



A Brief Review on Aquaculture Genetics, Machine Learning, and Their Convergence

Thisarani Ediriweera^{1*}, Prabuddha Manjula²

¹Department of Bio-AI Convergence, Chungnam National University, Daejeon 34134, South Korea

²Division of Animal and Dairy Science, Chungnam National University, Daejeon 34134, South Korea

ABSTRACT

This concise account on aquaculture, aquaculture genetics, and emerging trends of its convergence with machine learning, a sub-class of artificial intelligence provides, succinct overviews for each of the disciplines separately, their basics, and machine learning approaches in aquaculture genetics, in a consolidative manner to brief their status, applications and prospects.

Key words: Aquaculture, Convergence, Genetics, Machine learning

AQUACULTURE AND AQUACULTURE GENETICS

As the global population escalates with an expectation of 9 billion heads by the year 2050 with no increment or expansion in the world's natural resources, the adequate and sustainable supply of foods for all, has become a flaming issue (Garcia and Rosenberg, 2010). In the sense of Fish production, the capture fisheries which is having proof of commercial-scale practices even before the 1500s bared the sole responsibility (Lackey, 2005). However, due to the reasons of both increment of population and declining of natural fisheries stocks, the capture fisheries production became stagnated. Later, with the first attempt of commercial aquaculture practices emerged in Germany in the year 1733 (The Healthy fish, 2019) besides the small scale isolated and primitive fish farming practices remained, revolutionized the fish production gradually and felicitously.

In the present context, the aquaculture is affected by complex series of factors including, type of species, type of production system, the intensity of production system, water quality, temperature, feeds and feeding, health/disease management, stocking density, stress management, biosecurity measures, reproduction, harvesting, available human resources, economic concerns, legal framework, etc. throughout its production cycle (Losordo and Westerman, 1994, Pillay and Kutty, 2005, Moyo and Rapatsa, 2021).

The genetic background of aquatic organisms including both fish and shellfish (here onwards referred to as fish), interconnects with most of the above-mentioned biological factors. As it determines the desirable phenotypes of cultured fish by addressing the genetic factors along with environmental factors, its consideration is crucial for

***Corresponding author:** Thisarani Ediriweera

Department of Bio-AI Convergence, Chungnam National University, Daejeon 34134, South Korea
Tel: +82-42-821-5779 E-mail: tkediriweera@gmail.com

Received: 23 September, 2021, Revised: 29 September, 2021, Accepted: 30 September, 2021



© Journal of Animal Breeding and Genomics 2021. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

successful and profitable aquaculture practices (Lutz, 2008). Accordingly, growth rates, survival rates, muscle ratios, feed conversion ratios, breeding capacities, etc. can be listed as some direct spheres that are affected by genetic factors of an individual, population, or species of fish in culture practices (Wilkins, 1981, De Verdal et al., 2018).

Development of new aquaculture species including hybridization, production of transgenic fish, application of genomic technologies and genetic engineering with genome sequencing, genome editing (ex: CRISPR/Cas9) and gene knockouts on acquiring desired traits related to the disease resistance, stress resistance, high growth rates, etc. (Ex: AquAdvantage Salmon (AAS) (Sweet 2019), CRISPR Cas developed *Oreochromis niloticus* (Evans, 2018)), genetic diversity and allelic diversity by means of minisatellites, microsatellites, Single Nucleotide Polymorphisms (SNPs), Linkage mapping, Selective breeding, inbreeding and interspecific crossbreeding, Sex manipulation, gynogenesis, androgenesis and cloning, marker-assisted selection and genomic selection, polyploidy and even epigenetics and hologenomics applications have appertained with the field of aquaculture genetics (Dunham et al., 2000, Changadeya, 2003, Shen and Yue, 2019, Okoli et al., 2021).

However, limited or poorly annotated genomic data of aquatic organisms (Sundaram et al., 2017, Wargelius, 2019), retardation of identification, clarification, and confirmation of trait-related fish genes (Okoli et al., 2021), Genome duplications (Glasauer and Neuhaus, 2014), etc. remain as some constraints that are still possible to address or will be resolved in impending years.

