Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Machine learning insights into the oxidative and cytotoxic effects of Ceratophyllum demersum extracts on HepG2 and THLE2 cells

Mustafa Ari
Karamanoglu Mehmetbey University: Karamanoglu Mehmetbey Universitesi

Bugrahan Emsen
bugrahanemsen@gmail.com

Karamanoglu Mehmetbey University: Karamanoglu Mehmetbey Universitesi https://orcid.org/0000-0002-9636-2596

Muhammet Dogan
Karamanoglu Mehmetbey University: Karamanoglu Mehmetbey Universitesi

Kagan Sarlar
Karamanoglu Mehmetbey University: Karamanoglu Mehmetbey Universitesi

---

Research Article

Keywords: anticancer, artificial intelligence, cell viability, oxidative stress, phytochemical

Posted Date: September 19th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7511345/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

---

Abstract

This study evaluates the cytotoxic, antioxidant, and oxidative effects of methanol and water extracts of *Ceratophyllum demersum* on hepatocellular carcinoma (HepG2) and normal hepatocyte (THLE2) cell lines. Lactate dehydrogenase (LDH) release assays revealed dose-dependent cytotoxicity in both cell lines. IC_{50} values for the water extract were calculated as 95.29 µg/mL for HepG2 and 165.44 µg/mL for THLE2. For methanol extract, IC_{50} values were 196.82 µg/mL (HepG2) and 271.87 µg/mL (THLE2). Total antioxidant status (TAS) levels peaked at low doses (12.5–50 µg/mL). Conversely, total oxidant status (TOS) levels rose sharply, with maximum oxidative stress observed at 200 µg/mL water extract in HepG2 cells. Machine learning analyses using random forest regression, gradient boosting regression, and extreme gradient boosting identified extract concentration and TOS as key predictors of LDH levels, with SHapley Additive exPlanations analysis supporting these findings. Machine learning revealed strong oxidative stress–cytotoxicity links, aiding phytochemical bioactivity analysis.

Introduction

Liver cancer remains one of the most prevalent and lethal malignancies worldwide, with hepatocellular carcinoma constituting approximately 75% of all primary liver cancer cases (Singh et al. 2025). Despite advances in diagnostic and therapeutic approaches, the prognosis for liver cancer patients remains dismal, largely due to late diagnosis, high recurrence rates, and resistance to conventional chemotherapeutic agents (Xu ve Hu 2024; Wei et al. 2025). As a result, there is an urgent need for novel, effective, and selective anticancer agents that can target liver cancer cells while sparing normal hepatocytes (Mossenta et al. 2022; Mostafaei et al. 2023).

In recent years, increasing attention has been directed towards the utilization of aquatic macrophytes, particularly submerged freshwater plants, for their potential pharmacological properties (Arya et al. 2022; Verma et al. 2023). *Ceratophyllum demersum* L., commonly known as coontail, is a free-floating submerged aquatic plant (Ade et al. 2024). Traditionally used in phytoremediation due to its ability to absorb heavy metals and organic pollutants, *C. demersum* is highly effective in environmental cleanup (Dogan et al. 2018; Aasim et al. 2023; Ugur ve Dogan 2024). It also produces a wide range of biologically active compounds, including phenolics, flavonoids, alkaloids, and terpenoids (Syed et al. 2018; Awati et al. 2020; Duong et al. 2025). Notably, valuable findings have been reported regarding the antioxidant activity of *C. demersum*, which is particularly significant given the well-established role of oxidative stress in cancer development and progression (Emsen ve Dogan 2018; Awati et al. 2021; Tokgoz et al. 2024). This suggests that its antioxidant potential may contribute to its yet unexplored anticancer effects (Pammi et al. 2023; Banu et al. 2024). However, its potential anticancer properties remain largely underexplored, especially in the context of human hepatic cell lines.

This study aims to evaluate, for the first time, the cytotoxic and antiproliferative properties of in vitro cultivated *C. demersum* extracts on both human hepatocellular carcinoma cell lines and normal liver cells. *In vitro* cultivation ensures chemical consistency of the plant material by eliminating environmental

variability and enables sustainable production of bioactive compounds without impacting natural ecosystems (Riahi et al. 2022; Ahlawat et al. 2024).

One of the distinguishing aspects of this research lies in its integrative approach combining experimental cytotoxicity assays with advanced machine learning (ML) methodologies. Traditional bioassays, while valuable, often lack the predictive power and scalability required for high-throughput screening and mechanistic understanding. Recent studies have successfully applied machine learning techniques to biological datasets, highlighting the potential of ML in enhancing biological research (Hassan et al. 2024; Liu et al. 2024; Bulut et al. 2025; Doga Atıcı et al. 2025; Sert et al. 2025). The driving force behind the interplay between machine learning and the natural sciences is the need to make sense of the vast troves of data amassed through experiments and observations. Analytic solutions often fall short for complex physical, chemical, and biological systems, whereas machine-learning models can uncover hidden patterns in these data and propose highly predictive, generalisable laws. Data-driven approaches therefore complement theoretical frameworks, allowing new hypotheses to be tested more rapidly and at lower cost. Because of this ML has become a chief catalyst shaping the speed and depth of modern scientific research (Carleo et al. 2019; Roscher et al. 2020). By incorporating machine learning techniques, this study aims to uncover complex patterns in the data that may elude conventional statistical methods. Another novel aspect of this study is the comparative analysis of *C. demersum*'s effect on normal versus cancerous liver cells. By simultaneously assessing the cytotoxic effects on both cell types, this research aims to evaluate the therapeutic index of the plant extracts and identify fractions with high selectivity for cancer cells, contributing to the development of more targeted anticancer strategies.

In summary, this study represents a multidisciplinary approach that integrates plant biotechnology, oncology, analytical chemistry, and artificial intelligence to explore the anticancer potential of *C. demersum*. The outcomes could not only contribute to liver cancer research but also establish a methodological framework for investigating other aquatic plant species and cancer types. Ultimately, this work underscores the importance of merging natural product research with computational tools to advance the development of safe, effective, and sustainable anticancer therapies.

Materials and methods

In vitro growth conditions

For micropropagation studies, *C. demersum* was surface sterilized. Internodal explants were excised from sterile stock plants under aseptic conditions in a laminar airflow cabinet. These explants were transferred into culture vessels containing appropriate nutrient media for in vitro cultivation. The basal medium consisted of Murashige and Skoog (MS) salts and vitamin formulation (Murashige ve Skoog 1962). In all treatments, the medium was supplemented with 0.5 mg/L 6-benzylaminopurine (BAP) as a plant growth regulator and 30 g/L sucrose as a carbon source. Cultures of C. demersum were maintained in liquid (agar-free) media. The pH of the media was adjusted to 5.6–5.8 using 1 N NaOH or

HCl before autoclaving at 121°C and 1.2 atm pressure for 20 minutes to ensure sterilization. For shoot regeneration, internodal explants (specifically the first and second nodes below the apical meristem) were placed in MS medium containing BAP. Cultures were maintained for eight weeks in a growth chamber under a 16-hour light/8-hour dark photoperiod, at 24 ± 1°C, with white LED light providing an intensity of 1500 lux (Fig. 1).

Preparation of plant extracts

Following a gentle heat treatment under a mild heat evaporator, the entirety of plant samples, underwent pulverization using liquid nitrogen and a mortar. These powder samples (10 g) were then extracted separately in water and methanol solvents with a Soxhlet extractor setup for 2 days. The extract was purified using Whatman No. 1 filter paper, and the remaining solvent was evaporated in a rotary evaporator at 40°C until dryness, followed by lyophilization to yield ultra-dry powders. The residues were stored in a sterile container at 4°C until analysis (Doga Atıcı et al. 2025). The abbreviations and yield percentages of the extracts obtained from mushroom samples are shown in Table 1.

Table 1
Yield (%) of obtained the plant extracts

Extract	Abbreviation of the extracts	Yield (w/w)
Methanol extract obtained from *C. demersum*	ME	11.12
Water extract obtained from *C. demersum*	WE	14.25

Preparation of cell cultures

Cancerous (HepG2) and normal (THLE2) human liver cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) with high glucose, L-glutamine, and sodium pyruvate, supplemented with 10% fetal bovine serum (FBS). During the culture process, a 5% CO_2 and 95% humidity incubator (Sanyo MCO 17AIC, USA) was used. When the cells reached the desired confluency (90%), they were detached using trypsin/EDTA and washed with fresh medium. The cells were then resuspended in fresh growth medium and counted using a trypan blue staining method with a cell counter device (TC-10, Bio-RAD, Germany) (Emsen et al. 2021).

Lactate dehydrogenase (LDH) release analysis

The related experiment was conducted using a commercial kit (LDH Cytotoxicity Assay Kit, Cayman Chemical, USA). Cells were seeded into 96-well flat-bottom microplates with 200 µL of culture medium per well. For background control, three wells contained only 200 µL of medium. To wells containing cells, 20 µL of 10% Triton X-100 solution was added to three wells to induce maximum release, while 20 µL of

Assay Buffer was added to another three wells for spontaneous release. In appropriate wells, 20 µL of the experimental cytotoxic agent was added, and all conditions were performed in triplicate. The plate was incubated at 37°C in a CO_2 incubator for 48 hours to induce cytotoxicity. After incubation, 100 µL of the cell supernatant was transferred to a new 96-well assay plate. Then, 100 µL of LDH Reaction Solution was added to each well, and the plate was gently shaken on an orbital shaker at 37°C for 30 minutes. Subsequently, absorbance was measured at 490 nm using a plate reader, and background absorbance (A490) values were subtracted from all wells (Emsen et al. 2019). LDH activity (%) was calculated using following formula:

$$\text{LDH activity (\%)} = \frac{\text{Sample absorbance (490 nm) - Spontaneous release absorbance (490 nm)}}{\text{Maximum release absorbance (490 nm) - Spontaneous release absorbance (490 nm)}} \times 100$$

Total antioxidant status (TAS) analysis

Cells treated with the extracts and incubated for 48 hours were removed from the incubator. The cells were treated with a lysis buffer (Cell Signaling Technology, USA) and incubated for 5 minutes. The cells were then harvested using a cell scraper and centrifuged at 14,000 × g. The resulting supernatants were used for the analysis. The Total Antioxidant Status (TAS) analysis was performed using a commercial kit (Total Antioxidant Status Assay Kit, Rel Assay Diagnostics, Türkiye). According to the kit protocol, 18 µL of each sample (supernatant), standard, and H_2O was transferred into a new 96-well assay plate. Then, 300 µL of Reagent 1 was added to each well and mixed by pipetting. After 30 seconds, the first spectrophotometric reading was taken at 660 nm. Next, 45 µL of Reagent 2 was added to each well and mixed again. The plate was incubated at room temperature for 10 minutes. Following the incubation, the second spectrophotometric reading was taken at 660 nm (Emsen et al. 2019). The TAS levels induced by the extracts on the cells were calculated according to following formulas:

