

## **Machine learning tools can now help with improving Taxol production in plant cell cultures**

**Ranjana Sarma, PhD**  
**Editorial Board Member**  
**Biotechnology Kiosk, USA**

DOI: <https://doi.org/10.37756/bk.21.3.3.2>

Chemotherapeutic intervention for cancer care is an important step. One of the most effective chemotherapy agents in use today is Paclitaxel (PTX), sold under the common name Taxol and Oxanol. Due to its ability to inhibit microtubule formation in cells, PTX is effective at all stages of the cancer and is FDA approved for treatment of many types of cancer (ovarian cancer, esophageal cancer, breast cancer, lung cancer, Kaposi sarcoma, cervical cancer, and pancreatic cancer). PTX is a plant alkaloid in the taxane family of compounds obtained from bark of the Pacific Yew tree (*Taxus brevifolia*) [1]. Adequate market supply of PTX has remained a challenge, as paclitaxel represents only a minor proportion of the total taxoid content of the *Taxus* species. Over the years, research into finding an alternate to cutting down Yew trees for PTX harvesting has been on the forefront. It is estimated that up to 60 trees may need to be harvested for the treatment of one patient.

Genetic manipulation tools have offered researchers a handle to shift from cutting down trees to PTX production in cell suspension cultures. *Agrobacterium*

mediated transformation of taxus cell in suspension cultures has been one way towards developing high yield cell cultures for large scale biomanufacturing [2]. However, setting up in vitro cultures of *Taxus* species is anything but straightforward and the search for high yielding cell cultures is on. One promising candidate is cell suspension culture of hazel plant (*Corylus avellana*), which is being actively researched on. Given that hazel is a dicotyledonous plant, it is expected to be more amenable to genetic manipulation by the *Agrobacterium* mediated transformation method. Researchers at Tarbiat Modares University, Tehran, have recently published findings that offer new insights into improving PTX levels in the total toxoid content of the plant culture [3].

PTX biosynthesis is complex, multifactorial and non-linear, involving of many pathways. To get plants to produce bioactive molecules and secondary metabolites, many factors need to be optimized. Testing so many parameters sequentially by laboratory methods is not only time consuming, but also commercially

non-viable. PTX biosynthesis requires, among others parameters, fungal elicitors. The PTX concentration in the secondary metabolite produced by the cell suspension culture, is affected by the type and concentration of the fungal elicitor, and the specific adding day the fungal elicitors are added. The exact harvesting time of the cell suspension cultures is an additional factor that needs to be optimization but is difficult to predict based on the input parameters. Forecasting PTX bioproduction by calculating the optimized value of so many factors is a difficult as the prediction of the precise effects of each factor and how they interact needs multifactorial analysis. Artificial neural networks (ANNs) and deep learning methods offer promising leads to improving plant biotechnology methods.

ANNs are brain-inspired systems that consist of an input, an output, and several hidden layers [4]. Since the 1940s, ANNs have been researched as machine learning tool, to solve complex relationships amongst input and output data. Deep Neural Networks (DNN) are relatively recent branch of ANNs that are being robustly developed with more hidden layers, with much improved prediction power. Deep learning methods have become a mainstay in plant biotechnology, e.g., to identify multi-dimensional genome-wide phenotypes and create new genetic elements with desirable traits. For such predictions, specific types of DNNs are used called Convolutional Neural Networks (CNN) which have the ability automatic feature extraction from a continuous signal like weather data as a time series, or plant image or DNA/RNA sequence. Recurrent Neural Networks (RNNs) are another subset of ANNs that have

the advantage of handling inputs with different size. In RNNs, outputs of some layers are fed back into the inputs of a previous giving the ANN memory capabilities [4]. Each ANN has its own workflow and specific training sets are required for machine learning. (fig.1)