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Artificial intelligence which is considered one of the major driving forces of the fourth industrial revolution has been defined as 'the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of success at some goal' (Ongsulee, 2017) in which its implementation overlaps with various spheres of science including, computer science, mathematics, neuroscience, philosophy and psychology (Singh and Jain, 2018).

Machine learning is considered a subpart of the so-called paramount concept of Artificial intelligence, which involves the construction of algorithms that are capable of detecting/ identifying different meaningful patterns of furnished big data, without any explicit programming (Ongsulee, 2017, Aristodemou and Tietze, 2018). However, the machine learning models are usually being trained with training data for training, exercising, inculcating, and priming the models prior to use for the test data to ensure the sensitivity and accuracy of its detections and/or predictions (Shalev-Shwartz and Ben-David, 2014). Accordingly, its life cycle comprises seven steps including data gathering, preparation, wrangling, analysis, model training, testing, and deployment (Javapoint, 2021a).

Further, the classification of machine learning includes three major types namely, supervised, unsupervised, and reinforcement learning. Accordingly, supervised learning holds classification and regression algorithms including Decision Tree (DT), Random Forests (RF), Linear Regression, Logistic Regression, KNN, Support Vector Machine (SVM) whereas unsupervised learning consists of clustering and association algorithms (Apriori algorithm, K-means). However, reinforcement learning learns based on the feedback it receives for each of its actions (Markov Decision Process) (Javapoint, 2021b, AnalyticsVidya, 2017). These different types and different algorithms are employed on various applications suitably.

In the field of genetics, these can be employed in tasks like identification of transcription initiation sites (Ohler et al., 2002), promoter sites (Bucher, 1990), splicing sites (Degroeve et al., 2002), enhancer sites (Heintzman et al., 2007) in genomic sequences, gene annotations (Picardi and Pesole, 2010), recognition of patterns in DNA sequences (i.e., RNA-

seq, DNase-seq, MNase-seq, FAIRE-seq, CHIP-seq), identifying functional relationships (Libbrecht and Noble, 2015), breed identification and classification (Seo et al., 2021) and many other.

USE OF MACHINE LEARNING IN AQUACULTURE GENETICS

Genetic algorithms, a type of stochastic algorithms that are used as adaptive search techniques (Vafaie and De Jong, 1992, Mitchell, 1995, Shapiro, 1999, Gupta and Ghafir, 2012) along with the above-mentioned algorithms has been deployed in the studies associated with the applications of machine learning in aquaculture genetics.

In the aspect of selective breeding of fish especially in line with the health management/disease-related identifications and predictions, machine learning approaches have successfully been employed. Some of the great examples include the 'Predicting for disease resistance in aquaculture species using machine learning models with the use of DT, SVM, RM, AdaBoost (adaptive boosting), and XGB (extreme gradient boosting) models to analyze the resistance of carps over Koi Herpesvirus by Palaiokostas (2021), machine learning on genomic prediction with a focal target of identifying the disease resistance of gilthead sea bream (*Sparus aurata*) over photobacteriosis (Bargelloni et al., 2021) and adopting machine learning algorithms for developing a cost-effective and precise method of using SNPs of Genome-Wide Association Studies (GWAS) for genomic selections with special focus on disease resistance traits using *Litopenaeus vannamei*, Salmon and Gilthead Sea Bream by Luo et al. (2021).

On another aspect, parasites also play a major role in determining the success and profitability of aquaculture practices. Correspondingly, the Random Forest machine learning approach on *Lepeophtheirus salmonis*, a parasite of salmonids has been conducted for the development of DNA markers to distinguish *L. salmonis* populations (Jacobs et al., 2018). Moreover, the previously mentioned Luo et al. (2021) also has addressed *Vibrio parahaemolyticus* infection in *Litopenaeus vannamei* while Lin et al. (2020) have studied the resistance on heterobothriosis, a parasitic disease caused by *Heterobothrium okamotoi*, in pufferfish using machine learning procedures. Additionally, Gautam et al. (2016) have reported a prediction model based on machine learning to predict antimicrobial peptides in fish, using genomic and proteomic data.