ΔAbsorbance: Absorbance after second reading – Absorbance after first reading

$$\text{mmol Trolox equivalent/L} = \frac{(\Delta H_2O\ \text{Absorbance}) - (\Delta\text{Sample absorbance})}{(\Delta H_2O\ \text{Absorbance}) - (\Delta\text{Standard absorbance})}$$

Total oxidant status (TOS) analysis

Cells treated with the extracts and incubated for 48 hours were removed from the incubator. The cells were treated with a lysis buffer (Cell Signaling Technology, USA) and incubated for 5 minutes. The cells were then scraped using a cell scraper and centrifuged at 14,000 × g. The resulting supernatants were used for the analysis. The Total Oxidant Status (TOS) analysis was performed using a commercial kit (Total Oxidant Status Assay Kit, Rel Assay Diagnostics, Türkiye). According to the kit procedure, 45 µL of each sample (supernatant), standard, and H_2O was transferred into a new 96-well assay plate. Then, 300 µL of Reagent 1 was added to each well and mixed by pipetting. After 30 seconds, the first spectrophotometric reading was taken at 530 nm. Next, 15 µL of Reagent 2 was added to each well and

mixed again. The plate was incubated at room temperature for 10 minutes. After incubation, a second spectrophotometric reading was taken at 530 nm (Emsen et al. 2019). The TOS levels induced by the extracts on the cells were calculated according to following formulas:

ΔAbsorbance: Absorbance after second reading – Absorbance after first reading

$$\mu\text{mol } H_2O_2 \text{ equivalent/L} = \frac{(\Delta\text{Sample absorbance})}{(\Delta\text{Standard absorbance})} \times \text{Standard concentration } (10\ \mu\text{mol/L})$$

Machine-learning workflow

Data pre-processing

Categorical variables were transformed using one-hot encoding (Budholiya et al. 2022). Because decision trees and their derivatives (Random Forest, Gradient Boosting, XGBoost, etc.) are scale-invariant, explicit data scaling (normalisation or standardisation) is usually unnecessary. The threshold-based splits performed at each node depend solely on the relative ordering of values within the feature's own range; the unit or magnitude of a feature does not influence how the tree branches. Consequently, a wide-ranged variable such as "Concentration" does not receive undue priority over narrower-ranged variables like "TAS" or "TOS," keeping the model computationally efficient and its results easier to interpret.

Model development

Three ensemble-learning algorithms were employed in this study. Random Forest Regression (RFR), Gradient Boosting Regression (GBR) and Extreme Gradient Boosting (XGBoost). Each model was trained and evaluated on separate training and test splits and for each model, the hyper-parameters were optimised using grid search cross-validation. RFR predicts by building numerous decision trees and averaging their outputs. Hyper-parameters (e.g., number of trees, maximum depth) were tuned with grid search (Nguyen et al. 2021; Shi et al. 2023). GBR uses decision trees as weak learners and sequentially adds trees to reduce residual errors. Its hyper-parameters were optimised via grid search (Nguyen et al. 2021). XGBoost is an optimised implementation of gradient-boosted decision trees. Hyper-parameters were adjusted through grid search, genetic algorithms or Bayesian optimisation, and multithreading was leveraged to speed up training (Pan et al. 2022).

Hyper-parameter optimisation

Grid Search Cross-Validation (GridSearchCV) was used to fine-tune every model's hyper-parameter in this study. The method begins with a user-defined "grid" of candidate values (e.g., tree depth, learning rate, or regularisation strength). For each unique combination, the data are partitioned into *k* folds in this study, a 5-fold split in which 80% of the observations form the training folds while the remaining 20% constitute the validation fold. Each candidate model is then trained on four folds and scored on the fifth; the

process rotates until every fold has served once as the hold-out set. Because the procedure is embedded inside cross-validation, it simultaneously estimates out-of-sample performance and guards against over-fitting. After every grid point has been evaluated, GridSearchCV selects the parameter set that yields the highest mean validation score, thus maximising both accuracy and generalisation capability. In practice, the models tuned with this strategy consistently outperformed their untuned counterparts and produced more stable, reproducible results on the unseen test data (Chen et al. 2020).

Model interpretation

In feature-importance and SHAP analysis, several complementary techniques were employed to open the “black box” of the tree-based models and to quantify the relative influence of each predictor. (i) Tree-based feature importance ranks variables by the total reduction in node impurity (mean decrease in squared error) they achieve across all splits in the ensemble (Breiman 2001). Because these scores are computed during training, they impose virtually no additional cost and directly reflect the structure learned by the model; however, they can overstate the relevance of highly correlated or high-cardinality features. (ii) Permutation importance provides a model-agnostic counterpart: after the model is trained, each feature column is randomly permuted while the remaining data are left intact, and the ensuing increase in the cross-validated prediction error (ΔMSE) is recorded (Fisher et al. 2019). This procedure captures both linear and non-linear contributions and is less sensitive to multicollinearity, but it is computationally more demanding because it requires one additional inference pass per feature. (iii) SHapley Additive exPlanations (SHAP) offer a granular, game-theoretic decomposition of every individual prediction (Lundberg ve Lee 2017). Tree SHAP exploits the additive nature of decision-tree ensembles to compute exact Shapley values in polynomial time, thereby assigning to each feature an additive contribution that satisfies the axioms of local accuracy, consistency and missingness. Averaging the absolute Shapley values over all observations yields a global importance ranking, while the signed values reveal both the direction and magnitude of each feature’s effect for any given sample. Taken together, these three methods provide a robust hierarchy of predictors: tree-based gain delivers an inexpensive first estimate, permutation testing corroborates that ranking under perturbation, and SHAP affords interpretable, observation-level insight that links statistical importance to mechanistic hypotheses.

Performance evaluation

Each model was fitted to the training set and validated on the test set. To ensure a fair and methodologically consistent comparison, the hyper-parameter grids were sized so that each algorithm underwent a comparable number of cross-validation trials. Specifically, both the GBR and XGBoost grids comprised 288 candidate configurations, yielding 5 folds×288 = 1440 individual fits per model, while the more compact grid for RFR contained 360 candidates and thus 5 folds ×360 = 1800 fits. By keeping the total fit counts within the same order of magnitude, we avoided biasing the optimisation process toward any single algorithm and preserved the statistical balance required for reliable performance benchmarking. Performance of those three ML algorithm was tested with accuracy mean absolute error (MAE), root mean square error (RMSE) and the coefficient of determination (R^2) (Li et al. 2023; Shi et al.

2023). Feature importance analysis identified the variables to which each model assigned the greatest weight (Zhang et al. 2022).

Statistical analyses

The activities of the extracts were evaluated through a one-way ANOVA followed by the Duncan test. Probit regression analysis was applied to calculate the median inhibitory concentration (IC_{50}) values. To examine the relationships among LDH activity, TAS, and TOS activities across different celss, three-dimensional (3D) density analysis was performed. These statistical analyses were conducted using SPSS software (version 27.0, IBM Corporation, Armonk, NY, USA). Additionally, heatmap and hierarchical cluster analyses, employing Ward's minimum variance method, were used to identify similarities and differences among LDH activity, TAS, and TOS results. These analyses were executed in the Rstudio console with R software (version 4.1.0).

Results

Determination of LDH activity in cells

The LDH activities of HepG2 and THLE2 cells treated with different concentrations (12.5, 25, 50, 100, 200, and 400 µg/mL) of methanol and water extracts obtained from *C. demersum* were evaluated. Considering the studies conducted on HepG2 cells, the methanol extract of *C. demersum* caused a moderate increase in LDH activity. While the LDH activity was 4.11% ± 1.64 at a concentration of 12.5 µg/mL, it reached 69.04% ± 1.40 at 400 µg/mL. The water extract of *C. demersum*, on the other hand, showed an LDH activity of 10.39% ± 0.90 even at the lowest concentration (12.5 µg/mL), indicating its effectiveness at low doses. At 400 µg/mL, this value increased to 79.10% ± 1.16. These results reveal that both methanol and water extracts increased LDH activity in a concentration-dependent manner. Moreover, the water extract appears to trigger a higher level of LDH release in the cells (Fig. 2a).

Considering the studies conducted on THLE2 cells, the water extract of *C. demersum* revealed significant findings. Its effect reached 35.07% ± 1.50 at 100 µg/mL, 62.68% ± 1.46 at 200 µg/mL, and 67.28% ± 0.67 at 400 µg/mL. Overall, the plant led to a significant increase in LDH activity. While the methanol extract of *C. demersum* showed moderate effects, the water extract demonstrated more pronounced effects at higher concentrations (Fig. 2b).

According to the LDH assay results, the IC_{50} values of methanol and water extracts obtained from *C. demersum* on HepG2 and THLE2 cells were determined. The methanol extract exhibited high IC_{50} values in both cell lines (HepG2: 196.82 µg/mL, THLE2: 271.87 µg/mL). Additionally, based on the calculated selectivity index, the high values obtained for the water extracts indicated that they possess selective toxicity against cancer cells (Table 2).

Table 2
IC_{50} values (µg/mL) determined from LDH assay results in HepG2 and THLE2 cells treated with methanol and water extracts derived from *C. demersum*.

	HepG2 cells		**THLE2 cells**		
Treatment	IC_{50} (Limits)	Slope ± Standard error of the mean (Limits)	IC_{50} (Limits)	Slope ± Standard error of the mean (Limits)	Selectivity index
ME	196.82 (174.85–224.50)	1.47 ± 0.07 (1.32–1.62)	271.87 (239.04–314.88)	1.53 ± 0.08 (1.37–1.69)	1.38
WE	95.29 (86.49–105.29)	1.54 ± 0.07 (1.40–1.68)	165.44 (149.05–185.26)	1.58 ± 0.07 (1.43–1.73)	1.74

ME: Methanol extract of *C. demersum*; WE: Water extract of *C. demersum*.

Determination of changes in total antioxidant status (TAS) of the cells

TAS rates were calculated in mmol Trolox equivalent/L unit. Considering the studies conducted on HepG2 cells, the 12.5 µg/mL concentration of *C. demersum* methanol extract showed an activity level (4.83 ± 0.18) that was not significantly different from the control group ($p > 0.05$). A decrease was observed at 25 µg/mL (4.33 ± 0.20), and TAS further declined to 1.20 ± 0.24 at 50 µg/mL. At 400 µg/mL, TAS was determined as 1.62 ± 0.34. This indicates that the methanol extract of *C. demersum* is more effective at lower concentrations but its effect significantly decreases as the concentration increases. The 12.5 µg/mL application of the water extract of *C. demersum* exhibited a higher activity (4.88 ± 0.12) compared to the control group; however, this difference was not statistically significant ($p > 0.05$). At 50 µg/mL, the TAS decreased to 3.81 ± 0.21, and further dropped to 1.55 ± 0.38 at 400 µg/mL. These results suggest that the water extract is more effective at lower concentrations compared to the methanol extract but is less stable at higher concentrations. Overall, methanol extracts demonstrated comparable or higher TAS activity than the control group at low concentrations (12.5–25 µg/mL), but a marked decrease in activity was observed at higher concentrations (50–400 µg/mL). Water extracts showed more stable efficacy at low concentrations, while their activity declined as the concentration increased (Fig. 2c).