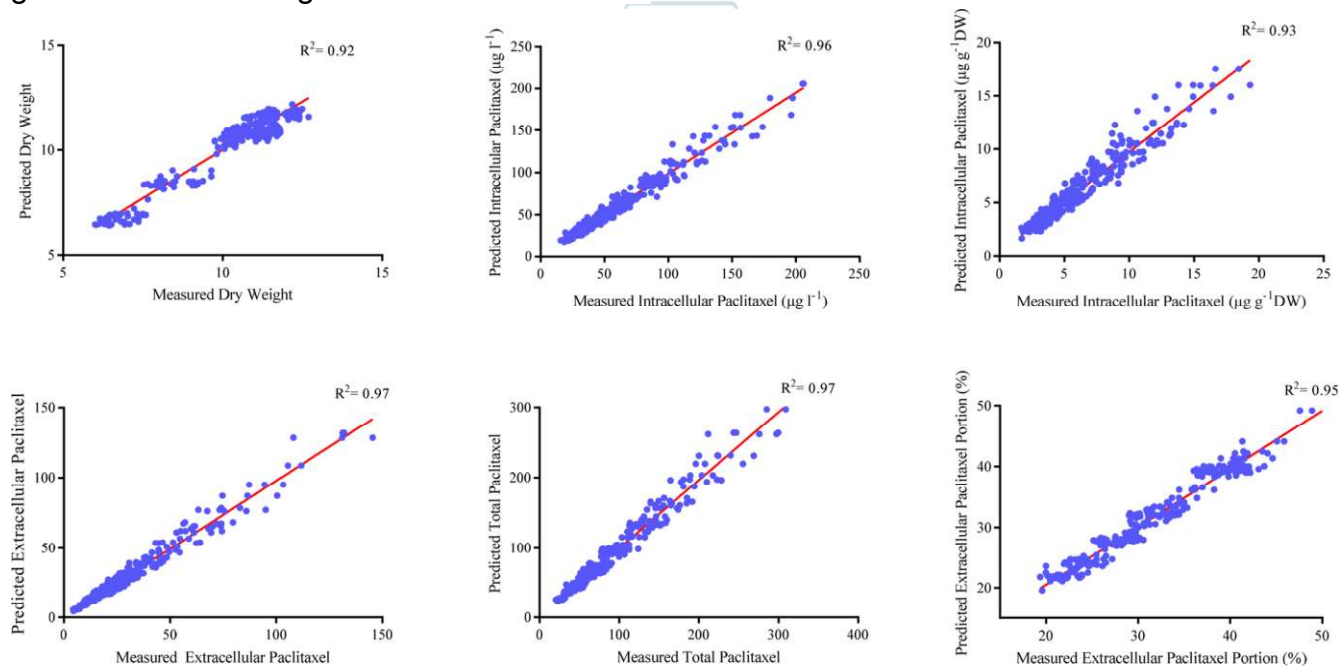
Conventional computational methods are inefficient for modelling secondary metabolite production in plants due to the requirement of multifactorial analysis, as mentioned above [3]. Problems requiring optimization, can be solved using search-based algorithms, like genetic algorithms (GAs) [5]. GAs are heuristic search algorithms that employ the concept of genetics and natural selection to provide solutions to problems, making GAs more intelligent than random search algorithms. For machine learning, it is a regression problem whenever the final output requires prediction of a quantity. PXT biosynthesis, being a non-linear regression problem, requires ANNs that can handle multifactorial analysis. The research group in the current study, had already tested machine learning algorithms such as multi-layer perceptron, genetic algorithm, adaptive neuro-fuzzy inference system to forecast and optimize paclitaxel biosynthesis with relative success. In the present study, the researchers expanded on their repertoire by testing General regression neural network (GRNN) optimized using fly optimization algorithm (FOA) [3].