Commercial phenotypes based on big data which are in the form of images are analyzed using machine learning models to achieve diverse demands through image analysis. Similarly, the machine learning ANNs (Artificial Neural Networks) have been deployed for obtaining scrupulous phenotypic data through this image analysis technology specifically for pearl oysters and *Penaeus monodon* to analyze the growth data in both and pearl quality data in pearl oysters, that alludes the individuals with good genetic composition on a concerning trait (Zenger, 2019).

Further, the gene analysis of copepod-associated Bacteriobiomes (CABs) has been conducted using machine learning models in a metanalysis-based approach to evaluate the biogeochemical properties like methanogenesis and nitrogen fixation that can also impact mostly on aquaculture practices that involved open marine waters (Sadaiappan et al., 2021).

Even though it is slightly different from the previous applications due to its consideration on residual aquatic organisms including macroinvertebrates, bacteria, etc. living in farming water than the main species which is being cultured, the approaches as follows are also could be considered under aquaculture genetics as the impact the aquaculture practices and involve with genetic materials. The application of Supervised Machine Learning (SML) using DNA metabarcoding to evaluates the impacts on aquaculture practices on communities of micro-faunal, micro-floral and the water itself has been tried, considering bacteria and ciliates as indicator organisms in a Salmon aquafarm (Fruehe et al., 2021).

FUTURE PROSPECTIVE

Aquaculture has become a crucial industry and genetics play a major role in shaping the industry without a doubt. The application of machine learning approaches is recently being applied in aquaculture genetics with an increasing trend and sooner it will uplift the industry to its next level. To facilitate the route for the revolution the machine learning approaches can further be used for genomic markers modeling, fecundity predictions, and expression fingerprinting of economically important aquaculture species by addressing proteomics and transcriptomics data (Abdelrahman et al., 2017) and for genes identifications, evolutionary genetics and population genetics suitably.

ACKNOWLEDGEMENT

This work was also partly supported and partly supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-01441, Artificial Intelligence Convergence Research Center (Chungnam National University)) and Korea Evaluation Institute of Industrial Technology (KEIT) grant funded by the Korea government (MOTIE).

REFERENCES

- Abdelrahman H, ElHady M, Alcivar-Warren A, Allen S, Al-Tobasei R, Bao L, Beck B, Blackburn H, Bosworth B, Buchanan J and Chappell J.
Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genom*, 18: 1-23.
- Aristodemou L. and Tietze F. 2018. The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Pat. Inf.* 55: 37-51.
- Bargelloni L, Tassiello O, Babbucci M, Ferraresso S, Franch R, Montanucci L and Carnier P. 2021. Data imputation and machine learning improve association analysis and genomic prediction for resistance to fish photobacteriosis in the gilthead sea bream. *Aquacul. Rep.* 20: p.100661.
- Bucher P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* 212: 563-578.
- Changadeya, W., Malekano, L.B. and Ambali, A.J.D., 2003. Potential of genetics for aquaculture development in Africa.
- De Verdal H, Vandeputte M, Mekki W, Chatain B and Benzie JA. 2018. Quantifying the genetic parameters of feed efficiency in juvenile Nile tilapia *Oreochromis niloticus*. *BMC genetics*, 19: 1-10.
- Degroeve S, De Baets, B, Van de Peer Y and Rouzé P. 2002. Feature subset selection for splice site prediction. *Bioinformatics*, 18: 75-83.
- Dunham, R.A., Majumdar, K., Hallerman, E., Bartley, D., Mair, G., Hulata, G., Liu, Z., Pongthana, N., Bakos, J., Penman, D. and Gupta, M., 2000, February. Review of the status of aquaculture genetics. In *Aquaculture in the Third Millennium. Technical Proceedings of the Conference on Aquaculture in the Third Millennium*, Bangkok, Thailand (pp. 137-166).
- Evans O (2018) GM salmon farmer receives exemption for gene-edited tilapia in Argentina. *Salmonbusiness.com*. <https://salmonbusiness.com/gm-salmon-farmer-receive-exemption-for-gene-edited-tilapia-in-argentina/> Accessed 14 Jan 2021
- FAO (Food and Agriculture Organization of the United Nations) (2018). *The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals* (p. 227). Rome. Retrieved from <http://www.fao.org/3/i9540en/i9540en.pdf>