Considering the studies conducted on THLE2 cells, the value measured for the water extract of *C. demersum* at a concentration of 12.5 µg/mL was 6.11 ± 0.17, indicating a significantly lower biological activity compared to the control group ($p < 0.05$). At 50 µg/mL, the value recorded was 3.78 ± 0.19, which also represents a significant decrease relative to the control. At 400 µg/mL, the value measured was 2.65 ± 0.15, indicating that the reduction in biological activity becomes more pronounced at higher concentrations (Fig. 2d).

Determination of changes in total oxidant status (TOS) of the cells

TOS rates were calculated in µmol H_2O_2 equivalent/L unit. Considering the studies conducted on HepG2 cells, the methanol extract of *C. demersum* showed limited activity at low concentrations (e.g., 5.41 ± 0.11 at 12.5 µg/mL), while a significant increase was observed at higher concentrations (e.g., 12.87 ± 1.17 at 200 µg/mL). The water extract exhibited its most pronounced activity at 200 µg/mL (15.22 ± 0.85). This suggests that the water-based extract may be more effective in terms of TOS activity. Both methanol and water extracts demonstrated increased biological activity with rising concentrations; however, it was generally observed that the water extracts were more active at higher concentrations (Fig. 2e).

Considering the studies conducted on THLE2 cells, the methanol extract of *C. demersum* showed slight increases in activity compared to the control group at low concentrations. At 12.5 µg/mL, an elevated effect of 2.68 ± 0.21 was recorded, which was not significantly different from the control value. At medium concentrations (50 and 100 µg/mL), TOS values were significantly higher than the control group (3.25 ± 0.23 and 3.91 ± 0.18, respectively). At 200 and 400 µg/mL, the activity increased significantly, showing higher effects compared to the control group ($p < 0.05$) with values of 9.91 ± 0.69 and 9.77 ± 0.21, respectively. The water extract showed effects close to the control group at low concentrations. The TOS values at 12.5, 25, and 50 µg/mL (2.71 ± 0.20, 2.99 ± 0.22, and 2.62 ± 0.13, respectively) were higher than the control group but were determined to be statistically insignificant within the 95% confidence interval. Effects increased rapidly at medium and high concentrations. Significantly higher effects compared to the control were observed between 100 and 400 µg/mL. The application of *C. demersum* water extract at 400 µg/mL showed the highest value (10.80 ± 0.37) recorded throughout the entire TOS assay (Fig. 2f).

General heatmap, 3-D density, and correlation analysis for all activities

The biochemical and cellular effects of plant extracts applied at different concentrations on HepG2 and THLE2 cells were evaluated using cluster analysis and heatmap visualization. In the heatmap analysis conducted on HepG2 cells, the effects of the substances on LDH activity, TOS, and TAS were clearly observed. Cluster analysis identified three distinct groups, classifying the treatments based on similarities in their biological effects. High concentrations (200–400 µg/mL) were generally associated with biological activities represented by red color for LDH activity and TOS, and purple color for TAS, indicating potential cell stress or toxicity. Methanol extracts at 50 and 100 µg/mL showed similar results across all activities and clustered together. Overall, this analysis demonstrates that the responses of HepG2 cells to various biochemical parameters are both concentration-dependent and extract-specific. Notably, high concentrations may induce toxic effects, while lower doses could support cell viability. Additionally, cluster analysis of the activities showed a close relationship between LDH activity and TOS, which formed one cluster, whereas TAS formed a separate cluster (Fig. 3).

The results on THLE2 cells revealed that the samples were grouped into different clusters, which varied according to the concentration and effects of the extracts. Specifically, methanol extracts applied at high concentrations (200–400 µg/mL) showed stronger effects represented by the red color, which could be associated with increased LDH activity and TOS levels. This suggests that high concentrations may enhance cell stress and toxicity. On the other hand, extracts used at lower concentrations (12.5–50 µg/mL) were linked to lower toxicity levels, represented by the purple color. Notably, the control group and low concentration extracts clustered together, showing lower LDH activity. These findings indicate that the plant extracts used have concentration-dependent effects on THLE2 cells. While low concentrations appear to have protective effects on cell viability, higher concentrations may exert inhibitory effects on the cells (Fig. 3).

3D density and correlation analyses revealed significant and strong associations between LDH activity, TAS, and TOS in both HepG2 and THLE2 cell lines. In HepG2 cells, a strong negative correlation was observed between LDH activity and TAS ($r = -0.80$, $p < 0.01$). The surface plot showed a downward trend, where increasing TAS levels were associated with a distinct reduction in LDH activity. This indicates that elevated antioxidant capacity may help limit cellular damage in HepG2 cells. A very strong positive correlation was found between LDH activity and TOS ($r = 0.97$, $p < 0.01$). The corresponding surface plot displayed a steep upward slope, with a pronounced peak at high TOS levels, suggesting that increased oxidative stress leads to enhanced cell membrane damage. TAS and TOS were negatively correlated ($r = -0.78$, $p < 0.01$). This inverse relationship was reflected in the surface topography, where higher oxidant levels coincided with lower antioxidant levels, forming a smooth gradient with declining TAS values as TOS increased (Fig. 4).

In THLE2 cells, a similar strong negative correlation was detected between LDH activity and TAS ($r = -0.82$, $p < 0.01$). The surface graph revealed a descending pattern, where higher TAS values were associated with reduced LDH activity, suggesting that antioxidant mechanisms also provide protective effects in non-cancerous hepatic cells. LDH activity and TOS showed a strong positive correlation ($r = 0.96$, $p < 0.01$), with the surface plot exhibiting a sharp peak in LDH activity at elevated TOS levels. This further supports the role of oxidative stress in promoting cytotoxic effects. The negative correlation between TAS and TOS ($r = -0.72$, $p < 0.01$) was also evident in the surface plot, with TOS increasing in regions where TAS decreased, forming valleys and sloped declines in antioxidant levels (Fig. 4).

Machine-learning analyses

The models were compared on predictive accuracy and overall performance. The R^2, RMSE, and MAE values of the three ML algorithms used in this study are shown in Table 3. As shown in Table 3 GBR demonstrated the best performance with the highest R^2 value, the lowest RMSE, and the lowest MAE.

Table 3
Performance metrics (R^2, RMSE, and MAE) of RFR, GBR, and XGBoost algorithms

ML algorithms	R^2	RMSE	MAE
RFR	0.98	2.74	2.20
GBR	0.99	1.97	1.50
XGBoost	0.98	2.76	2.15

Figure 5 visually corroborates the error statistics summarised in Table 3. Across the full span of observed LDH activity values, the GBR points cluster most tightly around the 1: 1 reference line, mirroring its top-ranked R^2, RMSE and MAE. RFR and XGBoost reproduce the overall linear trend, yet their scatter is slightly broader: RFR exhibits minor over-prediction at the lower end of the scale, whereas XGBoost shows a modest under-prediction in the mid-range. These deviations visible as wider vertical spreads from the dashed line translate directly into the higher error metrics reported for those models. Taken together, the parity plots and Table 3 converge on the same conclusion: GBR delivers the most reliable and precise LDH activity forecasts in this data set, while RFR and XGBoost introduce small but systematic discrepancies.

From a machine-learning standpoint, all three models isolate the same two variables TOS and extract concentration as the most informative predictors of LDH activity. In the tree-based feature-importance bar charts, these variables receive the largest weights, indicating that they account for most of the reduction in prediction error achieved by successive splits (Fig. 6). Permutation importance, which measures the loss of accuracy when each variable is randomly shuffled, reaches its peak for the same two features, confirming their dominant role. By contrast, TAS forms a secondary tier, while solvent-type flags (extract water/methanol) and cell-line indicators (HepG2, THLE2) make only minor contributions shuffling them produces little change in the model's mean-squared error.

SHAP values provide a complementary, observation-level perspective and clarify the direction of each effect. Higher TOS or concentration values (depicted in warmer colours) shift the prediction upward, whereas lower values (cooler colours) pull it downward; TAS follows the same but weaker monotonic trend (Fig. 7). Solvent and cell-line variables cluster tightly around zero, suggesting that they fine-tune the output only under specific conditions. The narrow vertical spread of SHAP points for each feature indicates limited interaction effects, reinforcing the conclusion that oxidative status and dosage are the primary data-driven determinants of LDH activity in this study.

Discussion

In this study, the cytotoxic and oxidative responses of HepG2 and THLE2 cells to methanol and water extracts of *C. demersum* were systematically evaluated using multiple biochemical markers, including LDH activity, TAS, and TOS. The experimental findings demonstrated that both extracts induced

concentration-dependent alterations in cellular redox balance and viability, with water extracts generally exhibiting stronger effects, particularly at higher doses. HepG2 cells were more susceptible to extract-induced oxidative stress and membrane damage compared to THLE2 cells, suggesting potential selectivity toward cancerous cells. Moreover, the integration of multivariate analyses—including heatmap visualization, cluster analysis and correlation modeling—offered a comprehensive overview of the biochemical interactions, reinforcing the role of oxidative stress in mediating cytotoxicity. These results collectively contribute to a deeper understanding of *C. demersum*'s bioactive potential and its differential effects on cancerous versus normal hepatic cells.

The cytotoxic effects of methanol and water extracts of *C. demersum* on HepG2 and THLE2 cells were evaluated through LDH release assay. The results demonstrate a dose-dependent increase in LDH activity for both extracts, indicating a loss of membrane integrity and cell viability. This observation aligns with previous studies reporting LDH release as a reliable marker of cell membrane damage and cytotoxicity (Barathan et al. 2021; Khalef et al. 2024; Nazarkina et al. 2025)

In HepG2 cells, the water extract of *C. demersum* was particularly potent, eliciting a notable LDH release (10.39%) even at the lowest concentration tested (12.5 µg/mL), and reaching 79.10% at 400 µg/mL. This suggests that bioactive compounds in the water extract may exhibit strong cytotoxic effects at lower doses. Comparatively, the methanol extract induced moderate LDH activity, indicating a weaker cytotoxic profile. These results are consistent with studies indicating that water extracts often retain more hydrophilic phenolic and flavonoid compounds with pronounced bioactivity (Premathilaka et al. 2022; Plaskova ve Mlcek 2023).