GRNN is a special case of radial basis function (RBF) that has been successfully used to find difficult and unknown solutions in various fields. Due to its highly parallel structure and a memory-

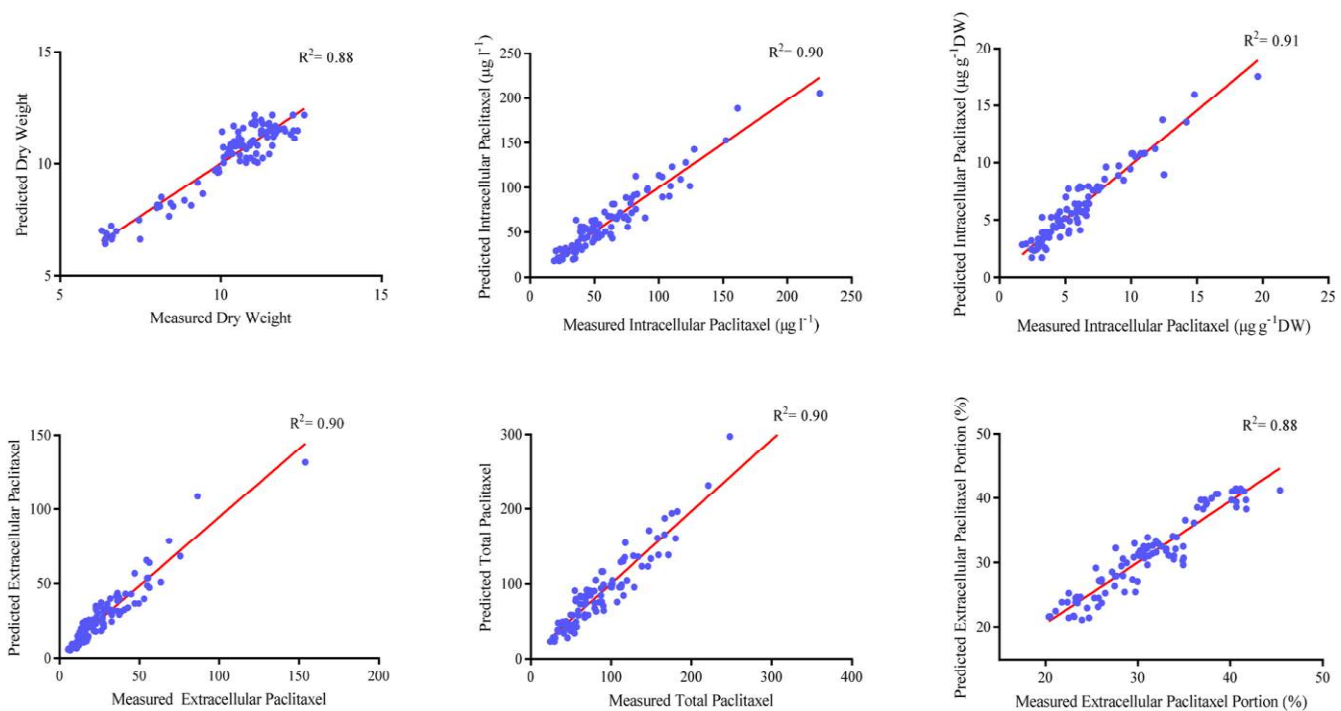
based robust network, GRNN has an ease of implementation. However, the challenge of the predictive performance depends on best estimate of smoothing (spread) parameter ( $\sigma$ ) in GRNN architecture. The research group for the current study used the Fly Optimization Algorithm (FOA) to solve the issue to finding the optimal smoothing parameter to be fed into GRNN. [3]. FOA is an evolutionary algorithm that uses fruit fly food searching behaviour to search for the best smoothing parameter to be fed into the GRNN architecture. Data from previous studies on PTX biosynthesis using different growth conditions were used as training sets for GRNN.

This study considered four different parameters as input variables into the ANN to generate forecasting data – concentration

levels of culture extract and culture filtrate, fungal elicitor adding day, and harvesting day of the culture. The success of growth conditions were evaluated using output variables to measure - dry weight (DW), intracellular ( $\mu\text{g g}^{-1}$  DW), intracellular ( $\mu\text{g l}^{-1}$ ), extracellular yield of PTX and total yield of PTX. GRNN-FOA was used to forecast the optimized input variables and expected the output variables of the predictions were validated against experimental data [3]. The quality of predictions made by the GRNN-FOA confirm the power of Artificial Intelligence in being able to provide solve multifactorial forecasting of growth and paclitaxel biosynthesis in hazel cell culture (fig 2&3).



