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.

Three-Dimensional Finite Element Modeling and Machine-Learning Surrogate Prediction of Stevia Assisted Osmotic Dehydration of Mango Slices

Ripa Muhury

22pmath006@student.cuet.ac.bd

Chittagong University of Engineering & Technology

Usama Ibn Aziz

University of Chittagong

Md. Sakib Ahmmed

University of Chittagong

Ferdusee Akter

Chattogram Veterinary and Animal Sciences University

Ujjwal Kumar Deb

Chittagong University of Engineering & Technology

Article

Keywords: Osmotic dehydration, Stevia, Moisture content, Mass transfer, Machine learning

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11490872/v1

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

Additional Declarations: No competing interests reported.

Three-Dimensional Finite Element Modeling and Machine-Learning Surrogate Prediction of Stevia Assisted Osmotic Dehydration of Mango Slices

Ripa Muhury[a*], Usama Ibn Aziz[b], Md.Sakib Ahmmed[b], Ferdusee Akter[c], Ujjwal Kumar Deb[a]

[a] Department of Mathematics, Chittagong University of Engineering & Technology, Chattogram - 4349, Bangladesh

[b] Department of Mathematics, University of Chittagong, Chattogram-4331, Bangladesh

[c] Department of Physical and Mathematical Sciences, Chattogram Veterinary and Animal Sciences University, Chattogram-4225, Bangladesh

[*]Corresponding author: 22pmath006@student.cuet.ac.bd (Ripa Muhury)

Abstract

Osmotic dehydration (OD) is a food preservation technique that entails the transfer of water from the fruit tissue to the osmotic medium, with the migration of soluble substances from the osmotic medium into the fruit. This research used a 3D finite-element (FE) model to investigate the simultaneous transport of water and solute from stevia, aiming to replicate water loss (WL) and solute gain (SG), and included machine-learning (ML) surrogate models for fast predictions of WL and SG. Transient mass transfer was computed in the COMSOL Multiphysics framework for 18 hours over solution concentrations of (5,7 and 10) °Brix, and temperatures of (304,314 and 323) K. The model successfully simulated the gradual development of concentration gradients from the mango-solution boundary to the center, with a decline in water concentration as solute concentration increased over time. The average water concentration decreased from about 49,700 to 47,300 mol m^{-3}, while the average concentration of solutes rose from about 188 to 250 mol m^{-3}. The extraction of water increased with higher solution concentration and temperature, maximum at 304 K under the examined model circumstances with total solute accumulation. In the grouped cross-validation, CatBoost's RMSE values were 0.58411 on WL and 0.10318 on SG; in the independent test, they were 0.4772 on WL and 0.1028 on SG. The integrated FE - ML approach enables spatially interpretable transport analysis and efficient response prediction to help design and screen processes throughout the explored operating region.

Keywords: Osmotic dehydration; Stevia; Moisture content; Mass transfer; Machine learning.

Nomenclature

Symbol	Definition	Unit
C_i	Molar concentration of species *i*	mol m^{-3}
$C_{i,0}$	Initial molar concentration of species *i*	mol m^{-3}
$C_{i,s}$	Molar concentration of species *i* at the mango surface	mol m^{-3}
$C_{i,b}$	Bulk concentration of species *i* in the osmotic solution	mol m^{-3}
D_i	Intrinsic diffusion coefficient of species *i*	m^2 s^{-1}
$D_{i,\mathrm{eff}}$	Effective diffusion coefficient of species *i* in mango tissue	m^2 s^{-1}
$D_{i,\mathrm{ref}}$	Reference diffusion coefficient	m^2 s^{-1}
$E_{a,i}$	Apparent activation energy for diffusion of species *i*	J mol^{-1}
J_i	Diffusive molar flux of species *i*	mol m^{-2} s^{-1}
k_i	Interfacial mass-transfer coefficient of species *i*	m s^{-1}
M_w	Molar mass of water	kg mol^{-1}
MR	Moisture ratio	–
n	Outward unit normal vector at the mango surface	–
R	Universal gas constant	J mol^{-1} K^{-1}
R^2	Coefficient of determination	–
$RMSE$	Root mean square error	–
MAE	Mean absolute error	–
SG	Solute gain	–
T	Absolute temperature	K
t	Osmotic dehydration time	s
WL	Water loss	–
ε	Effective porosity of mango tissue	–
ρ_w	Density of water	kg m^{-3}
B	Osmotic-solution concentration	°Brix
i	Species index (water or stevia solute)	–
b	Bulk osmotic solution	–
s	Surface condition or stevia solute, according to context	–
w	Water	–
OD	Osmotic dehydration	–
FEM	Finite element method	–
ML	Machine learning	–
MSI	Mango–solution interface	–
SSR	Subsurface region	–
IR	Intermediate region	–
CR	Core region	–
RF	Random forest	–

1. Introduction

Osmotic dehydration (OD) is one of the most significant food preservation and pre-treatment methods in which water is partially extracted from the plant tissues by immersing them in a hypertonic solution [1] [2] [3]. It is a simultaneous counter-current mass transfer process with the transfer of water from the food matrix to the osmotic solution, and the transfer of soluble components from the osmotic solution to the food matrix [4] [5] [6]. Due to the removal of water without complete vaporisation, OD can lower the thermal and energy demands of the subsequent drying stage and can be used to help retain desirable product properties [7] [8] . The performance depends on the food structure, properties of osmotic agents, its concentration, temperature and processing time [1] [9] [10].