In THLE2 cells, the water extract also induced LDH release, particularly at higher concentrations (200–400 µg/mL) though the magnitude of response was lower than in HepG2 cells. This difference suggests that the extract may exhibit selective cytotoxicity toward cancerous cells—a desirable property in anticancer drug development (Grigoletto et al. 2021; Yuan et al. 2021). The selectivity index values further support this, indicating that the water extract exhibits preferential toxicity to HepG2 cells over normal THLE2 cells. The higher IC_{50} values observed for the methanol extract in both HepG2 (196.82 µg/mL) and THLE2 (271.87 µg/mL) cell lines further support its lower cytotoxic potency. Previous research has shown that methanol extracts may contain less active or more lipophilic compounds, which may be less effective in disrupting cancer cell function (Anoor et al. 2022). By contrast, the water extract likely contains phenolic acids, flavonoids, and possibly alkaloids that could disrupt mitochondrial function and trigger apoptotic signaling via oxidative stress (Zheng et al. 2022; Albahri et al. 2024). Similar cytotoxic effects have been observed in other aquatic plant extracts, such as *Hydrilla verticillata* and *Eichhornia crassipes*, which have been shown to induce apoptosis in HepG2 cells via ROS generation and caspase activation (Eden et al. 2023; Tang et al. 2024). These studies support the idea that aquatic macrophytes like *C. demersum* can be rich sources of cytotoxic and antioxidant compounds. Collectively, our findings demonstrate that the water extract of *C. demersum* exhibits significant and selective cytotoxicity toward HepG2 cells, suggesting its potential as a natural antitumor

agent. The data also highlight the importance of extraction solvent in determining the biological activity of plant materials, as the water extract consistently outperformed the methanol extract in this study.

The TAS assay results revealed significant insights into the antioxidant behavior of *Ceratophyllum demersum* extracts in both HepG2 and THLE2 cells. Overall, both methanol and water extracts exhibited a concentration-dependent decrease in antioxidant status, particularly at higher concentrations, indicating a potential shift from antioxidant to pro-oxidant behavior as dose increases. In HepG2 cells, the methanol extract exhibited a TAS value of 4.83 ± 0.18 at 12.5 µg/mL, which was not significantly different from the control. However, as the concentration increased to 50 µg/mL and 400 µg/mL, the TAS values sharply declined (1.20 ± 0.24 and 1.62 ± 0.34, respectively). This decline suggests that while low concentrations of methanol extract may contribute to antioxidant defense, higher concentrations could impair redox balance or potentially act through cytotoxic oxidative mechanisms. Similar biphasic effects of plant extracts have been previously reported, where phenolic compounds exhibit antioxidant properties at low concentrations but generate reactive species at high doses (Xi et al. 2022; Doga Atıcı et al. 2025). The water extract followed a similar trend. Although the 12.5 µg/mL application showed slightly higher TAS than the control (4.88 ± 0.12), this difference was not statistically significant. At 400 µg/mL, TAS dropped to 1.55 ± 0.38, indicating that the antioxidant activity of the water extract also declines substantially at high concentrations. These results support the hypothesis that some compounds in *C. demersum* extracts may become pro-oxidant at elevated levels, a behavior commonly observed in polyphenolic-rich extracts (Emsen ve Dogan 2018; Masłyk et al. 2024).

In THLE2 cells, a more pronounced reduction in TAS was observed with the water extract, especially at 12.5 µg/mL (6.11 ± 0.17), which was significantly lower than the control group ($p < 0.05$). The antioxidant capacity continued to decline at higher concentrations (3.78 ± 0.19 at 50 µg/mL and 2.65 ± 0.15 at 400 µg/mL). These findings suggest that normal hepatic cells may be more sensitive to oxidative changes induced by *C. demersum* water extract, particularly in the presence of hydrophilic bioactives that may affect redox homeostasis (Aravind ve Prasad 2005). Interestingly, while both extracts demonstrated antioxidant activity at lower doses, their efficacy decreased markedly with increased concentrations. This trend may be attributed to the auto-oxidation of polyphenols, redox cycling of flavonoids, or cellular stress responses to phytochemicals, as reported in other plant-based studies (Aravind ve Prasad 2005; Kumar et al. 2023). Taken together, these results suggest that *C. demersum* extracts possess dose-sensitive antioxidant activity, with potential antioxidant benefits at low concentrations and possible pro-oxidant or cytotoxic effects at higher doses. These findings are in line with the dual role of phytochemicals, which can act as either protectors or stress inducers depending on their concentration and cellular environment (Khouzani ve Taghizadeh 2024; Muscolo et al. 2024).

TOS assay results provide important insights into the oxidative potential of *C. demersum* extracts in both HepG2 and THLE2 cells. The findings reveal a dose-dependent increase in TOS, particularly with water extracts, indicating an elevation in oxidative stress markers as extract concentration increases. In HepG2 cells, the methanol extract showed only a modest increase in TOS at low concentrations (e.g., 5.41 ± 0.11 at 12.5 µg/mL), but a significant elevation was observed at higher doses, reaching 12.87 ± 1.17 at

200 µg/mL. The water extract exhibited its most pronounced oxidant activity at the same concentration (15.22 ± 0.85), suggesting that the water-based phytochemicals may possess stronger pro-oxidant effects at higher concentrations. This pattern aligns with studies reporting that polyphenolic-rich plant extracts, while antioxidant at low doses, can induce oxidative stress at high concentrations through redox cycling or mitochondrial disruption (Castejon-Vega et al. 2023; Chen et al. 2023). In THLE2 cells, a similar trend was observed. The methanol extract showed slight, non-significant increases in TOS at low concentrations, but significant elevations were recorded at 100–400 µg/mL. The water extract also remained close to control levels at low doses but induced a sharp increase in TOS at medium to high concentrations, reaching the highest overall value of 10.80 ± 0.37 at 400 µg/mL. These findings suggest that both extracts can enhance oxidative stress in a concentration-dependent manner, but the water extract may have a greater oxidative potential, particularly in normal liver epithelial cells.

Such a dual behavior is characteristic of plant extracts containing flavonoids, phenolic acids, and tannins, which can act as either antioxidants or pro-oxidants depending on the cellular context, redox balance, and compound concentration (Rudrapal et al. 2022; Lescano et al. 2025). In the context of cancer therapy, this behavior may be beneficial, as selective pro-oxidant stress in cancer cells can trigger apoptosis, especially in redox-sensitive cancers like hepatocellular carcinoma (Krakowiak ve Pietrasik 2023; Catalano et al. 2025). However, the similar oxidative effects observed in THLE2 cells at higher doses raise concerns about cytotoxicity to non-cancerous tissue, and highlight the need for careful dose optimization. Notably, the sharp increase in TOS at high concentrations of water extract supports previous observations from the TAS assay that these extracts may transition from antioxidant to pro-oxidant behavior depending on dose. The oxidative stress generated at high concentrations may underlie the cytotoxic effects observed in LDH assays, especially in HepG2 cells, further suggesting that *C. demersum* contains bioactives with redox-modulating potential.

ML analysis provided valuable insights into the predictive dynamics of LDH activity in relation to biochemical and experimental variables. Among the three algorithms employed—GBR, RFR, and XGBoost—GBR outperformed the others, as evidenced by its superior R^2, RMSE, and MAE values. This suggests that GBR's ensemble structure and sequential learning approach offer enhanced modeling capacity for complex, nonlinear biological data, in line with previous reports emphasizing GBR’s efficiency in biomedical prediction tasks (Li et al. 2024; Mim et al. 2025).

Feature importance metrics consistently identified TOS and extract concentration as the most influential predictors of LDH activity across all models, underscoring the central role of oxidative stress and dose-response relationships in cytotoxicity mechanisms. These findings are further supported by permutation importance and SHAP analyses, which highlighted strong positive associations between increased TOS/concentration and elevated LDH activity—indicating enhanced cellular membrane damage at higher oxidative stress levels. TAS (total antioxidant status), while showing a weaker effect, maintained an inverse relationship, consistent with its proposed protective role against oxidative damage. Notably, solvent type (water vs. methanol) and cell line identity (HepG2 vs. THLE2) contributed minimally to the predictive models, suggesting that cytotoxic responses were largely governed by redox-related variables

rather than intrinsic differences between extract type or cell line. This is particularly important as it indicates that the biochemical environment, rather than the delivery matrix, drives LDH release, aligning with biochemical assay results and highlighting the robustness of ML interpretability in biological systems (Alves et al. 2021; Wang et al. 2024).

Together, these results validate the application of interpretable machine learning as a complementary tool in biomedical research. By identifying the most influential variables and quantifying their individual contributions, ML approaches not only improve predictive accuracy but also deepen mechanistic understanding—especially in multi-parametric systems involving dose, redox status, and cellular context.

Conclusions

The findings of this study demonstrate that *C. demersum* exhibits significant and concentration-dependent biological effects on both HepG2 hepatocellular carcinoma cells and non-tumorigenic THLE2 liver epithelial cells. The LDH assay revealed that both methanol and water extracts increased membrane damage, with the water extract inducing a higher cytotoxic response at elevated doses, particularly in HepG2 cells. Antioxidant evaluations through TAS assays indicated that low concentrations of the extracts, especially the water-based ones, maintained or slightly enhanced the antioxidant capacity of the cells; however, higher doses led to a marked decline, suggesting oxidative imbalance. Correspondingly, TOS assays confirmed that oxidative stress levels increased significantly with concentration, with the water extract producing the highest oxidant response at 400 µg/mL. These results point to a dual antioxidant/pro-oxidant nature of *C. demersum*, where low doses may offer protective or therapeutic benefits, while higher concentrations elicit oxidative and cytotoxic stress. Importantly, the selective toxicity observed—especially in HepG2 cells—suggests potential anticancer activity, though the effects on normal cells warrant cautious dose optimization for therapeutic applications.

This study compared three ensemble regression models under equal cross-validation settings to ensure unbiased performance evaluation. Among the models, the gradient boosting approach showed the highest predictive accuracy, with excellent performance metrics on the independent test set. The results indicate that this method is more effective in capturing the underlying structure of the data compared to bagging or second-order boosting techniques, within the limits of the current dataset. Interpretability analyses consistently highlighted oxidative status and extract concentration as the most influential factors affecting the target variable, while antioxidant status played a secondary role. Other variables, such as solvent type and cell-line information, had minimal impact. Observation-level explanations further confirmed that high oxidative status and concentration values tend to increase predicted values, with little evidence of complex interactions among features. In summary, this work demonstrates that (i) balanced model tuning prevents selection bias, (ii) gradient boosting delivers the most reliable predictions, and (iii) combining multiple interpretability tools yields coherent insights into model behavior. Despite the limited sample size and potential multicollinearity, the findings offer a solid foundation for emphasizing oxidative markers and dosage in future studies.

Declarations

Funding This work was supported by the Karamanoğlu Mehmetbey University Scientific Research Projects Commission (Project No. 11-YL-23).

Data availability The whole datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethical approval This article does not contain any studies with human participants or animals (vertebrates) performed by any of the authors.

