeling for cold sea fish farming

5.1. Procedure of data labeling

Data labeling is mainly divided into image data and sensor/management time series data. This study uses a self-developed labeling tool to apply the Polygon labeling technique to refined images generated by Hollow spender instruments installed in the tank. Sensor/Management data is not subject to manual labeling. When labeling an image is complete, the system automatically enters information in the Annotation information for that image. The input criteria are based on when the image was taken, and the most recent (past) input of sensor/management data is entered.

5.2. Considerations of data labeling

Accurate and clear object identification is key to the processing of pollack and salmon fish data. There are some criteria for this. First, for normal object identification, the outline of the object must be clearly visible or at least 50% of the image must be clearly visible. In addition, it should not overlap with other objects and should not be affected by external elements such as fluorescent light or rope. If affected, the object is excluded from processing. Labeling operations require regular intervals for straight lines along the edges of the object, and relatively tight labels for curves. In addition, each individual must be labeled with at least 30 dots. This allows you to accurately understand the shape and structure of the object. In particular, certain parts, such as side fins or tail fins, should be excluded from labeling. This is because that part does not represent the characteristics of the overall object. These processing operations will enable accurate identification and labeling of pollack and salmon fish data.

To effectively manage processed (labeling) data, several fundamentals must be considered. First of all, data annotation must be carried out by establishing standards suitable for the purpose. Depending on the purpose of data use, you should clearly define what information you want to annotate and perform the annotation based on this. In addition, the annotation items in the data should be designed to be difficult to change. However, changes should only be made if the purpose of data use changes and modifications are required. Annotation information should be easily understood and confusion caused by unclear meaning should be minimized. Clear and consistent annotations increase data utilization. You should check from time to time whether the data meets the purpose of use and maintain consistent data. Consistency of annotation has a significant impact on model learning and evaluation. It is necessary to balance the bias of the data by continuously adding it as needed. Complementation of biased data is needed to improve the performance of the model. It is important to adhere to the retention schedule and regulations of your data. It is necessary to preserve data according to legal requirements, and to properly manage storage and storage methods. Effective management of processed data taking into account these fundamentals can contribute to the learning and performance of the model. An example of data labeling schema for AI learning for cold sea fish farming is Table 3.

Table 3. Data schema for data labeling

Attribute name	Attribute description	Data type	Mandatory/ Optional	Example
filename	file name	string	mandatory	D01_T1S_210701_1.JPG
id	id	string	mandatory	21547
date	date of filming	string	mandatory	2023.01.01 01:01:01
file_format	data format	string	mandatory	JPG
imsize	file size	string	mandatory	1563(KB)
region_name	filming area name	string	mandatory	A farm
images_location	filming location	string	mandatory	2 water tank cleaner
copyright	copyright information	string	mandatory	James
width	horizontal size	number	mandatory	1920
height	vertical size	number	mandatory	1080
resolution	resolution	string	mandatory	HD
bit	color	number	mandatory	24
aspect_ratio	aspect ratio	string	mandatory	04:03
pixel	pixel	string	mandatory	1K
color_depth	depth of color	string	mandatory	sRGB
iso	iso sensitivity	string	mandatory	3200
white_balance	white balance	string	mandatory	5500K
exposure_time	exposure time	string	mandatory	f2.8 1/80
f-stop	aperture value	string	mandatory	f2.8
flash	flash	string	mandatory	-
filter	filter status	string	mandatory	-
focal_length	focal length	string	mandatory	50mm
angle	filming angle	string	optional	120 degrees
weather	weather information	string	optional	indoor
CCTV_name	CCTV name	string	mandatory	CCTV #1
farm_name	name of fish farm	string	mandatory	farm A
tank_id	water tank id	string	mandatory	1
tank_size	water tank size	number	mandatory	6
do_value	do value	number	mandatory	7.5
water_temp_value	water temperature value	number	mandatory	30.1
PH_value	PH value	number	mandatory	5.7
CO2_value	CO2 value	number	mandatory	3.5
OPR_value	OPR value	number	mandatory	12
flux_value	flow rate value	number	mandatory	30
influentwatertemp_value	inflow water temperature data	number	mandatory	7
temperatures_value	temperature	number	mandatory	23
seawatertemp_value	sea temperature data	number	mandatory	12
illuminance_value	illumination value	number	mandatory	30
ammonia_value	ammonia nitrogen value	number	mandatory	45
nitrite_value	nitrite value	number	mandatory	12
nitrate_value	nitrate value	number	mandatory	54
alkalido_value	alkaline degree value	number	mandatory	23
totalbacteria_value	total bacterial count value	number	mandatory	5
totalgaspressure_value	total gas pressure value	number	mandatory	12
dead_value	daily death population value	number	mandatory	10
feed_supply_value	feeding supply	number	mandatory	35.7
feed_supply_time	feeding time	string	mandatory	2023.01.01 01:01:01
stock_date	entry date	string	mandatory	2021.04.01 12:05:34
harvest_value	shipment date	string	mandatory	2023.01.01 01:01:01
grade_value	selection date	string	mandatory	2023.01.01 01:01:01