- Fruehe L, Cordier T, Dully V, Breiner HW, Lentendu G, Pawlowski J, Martins C, Wilding TA and Stoeck T. 2021. Supervised machine learning is superior to indicator value inference in monitoring the environmental impacts of salmon aquaculture using eDNA metabarcodes. *Mol. Ecol.* 30: 2988-3006.
- Garcia SM and Rosenberg AA. 2010. Food security and marine capture fisheries: characteristics, trends, drivers and future perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365: 2869-2880.
- Gautam A, Sharma, A, Jaiswal S, Fatma S, Arora V, Iquebal MA, Nandi S, Sundaray JK, Jayasankar P, Rai A and Kumar D. 2016. Development of antimicrobial peptide prediction tool for aquaculture industries. *Probiotics. Antimicrob. Proteins.* 8: 141-149.
- Glasauer SM and Neuhauss SC. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genom.* 289: 1045-1060.
- Gupta D and Ghafir S. 2012. An overview of methods maintaining diversity in genetic algorithms. *Int. J. Emerg. Technol. Adv. Eng.* 2: 56-60.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA. and Wang W. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39: 311-318.
- Julie K. (2019) What is aquaculture? A brief history of Fish Farming. Available at: <https://thehealthyfish.com/aquaculture-brief-history-fish-farming/> (Accessed: 14/09/2021)
- Ray S. (2017) Commonly used Machine Learning Algorithms. (with Python and R Codes). Available at: <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/> (Accessed: 15/09/2021)
- Javapoint (2021a) Machine learning Life cycle. Available at: <https://www.javatpoint.com/machine-learning-life-cycle> (Accessed: 14/09/2021)
- Javapoint (2021b) Machine Learning Tutorial. Available at: <https://www.javatpoint.com/machine-learning> (Accessed: 14/09/2021)
- Jacobs A, De Noia M, Praebel K, Kanstad-Hanssen Ø, Paterno M, Jackson D, McGinnity P, Sturm A, Elmer KR and Llewellyn MS. 2018. Genetic fingerprinting of salmon louse (*Lepeophtheirus salmonis*) populations in the North-East Atlantic using a random forest classification approach. *Sci. Rep.* 8: 1-9.
- Kok JN, Boers EJ, Kosters WA, Van der Putten P and Poel M. 2009. Artificial intelligence: definition, trends, techniques, and cases. *Artif. intel.* 1: 270-299.
- Lackey, R.T., 2005. Fisheries: history, science, and management. *Water encyclopedia: surface and agricultural water*, pp.121-129.
- Libbrecht MW and Noble WS. 2015. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16: 321-332.
- Lin Z, Hosoya S, Sato M, Mizuno N, Kobayashi Y, Itou T and Kikuchi K. 2020. Genomic selection for heterobothriosis resistance concurrent with body size in the tiger pufferfish, *Takifugu rubripes*. *Sci. Rep.* 10: 1-13.
- Losordo TM and Westerman PW. 1994. An analysis of biological, economic, and engineering factors affecting the cost of fish production in recirculating aquaculture systems. *J. World. Aquacult Soc.* 25: 193-203.
- Luo Z, Yu Y, Xiang J and Li F. 2021. Genomic selection using a subset of SNPs identified by genome-wide association analysis for disease resistance traits in aquaculture species. *Aquaculture*, 539: p.736620.
- Lutz CG. 2008. *Practical genetics for aquaculture*. John Wiley & Sons.
- Moyo NA and Rapatsa MM. 2021. A review of the factors affecting tilapia aquaculture production in Southern Africa. *Aquaculture*, p.736386.
- Ohler U, Liao, GC, Niemann H and Rubin GM. 2002. Computational analysis of core promoters in the *Drosophila* genome. *Genome Bio.* 3: 1-12.
- Okoli AS, Blix T, Myhr AI, Xu W and Xu X. 2021. Sustainable use of CRISPR/Cas in fish aquaculture: the biosafety perspective. *Transgenic. Res.* 1-21.