Ethical responsibility The submitted work is original and has not been published elsewhere in any form or language.

Competing interests The authors have no relevant financial or nonfinancial interests to disclose.

References

1. Aasim M, Ali SA, Aydin S, et al (2023) Artificial intelligence–based approaches to evaluate and optimize phytoremediation potential of in vitro regenerated aquatic macrophyte *Ceratophyllum demersum* L. Environ Sci Pollut Res 30:40206–40217
2. Ade S, Sanubi JO, Obakanurhe O, et al (2024) Evaluating growth performance, haematological indices, and serological responses in weaner rabbits fed flesh coontail leaf (*Ceratophyllum demersum*). J Agric Rural Dev Trop Subtrop 125:219–224
3. Ahlawat YK, Yadav K, Samani M, Chaudhary D (2024) Harnessing in-vitro propagation for the sustainable conservation of medicinal plants: challenges and prospects. Içinde: Medicinal and Aromatic Plants: Current Research Status, Value-Addition to Their Waste, and Agro-Industrial Potential. Springer, ss 27–37
4. Albahri G, Badran A, Abdel Baki Z, et al (2024) Potential anti-tumorigenic properties of diverse medicinal plants against the majority of common types of cancer. Pharmaceuticals 17:574
5. Alves MA, Castro GZ, Oliveira BAS, et al (2021) Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. Comput Biol Med 132:104335
6. Anoor PK, Yadav AN, Rajkumar K, et al (2022) Methanol extraction revealed anticancer compounds quinic acid, 2 (5H)-furanone and phytol in *Andrographis paniculata*. Mol Clin Oncol 17:151
7. Aravind P, Prasad MNV (2005) Modulation of cadmium-induced oxidative stress in *Ceratophyllum demersum* by zinc involves ascorbate–glutathione cycle and glutathione metabolism. Plant Physiol Biochem 43:107–116
8. Arya AK, Durgapal M, Bachheti A, et al (2022) Ethnomedicinal use, phytochemistry, and other potential application of aquatic and semiaquatic medicinal plants. Evidence-Based Complement Altern Med 2022:4931556. https://doi.org/10.1155/2022/4931556

9. Awati SS, Gilhotra RM, Singh SK, et al (2020) Plant profile, phytochemistry and pharmacological properties of *Ceratophyllum demersum* Linn.: a review. SGVU J Pharm Res Educ 5:409–422
10. Awati SS, Singh SK, Wadkar KA (2021) In vitro antioxidant potential and anticancer activity of *Ceratophyllum demersum* Linn. extracts on HT-29 human colon cancer cell line. Res J Pharm Technol 14:28–36. https://doi.org/10.5958/0974-360X.2021.00006.8
11. Banu HS, Parvin I sheriff, Priyadharshini SD, et al (2024) Molecular insights into the antioxidant and anticancer properties: a comprehensive analysis through molecular modeling, docking, and dynamics studies. J Cell Biochem 125:e30564
12. Barathan M, Zulpa AK, Hoong SM, et al (2021) Synergistic effect of hyperforin and paclitaxel on growth inhibition, apoptotic mediator activation in MCF-7 human breast cancer cells. J Taibah Univ Sci 15:918–927. https://doi.org/10.1080/16583655.2021.2010910
13. Breiman L (2001) Random forests. Mach Learn 45:5–32
14. Budholiya K, Shrivastava SK, Sharma V (2022) An optimized XGBoost based diagnostic system for effective prediction of heart disease. J King Saud Univ Inf Sci 34:4514–4523
15. Bulut S, Aasim M, Emsen B, et al (2025) Machine learning modeling and response surface methodology driven antioxidant and anticancer activities of chitosan nanoparticle-mediated extracts of *Bacopa monnieri*. Int J Biol Macromol 310:143470. https://doi.org/10.1016/j.ijbiomac.2025.143470
16. Carleo G, Cirac I, Cranmer K, et al (2019) Machine learning and the physical sciences. Rev Mod Phys 91:45002
17. Castejon-Vega B, Cordero MD, Sanz A (2023) How the disruption of mitochondrial redox signalling contributes to ageing. Antioxidants 12:831
18. Catalano T, Selvaggi F, Cotellese R, Aceto GM (2025) The role of reactive oxygen species in colorectal cancer initiation and progression: perspectives on theranostic approaches. Cancers (Basel) 17:752
19. Chen M, Chen Y, Zhu W, et al (2023) Advances in the pharmacological study of Chinese herbal medicine to alleviate diabetic nephropathy by improving mitochondrial oxidative stress. Biomed Pharmacother 165:115088
20. Chen R-C, Dewi C, Huang S-W, Caraka RE (2020) Selecting critical features for data classification based on machine learning methods. J Big Data 7:52
21. Doga Atıcı Y, Dogan M, Emsen B, Yıldırım H (2025) Machine learning-assisted evaluation of antioxidant and metal chelating capacities in in vitro propagated *Ceratophyllum demersum* L. under different LED light conditions. Plant Cell Tissue Organ Cult 161:41
22. Dogan M, Karatas M, Aasim M (2018) Cadmium and lead bioaccumulation potentials of an aquatic macrophyte *Ceratophyllum demersum* L.: a laboratory study. Ecotoxicol Environ Saf 148:431–440
23. Duong T-H, Aree T, Dang V-S, et al (2025) Chemical constituents with their alpha-glucosidase inhibitory activity from the whole plant of *Ceratophyllum demersum*. Phytochemistry 229:114290

24. Eden WT, Wahyuono S, Cahyono E, Astuti P (2023) Phytochemical, antioxidant, and cytotoxic activity of water hyacinth (*Eichhornia crassipes*) ethanol extract. Trop J Nat Prod Res 7:3606
25. Emsen B, Dogan M (2018) Evaluation of antioxidant activity of in vitro propagated medicinal *Ceratophyllum demersum* L. extracts. Acta Sci Pol Cultus 17:23–33. https://doi.org/10.24326/asphc.2018.1.3
26. Emsen B, Ozdemir O, Engin T, et al (2019) Inhibition of growth of U87MG human glioblastoma cells by *Usnea longissima* Ach. An Acad Bras Cienc 91:e20180994. https://doi.org/10.1590/0001-3765201920180994.1
27. Emsen B, Sadi G, Bostanci A, Aslan A (2021) In vitro evaluation of cytotoxic, oxidative, genotoxic, and apoptotic activities of physodic acid from *Pseudevernia furfuracea* in HepG2 and THLE2 cells. Plant Biosyst 155:1111–1120. https://doi.org/10.1080/11263504.2020.1852329
28. Fisher A, Rudin C, Dominici F (2019) All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J Mach Learn Res 20:1–81
29. Grigoletto A, Martinez G, Gabbia D, et al (2021) Folic acid-targeted paclitaxel-polymer conjugates exert selective cytotoxicity and modulate invasiveness of colon cancer cells. Pharmaceutics 13:929
30. Hassan J, Saeed SM, Deka L, et al (2024) Applications of machine learning (ML) and mathematical modeling (MM) in healthcare with special focus on cancer prognosis and anticancer therapy: current status and challenges. Pharmaceutics 16:260
31. Khalef L, Lydia R, Filicia K, Moussa B (2024) Cell viability and cytotoxicity assays: biochemical elements and cellular compartments. Cell Biochem Funct 42:e4007
32. Khouzani ZK, Taghizadeh M (2024) In vitro application of *Methyl jasmonate* to induce phytochemical production and antioxidant activity in *Dracocephalum polychaetum* Bornm. cells. Plant Cell Tissue Organ Cult 159:47. https://doi.org/10.1007/s11240-024-02865-8
33. Krakowiak A, Pietrasik S (2023) New insights into oxidative and reductive stress responses and their relation to the anticancer activity of selenium-containing compounds as hydrogen selenide donors. Biology (Basel) 12:875
34. Kumar H, Dhalaria R, Guleria S, et al (2023) Anti-oxidant potential of plants and probiotic spp. in alleviating oxidative stress induced by H2O2. Biomed Pharmacother 165:115022
35. Lescano L, Cziaky Z, Custodio L, Rodrigues MJ (2025) Yeast extract elicitation enhances growth and metabolite production in *Limonium algarvense* callus cultures. Plant Cell Tissue Organ Cult 160:45. https://doi.org/10.1007/s11240-025-02991-x
36. Li H, Wang Y, Zhang J, et al (2023) Prediction of the freshness of horse mackerel (*Trachurus japonicus*) using E-nose, E-tongue, and colorimeter based on biochemical indexes analyzed during frozen storage of whole fish. Food Chem 402:134325
37. Li Y, Cui X, Yang X, et al (2024) Artificial intelligence in predicting pathogenic microorganisms' antimicrobial resistance: challenges, progress, and prospects. Front Cell Infect Microbiol 14:1482186

38. Liu M, Srivastava G, Ramanujam J, Brylinski M (2024) Augmented drug combination dataset to improve the performance of machine learning models predicting synergistic anticancer effects. Sci Rep 14:1668
39. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Içinde: Advances in Neural Information Processing Systems. ss 1–10
40. Masłyk M, Lenard T, Olech M, et al (2024) *Ceratophyllum demersum* the submerged macrophyte from the mining subsidence reservoir Nadrybie Poland as a source of anticancer agents. Sci Rep 14:6661
41. Mim MA, Khatun MR, Hossain MM, Rahman W (2025) Advanced PCF-SPR biosensor design and performance optimization using machine learning techniques. Optik (Stuttg) 172391
42. Mossenta M, Busato D, Dal Bo M, et al (2022) Novel nanotechnology approaches to overcome drug resistance in the treatment of hepatocellular carcinoma: glypican 3 as a useful target for innovative therapies. Int J Mol Sci 23:10038
43. Mostafaei F, Mahdinloo S, Valizadeh H, et al (2023) An update review of smart nanotherapeutics and liver cancer: opportunities and challenges. Nanomedicine 18:1855–1873
44. Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15:473
45. Muscolo A, Mariateresa O, Giulio T, Mariateresa R (2024) Oxidative stress: the role of antioxidant phytochemicals in the prevention and treatment of diseases. Int J Mol Sci 25:3264
46. Nazarkina ZK, Stepanova AO, Savostyanova TA, Laktionov PP (2025) Application of a modified lactate dehydrogenase assay to evaluate the viability of cells cultured on 3D scaffolds when commonly used assays fail. Cytotechnology 77:9
47. Nguyen H, Vu T, Vo TP, Thai H-T (2021) Efficient machine learning models for prediction of concrete strengths. Constr Build Mater 266:120950
48. Pammi SSS, Suresh B, Giri A (2023) Antioxidant potential of medicinal plants. J Crop Sci Biotechnol 26:13–26
49. Pan S, Zheng Z, Guo Z, Luo H (2022) An optimized XGBoost method for predicting reservoir porosity using petrophysical logs. J Pet Sci Eng 208:109520
50. Plaskova A, Mlcek J (2023) New insights of the application of water or ethanol-water plant extract rich in active compounds in food. Front Nutr 10:1118761
51. Premathilaka R, Rashidinejad A, Golding M, Singh J (2022) Oral delivery of hydrophobic flavonoids and their incorporation into functional foods: Opportunities and challenges. Food Hydrocoll 128:107567
52. Riahi L, Chograni H, Ben Rejeb F, et al (2022) Efficient in vitro regeneration of the endangered species *Artemisia arborescens* L. through direct organogenesis and impact on secondary metabolites production. Hortic Environ Biotechnol 63:439–450