**Figure 1:** Scatter plot of actual data against predicted values of dry weight, intracellular ( $\mu\text{g g}^{-1}$  DW), intracellular ( $\mu\text{g l}^{-1}$ ), extracellular and total yield of paclitaxel, and extracellular paclitaxel portion in *Corylus avellana* cell cultures using general regression neural network-fruit fly optimization algorithm (GRNN-FOA) models in training subset. The solid line shows fitted simple regression line on scatter points [Source: Plant Methods (2021)].



**Figure 2:** Scatter plot of actual data against predicted values of dry weight (DW), intracellular ( $\mu\text{g g}^{-1}\text{DW}$ ), intracellular ( $\mu\text{g l}^{-1}$ ), extracellular and total yield of paclitaxel and extracellular paclitaxel portion in *Corylus avellana*(hazel) cell cultures using general regression neural network-fruit fly optimization algorithm (GRNN-FOA) models in testing subset. The solid line shows fitted simple regression line on scatter points [Source: Plant Methods (2021)].

The data mining approach to of the present study improves plant biotechnology process using ANNs to help reduce the infrastructural cost of trials in laboratory setting. The research group also compared the prediction accuracy of GRNN vs the predictive power of multilayer perceptron-genetic algorithm (MLP-GA) to be able to select the most efficient ANN for future cases. The results of the study show great accordance between what the algorithm predicted and what was harvested as dry weight of PTX in hazel cells suspension cultures. One only has to look the

shift in the cancer care scenario during the pandemic to realise the importance of tools that offer ways to improve industrial scale production of an important chemotherapy drug like PTX. With many missed screenings, early-stage cancers would have been missed and the demand for PTX is bound to rise in the coming years. ANN driven boost of plant biotechnology methods are taking the ‘unknown’ out of production and replacing it with predicted outcomes and a promising future.

## References

1. Mubeen, S., He, CT., and Yang, ZY., Plant cell culture technology: An effective biotechnology tool for taxol and related taxanes production. J Cell Sci Ther 2018, 9 doi: <https://doi.org/10.4172/2157-7013-C2-04>.
2. Wilson, S.A., Keen, P., McKee, M.C. et al. Development of an Agrobacterium-mediated transformation method for Taxus suspension cultures. In Vitro Cell.Dev.Biol.-Plant 2018, 54, 36–44 doi: <https://doi.org/10.1007/s11627-017-9876-8>.
3. Salehi, M., Farhadi, S., Moieni, A. et al. A hybrid model based on general regression neural network and fruit fly optimization algorithm for forecasting and optimizing paclitaxel biosynthesis in Corylus avellana cell culture. Plant Methods 2021, 17, 13, doi:<https://doi.org/10.1186/s13007-021-00714-9>.
4. Wang H, Cimen E, Singh N, Buckler E. Deep learning for plant genomics and crop improvement. Curr Opin Plant Biol. 2020, 54:34-41. <https://doi.org/10.1016/j.pbi.2019.12.010>
5. <https://www.section.io/engineering-education/the-basics-of-genetic-algorithms-in-ml/>
6. D. F. Specht, "A general regression neural network," in IEEE Transactions on Neural Networks, vol. 2, no. 6, pp. 568-576, Nov. 1991, doi: [10.1109/72.97934](https://doi.org/10.1109/72.97934)



## Rights and Permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

**Creative Commons** This is an open access article distributed under the terms of the [Creative Commons CC BY](https://creativecommons.org/licenses/by/4.0/) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. You are not required to obtain permission to reuse this article. To request permission for a type of use not listed, please contact Biotechnology Kiosk.