- Oliveira LS, Sabourin R, Bortolozzi F and Suen CY. 2003. A methodology for feature selection using multiobjective genetic algorithms for handwritten digit string recognition. *Int. J. Pattern Recognit. Artif. Intell.* 17: 903-929.
- Palaiokostas C. 2021. Predicting for disease resistance in aquaculture species using machine learning models. *Aquac. Rep.* 20: p.100660.
- Picardi E and Pesole G. 2010. Computational methods for ab initio and comparative gene finding. *Data mining techniques for the life sciences.* 269-284.
- Pillay TVR and Kutty MN. 2005. *Aquaculture: principles and practices* (No. Ed. 2). Blackwell publishing.
- Sadaipappan B, PrasannaKumar C, Nambiar VU, Subramanian M and Gauns MU. 2021. Meta-analysis cum machine learning approaches address the structure and biogeochemical potential of marine copepod associated bacteriobiomes. *Sci. Rep.* 11: 1-17.
- Seo D, Cho S, Manjula P, Choi N, Kim YK, Koh YJ, Lee SH, Kim HY, Lee JH. 2021. Identification of Target Chicken Populations by Machine Learning Models Using the Minimum Number of SNPs. *Animals (Basel)*.11: 241.
- Shalev-Shwartz S and Ben-David S. 2014. *Understanding machine learning: From theory to algorithms.* Cambridge university press.
- Shen Y and Yue G. 2019. Current status of research on aquaculture genetics and genomics-information from ISGA 2018. *Aquac. Fish.* 4: 43-47.
- Sundaram A, Tengs T and Grimholt U. 2017. Issues with RNA-seq analysis in non-model organisms: a salmonid example. *Dev. Comp. Immunol.* 75: 38-47.
- Sweet JB. (2019) Draft study on risk assessment: application of annex 1 of decision CP 9/13 to living modified fish. Report for the secretariat of the convention on biological diversity, UN Environmental Programme. Available at: https://bch.cbd.int/protocol/risk_assessment/report%20-%20study%20on%20risk%20assessment%2020.12_final%20for%20posting.pdf. (Accessed 13 Sep 2021)
- Wilkins NP. 1981. The rationale and relevance of genetics in aquaculture: an overview. *Aquaculture*, 22: 209-228.
- Zenger KR, Khatkar MS, Jones DB, Khalilisamani N, Jerry DR and Raadsma HW. 2019. Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front. Genet.* 9:693.
- Vafaie H and De Jong KA. 1992. November. Genetic Algorithms as a Tool for Feature Selection in Machine Learning. In *ICTAI* (pp. 200-203).
- Mitchell M. 1995. September. Genetic algorithms: An overview. In *Complex*. (Vol. 1, No. 1, pp. 31-39).
- Shapiro J. 1999. July. Genetic algorithms in machine learning. In *Advanced Course on Artificial Intelligence* (pp. 146-168). Springer, Berlin, Heidelberg.
- Altun AA. and Allahverdi N. 2007. April. Neural network based recognition by using genetic algorithm for feature selection of enhanced fingerprints. In *International Conference on Adaptive and Natural Computing Algorithms* (pp. 467-476). Springer, Berlin, Heidelberg.
- Ongsulee P. 2017. November. Artificial intelligence, machine learning and deep learning. In *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)* (pp. 1-6). IEEE.
- Singh, DEEPAK and Jain ANKIT. 2018. February. A look into the artificial intelligence and its application in various fields of life. In *International Conference on Advances in Computer Technology and Management (ICACTM)*, Pune, Maharashtra.
- Wargelius A. 2019. August. Application of genome editing in aquatic farm animals: Atlantic salmon. In *Transgenic research* (Vol. 28, No. 2, pp. 101-105). Springer International Publishing.