53. Roscher R, Bohn B, Duarte MF, Garcke J (2020) Explainable machine learning for scientific insights and discoveries. Ieee Access 8:42200–42216
54. Rudrapal M, Khairnar SJ, Khan J, et al (2022) Dietary polyphenols and their role in oxidative stress-induced human diseases: insights into protective effects, antioxidant potentials and mechanism(s) of action. Front Pharmacol 13:806470
55. Sert E, Kalındemirtaş FD, Karakuş EE, et al (2025) Machine learning regression for assessing sensing performance and anticancer potential of oolong tea-derived silver nanoparticles. J Electrochem Soc 172:37503
56. Shi Z, Du L, He X, et al (2023) Prediction model of yield strength of V–N steel hot-rolled plate based on machine learning algorithm. JOM 75:1750–1762
57. Singh SP, Madke T, Chand P (2025) Global epidemiology of hepatocellular carcinoma. J Clin Exp Hepatol 15:102446
58. Syed I, Fatima H, Mohammed A, Siddiqui MA (2018) *Ceratophyllum demersum* a free-floating aquatic plant: a review. Indian J Pharm Biol Res 6:10–17
59. Tang Y, Liang F, Yan Y, et al (2024) Purification and identification of peptides from *Hydrilla verticillata* (Linn. f.) Royle with cytoprotective and antioxidative effect against H2O2-treated HepG2 Cells. J Agric Food Chem 72:4170–4183
60. Tokgoz A, Emsen B, Dogan M (2024) The efficacy of allelopathy of select lichens on antioxidant potential and multiplication of *Ceratophyllum demersum* L. Plant Cell Tissue Organ Cult 159:26. https://doi.org/10.1007/s11240-024-02894-3
61. Ugur K, Dogan M (2024) Effectiveness of light-emitting diodes for arsenic and mercury accumulation by *Ceratophyllum demersum* L.: an innovative advancement in phytoremediation technology. Chemosphere 358:142064
62. Verma S, Nakin MD V, Makhosayafana Z, Lall N (2023) The role of aquatic plants in natural products and drug discovery. Lett Drug Des Discov 20:386–407
63. Wang S, Shen Y, Zeng F, et al (2024) Exploiting biochemical data to improve osteosarcoma diagnosis with deep learning. Heal Inf Sci Syst 12:31
64. Wei F, Huang P, Zhang B, et al (2025) Machine learning analysis identified NNMT as a potential therapeutic target for hepatocellular carcinoma based on PCD-related genes. Sci Rep 15:7494
65. Xi X, Wang J, Qin Y, et al (2022) The biphasic effect of flavonoids on oxidative stress and cell proliferation in breast cancer cells. Antioxidants 11:622
66. Xu X, Hu L (2024) Impact of anesthesia-related genes on prognosis and tumor microenvironment in hepatocellular carcinoma: a comprehensive analysis. Environ Toxicol 39:4700–4711
67. Yuan K, Wang X, Dong H, et al (2021) Selective inhibition of CDK4/6: a safe and effective strategy for developing anticancer drugs. Acta Pharm Sin B 11:30–54
68. Zhang W, Li H, Han L, et al (2022) Slope stability prediction using ensemble learning techniques: a case study in Yunyang County, Chongqing, China. J Rock Mech Geotech Eng 14:1089–1099

69. Zheng Z, Zhang L, Hou X (2022) Potential roles and molecular mechanisms of phytochemicals against cancer. Food Funct 13:9208–9225

Figures

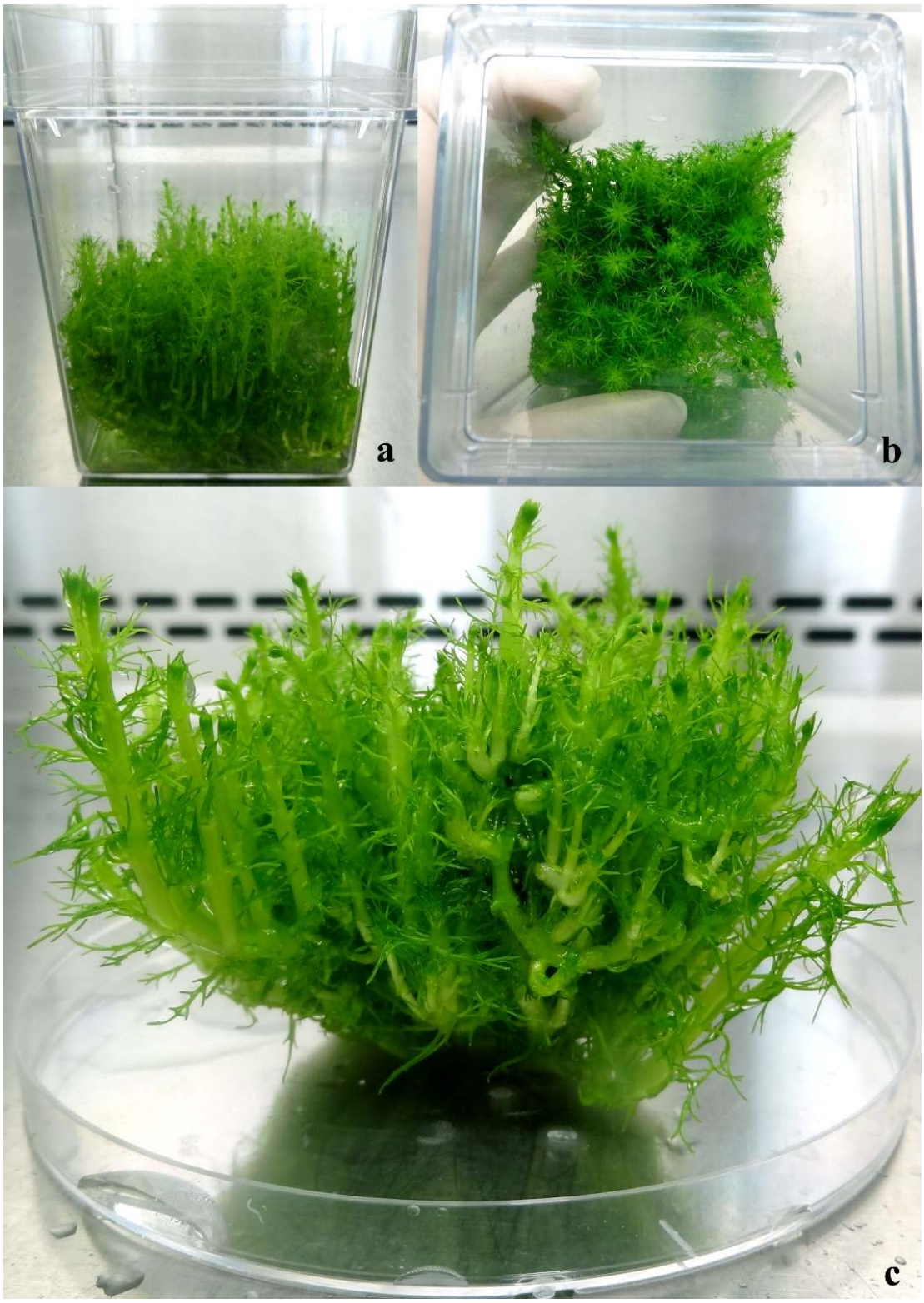

Figure 1

Shoot development of *C. demersum* after eight weeks of in vitro culture. Dense shoot proliferation within the culture vessel (a), top view showing the compact structure of the regenerated shoots (b), and overall appearance of well-developed and healthy shoots in a Petri dish (c).

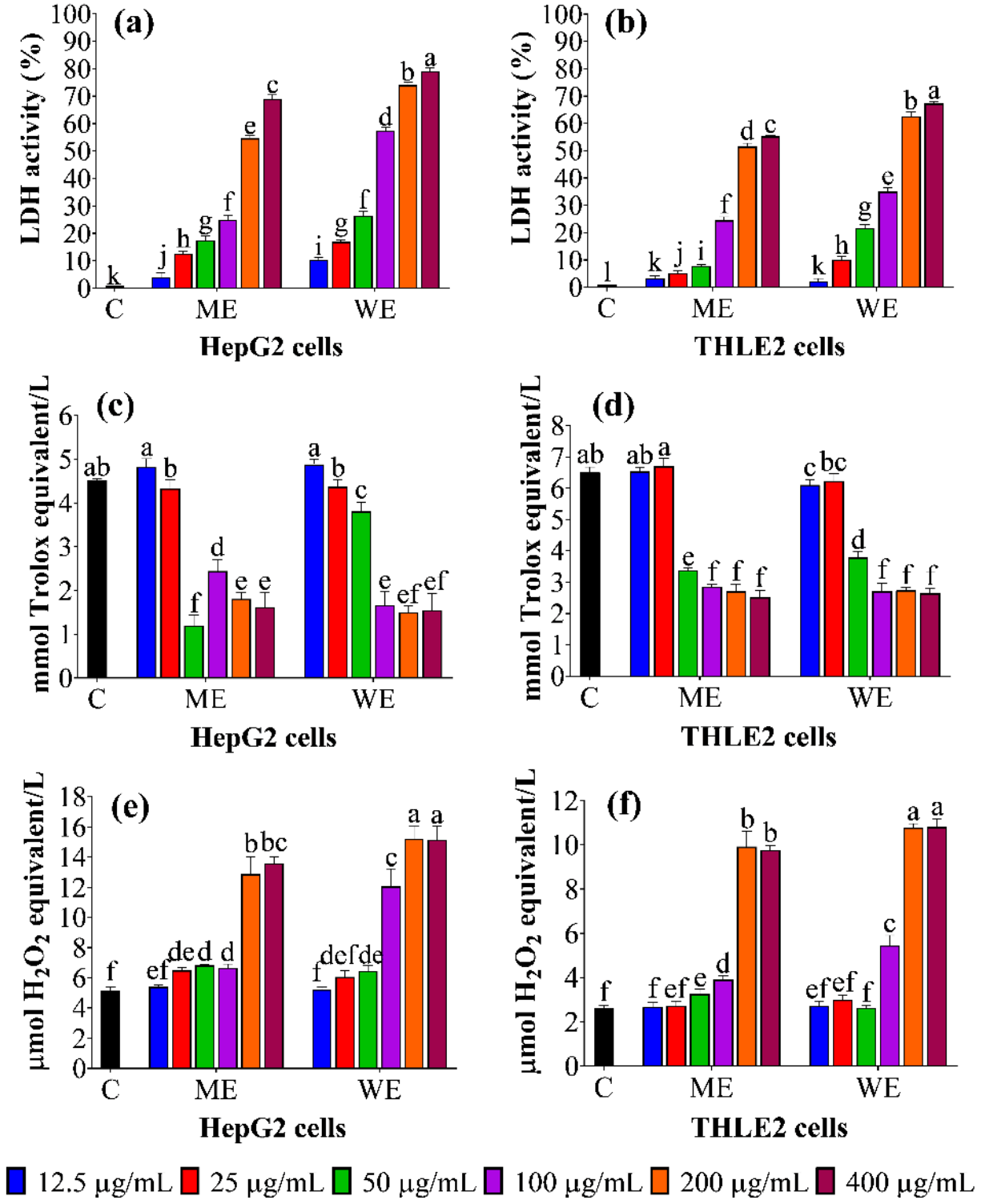

Figure 2

LDH (a, b), total antioxidant status (c, d), and total oxidant status (e, f) activities in cells exposed to different extracts. Data are presented as mean ± standard deviation (*n*= 3). Groups marked with different

letters are significantly different ($p < 0.05$). LDH: Lactate dehydrogenase; C: Control; ME: Methanol extract of *C. demersum*; WE: Water extract of *C. demersum*.