6. Conclusions

This study aimed to establish an effective strategy through data utilization in the field of fish farming in Gangwon-do. The establishment of datasets and AI models targeting salmon and pollack can provide important information for predicting growth and predicting ecological variables in fish farming. It is also possible to present technical strategies for managing food and water quality according to the period of growth in salmon and pollack aquaculture, and improving the survival rate of pollack. Through the continuous use of datasets, the aquaculture sector of the fisheries industry can improve data quality by receiving feedback on data that require de-identification processing, and derive specific directions for future research and service improvement. Consequently, data utilization and technical support in fish farming are expected to be utilized as effective tools for sustainable fisheries and fisheries resource management. In the future, by examining user satisfaction with the developed service and studying the quality and utilization of datasets to improve service performance, it is believed that effective implementation of data-based artificial intelligence model development strategies in the domestic fish farming field will be possible.

Acknowledgement

This paper was supported by Wonkwang University in 2023.

References

[1] M. A. Boden, "Artificial intelligence", Elsevier, 2023.

[2] Y. Mintz, and R. Brodie, "Introduction to artificial intelligence in medicine", Minimally Invasive Therapy & Allied Technologies, Vol. 28, No. 2, pp. 73-81, 2019.
DOI: https://doi.org/10.1080/13645706.2019.1575882

[3] S.H. Kim, J.K. Choi, J.S. Kim, A.R. Jang, J.H. Lee, K.J. Cha, and S.W. Lee, "Animal Infectious Diseases Prevention through Bigdata and Deep Learning", Journal of Intelligence and Information Systems, Vol. 24. No. 4. pp. 137-154, 2018.
DOI: https://doi.org/10.13088/jiis.2018.24.4.137

[4] C. H. Lee, K. Park, and S. Lee, "Development of IoT Healthcare Platform Model for the Elderly using Bigdata and Artificial Intelligence", International Journal of Membrane Science and Technology, Vol. 10, No. 1, pp. 108-113, 2023.
https://doi.org/10.15379/ijmst.v10i1.1435

[5] D. Fasel, and A, Meier, Big data, Springer Vieweg, 2014

[6] S. Lee and J. Kim, "Data Design Strategy for Data Governance Applied to Customer Relationship Management", International Journal of Advanced Culture Technology, Vol. 11, No. 3, pp. 338-345, 2023.
DOI: https://doi.org/10.17703/IJACT.2023.11.3.338

[7] S.H. Kim, S. Chang, and S.W. Lee, "Consumer Trend Platform Development for Combination Analysis of Structured and Unstructured Bigdata", Journal of Digital Convergence, Vol. 15. No. 6. pp. 133-143, 2017.
DOI: https://doi.org/10.14400/JDC.2017.15.6.133

[8] S. Sagiroglu, and D. Sinanc, "Big data: A review", The 2013 international conference on collaboration technologies and systems (CTS), IEEE, pp. 42-47. 2013, May.
DOI: http://doi.org/10.1109/CTS.2013.6567202

[9] J. A. Nichols , H. W. Herbert Chan, and M. A. Baker, "Machine learning: applications of artificial intelligence to imaging and diagnosis", Biophysical reviews, Vol. 11, pp. 111-118, 2019.
DOI: https://doi.org/10.1007/s12551-018-0449-9

[10] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques. Pattern Recognition", Vol. 45, No. 1, pp. 346-362, 2012.
DOI: https://doi.org/10.1016/j.patcog.2011.05.013

[11] M. Desmond, M. Muller, Z. Ashktorab, C. Dugan, E. Duesterwald, K. Brimijoin, ... and Q. Pan, "Increasing the speed and accuracy of data labeling through an ai assisted interface", The 26th International Conference on Intelligent User Interfaces, pp. 392-401, 2021, April.
DOI: https://doi.org/10.1145/3397481.3450698