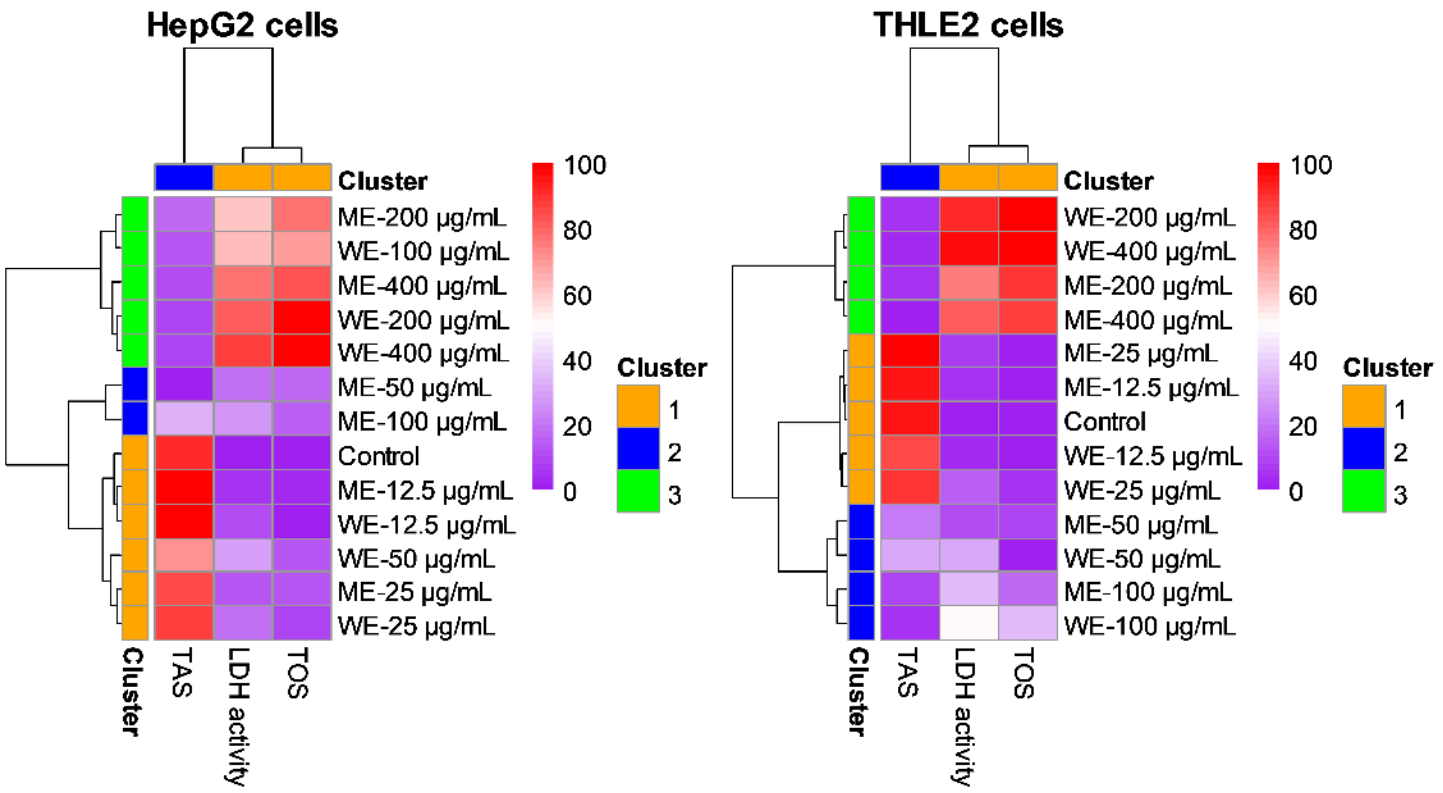

Figure 3

Heatmap and hierarchical clustering analyses of normalized variable data in the cells treated with different concentrations of plant-derived extracts. Red represents high activity levels, whereas purple corresponds to low activity levels. LDH: Lactate dehydrogenase; TAS: Total antioxidant status; TOS: Total oxidant status; C: Control; ME: Methanol extract of *C. demersum*; WE: Water extract of *C. demersum*.

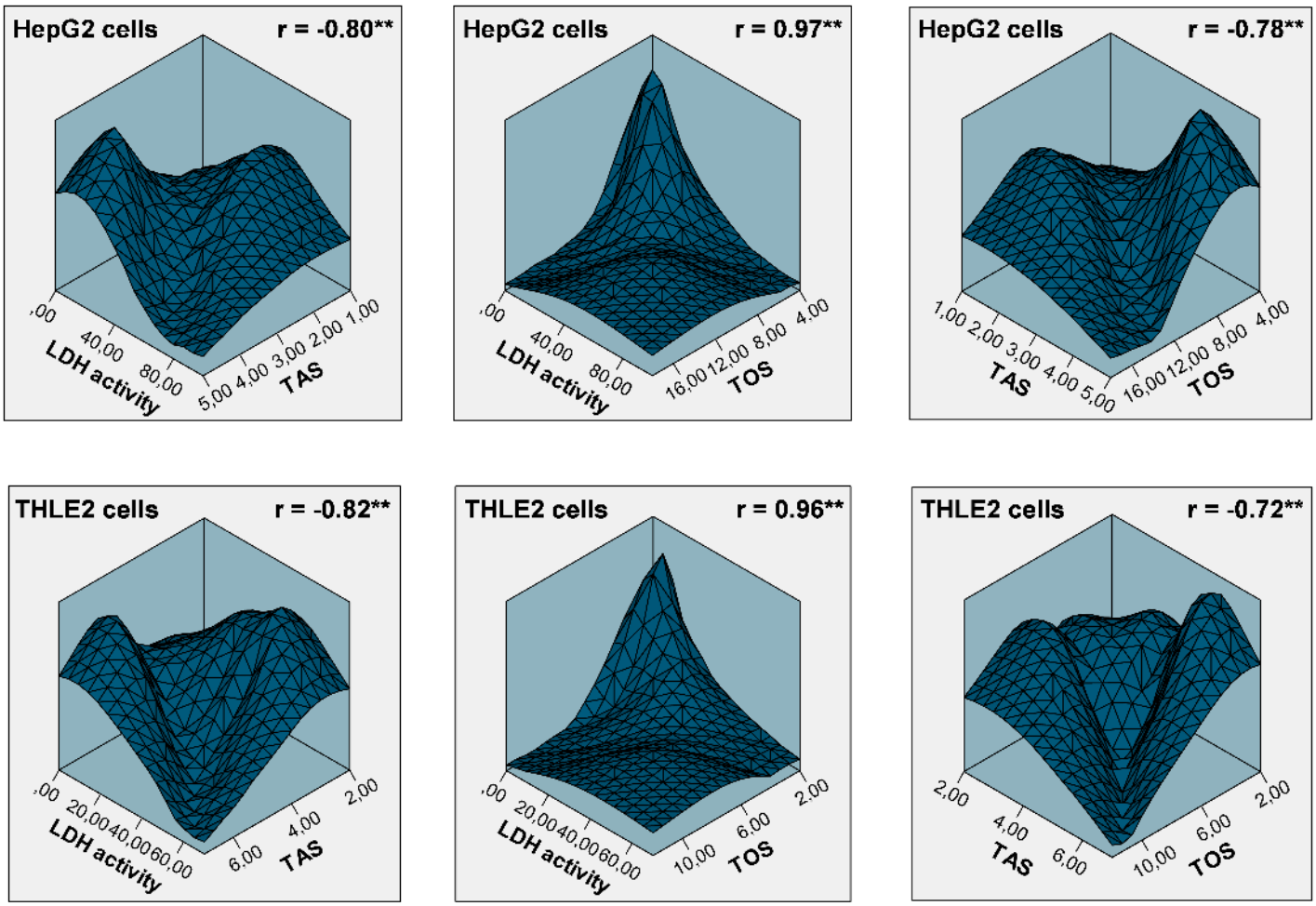

Figure 4

3-D density analysis was carried out to examine the impact of the extracts on cellular LDH, TAS, and TOS activities. Correlation coefficients (r) were determined, with statistical significance set at $p < 0.01$, as indicated by double asterisks (**). LDH: Lactate dehydrogenase; TAS: Total antioxidant status; TOS: Total oxidant status.

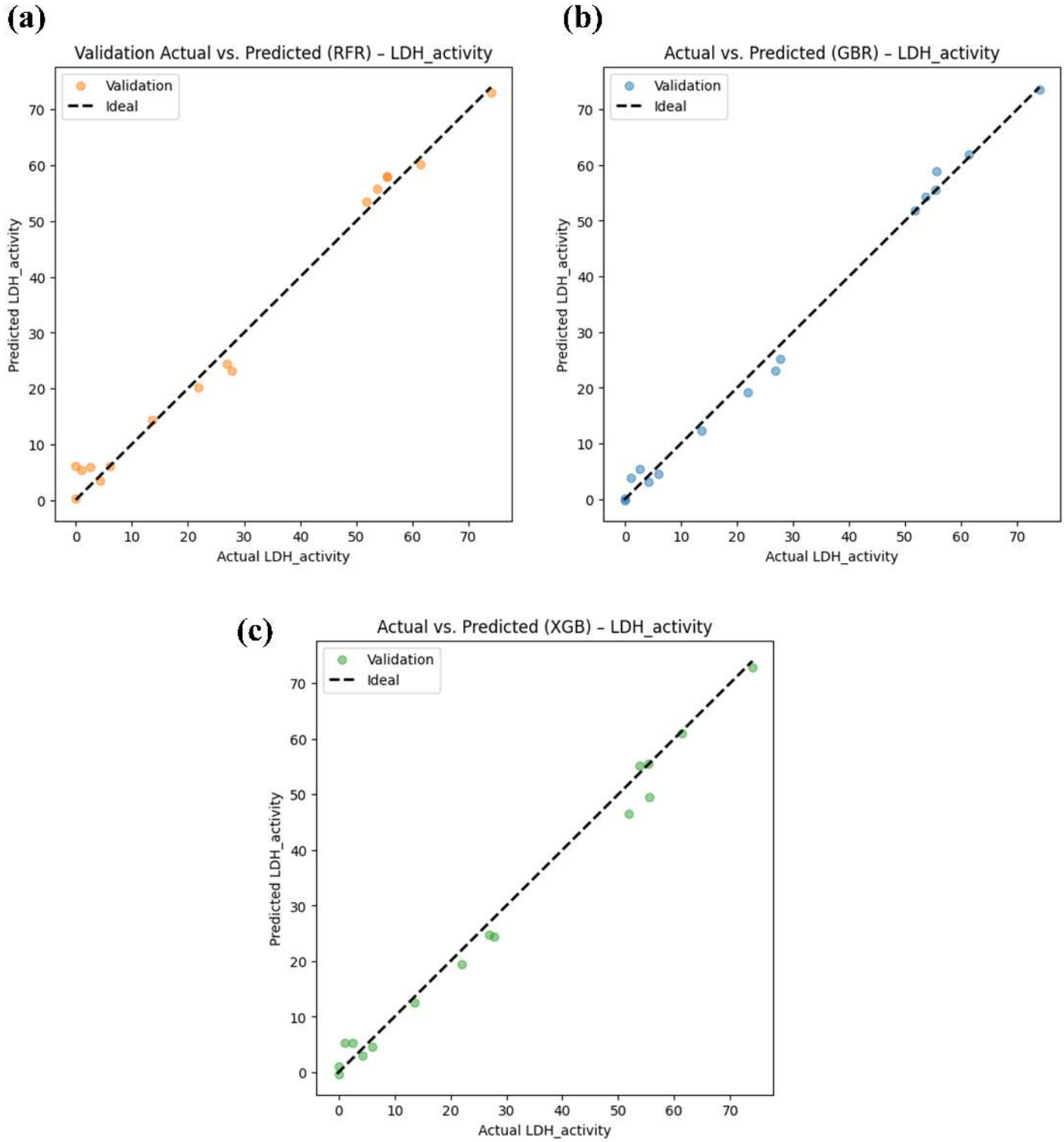

Figure 5

Parity plots of predicted versus observed LDH activity for Random-Forest Regressor (a), Gradient Boosting Regressor (b), and XGBoost (c). The dashed 1:1 line denotes perfect agreement; tighter clustering around this line indicates greater predictive accuracy.

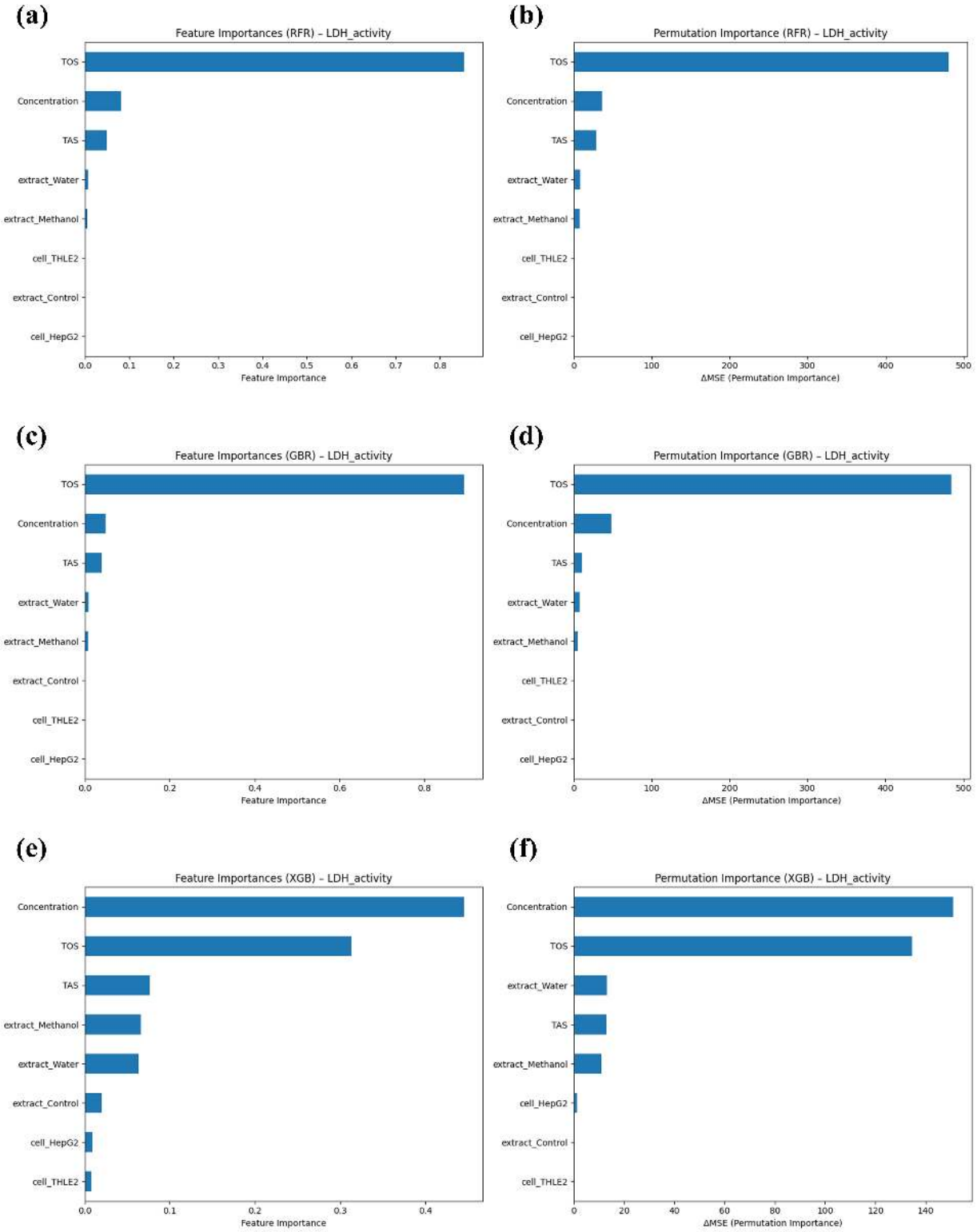

Figure 6

Global rankings obtained from tree-based feature-importance (split gain) and permutation tests for RFR (a, b), GBR (c, d), and XGBoost (e, f).

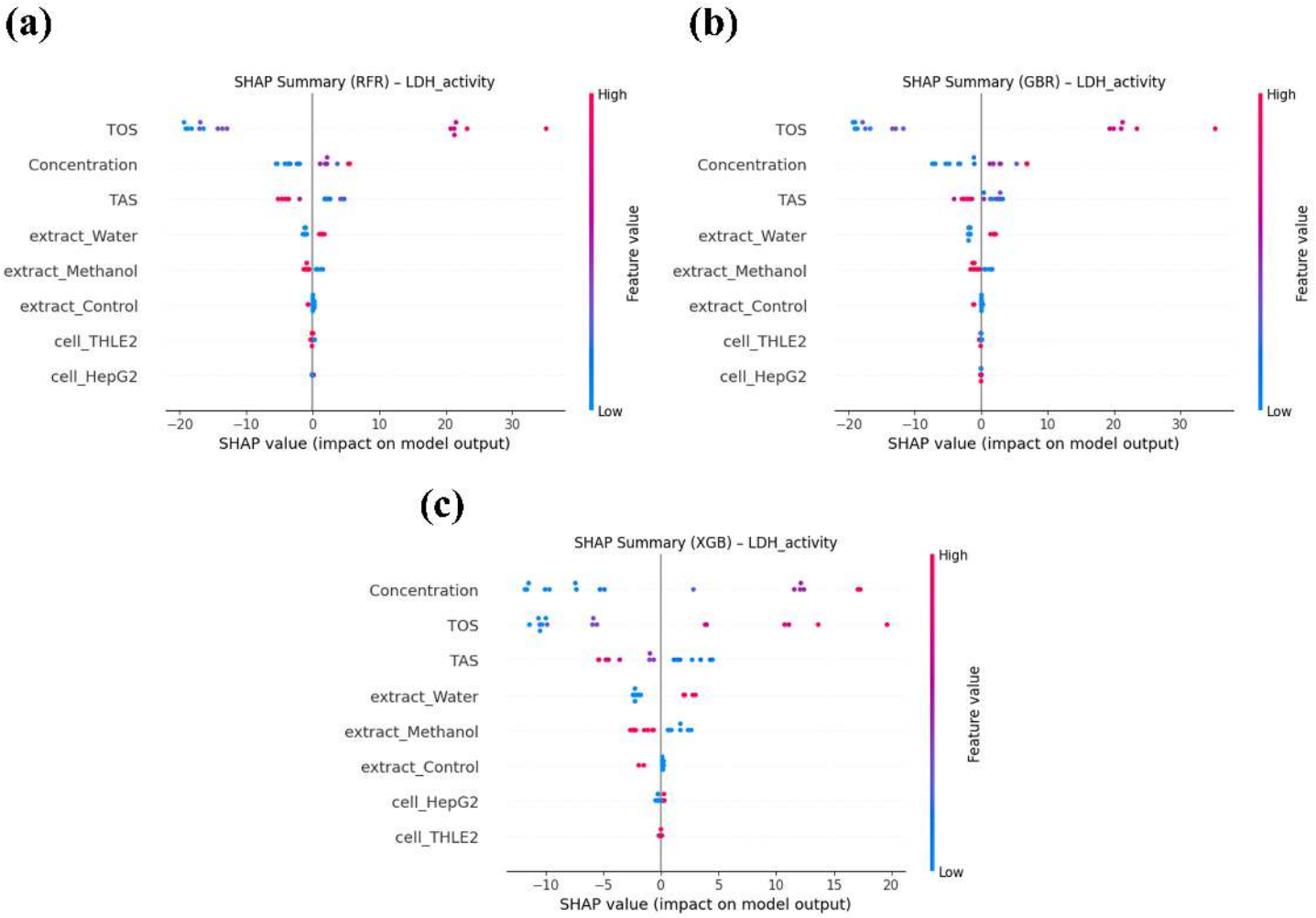

Figure 7

SHAP summary plots for the RFR (a), GBR (b), and XGB (c) illustrating both magnitude and direction of each feature's contribution at the observation level.