The choice of mango (Mangifera indica L.) as a substrate for OD is due to its high moisture content and structure. But the simple independent-diffusion assumption is not sufficient to represent the mass transfer in mango. Water loss (WL) and solute gain (SG) are both happening at the same time, concentration gradients are continuously changing, and tissue structure can change the internal transport resistance[1] . The transfer of moisture from a food product during drying process is a multi-component phenomenon consisting of coupled heat and mass transfer, internal moisture resistance and structural changes, which can be analyzed and predicted by the mathematical modeling of the phenomenon[11] [5] [7]. Floury et al. created a 3D COMSOL model of mango, with intracellular and extracellular sucrose and water transport formulated separately and then coupled together across the cell membranes; in the model, they found good quantitative agreement with experimental WL and sugar-gain data[12]. Likewise, using finite-element modeling COMSOL has been shown to predict simultaneous water and solute transport in melon, and to estimate the diffusion and interfacial mass-transfer parameters in the multicomponent diffusion case [13].

On the other hand, OD is a fundamentally connected process. The substantial overestimation of diffusivities is caused by partly linked or uncoupled formulations, according to Aguirre-García et al., which is the most major impact of water and solute diffusion when coupled to product shrinkage [4] [14]. It has been noted from mango experiments that WL and SG could show distinct kinetic behaviors; for example, in the circumstances utilized by Sulistyawati et al., the Peleg model was more suited to SG than the Weibull model. These findings indicate that lumped empirical models may well characterize overall kinetics, although they provide little insight into the spatial progression of mass transfer inside fruit tissue [15]. It is also important to choose an appropriate osmotic agent for osmotic OD. Sucrose is the most thoroughly examined osmotic agent; yet, natural sweeteners are favored [14] . Stevia and other natural sweeteners have been identified as possible osmotic agents in evaluations due to interest in reducing the usage of conventional sugar [16] [10]. Crucially, the osmotic agent may influence mass-transfer properties; studies

comparing sucrose and sorbitol have shown that the osmotic system affects both the rate of water loss and mass transfer features [6] [14] .

An further opportunity is presented by machine learning (ML). To alleviate some of the numerical strain on the high-fidelity simulation model, ML models may describe nonlinear interactions between processing variables and process responses[17]. With the use of several input parameters like temperature, sample thickness, initial moisture content, processing time, and pre-treatment, ANN models have recently been demonstrated to be useful for predicting food-drying processes, including OD[18]. Typically, ML-oriented surrogate modeling has been advocated to serve as an alternative to physics-based simulations without diminishing their predictive capabilities. Recent deep-learning research has also shown that surrogate models may mimic numerical process responses with much lower computing cost[17] . To accomplish this, three components must be incorporated: (i) spatially resolved mechanistic modeling of the interconnected water-solute transport; (ii) employing the spatially resolved mechanistic model for stevia-assisted mango OD; and (iii) transforming the high-fidelity simulation data into a machine learning-based rapid prediction framework.

While most of the mango COMSOL studies conducted have been based on conventional osmotic agents such as sucrose, studies have been focused on experimental kinetics, and semi-empirical and empirical models have been reported in a few studies with alternative sweeteners. The present study thus proposes a coupled FE-ML framework especially designed for the stevia-based OD of mango. In this study, a three-dimensional finite-element model was developed in COMSOL Multiphysics to simulate transient water and stevia transport within mango tissue. The model analyzes the effects of stevia concentration, processing time, and temperature on WL and SG, producing concentration fields from a spatial perspective. A systematic simulation was conducted using three various concentrations of Stevia solution (5 °Brix, 7 °Brix, and 10 °Brix) at three varying temperatures (304 K, 314 K, and 323 K), with a structured dataset used to construct the machine learning algorithm. Subsequently, machine learning models were created using the simulation results for rapid forecasting of mass-transfer reactions, and a physics-to-data framework was instituted for process evaluation.

The novelty of the study lies in integrating transport physics principles, COMSOL simulations with spatial precision, systematic parameter change, and machine learning-based surrogate predictions into a cohesive computational process. The numerical simulations do not replace the physical model by ML, but they are used to provide physically interpretable training data for the ML model which is a computationally efficient way to predict process responses when untested conditions occur. The simulations indicated that the absorption of stevia escalated with extended processing duration, and that the stevia uptake approached an

asymptotic value as function of time, with the magnitude and rate of uptake dependent on the osmotic concentration and temperature.

Along with the traditional combined kinetic models, the acquired concentration fields revealed the time-dependent development of internal concentration gradients. Next, we utilize the simulation database to make ML predictions of WL and SG and to find out how the processing factors affected each other. This study may be extended to other systems involving osmotic agents in which experimental data is sparse, and it proposes a framework to link mass-transfer processes with rapid prediction models. The framework will function not only to elucidate the fundamental mass-transfer dynamics but will also act as a computationally efficient resource for process evaluation, parameter examination, WL and SG forecasting, and future process enhancement without a comprehensive experimental dataset. Within this context, the goals of the ongoing research were to: (i) devise a three-dimensional finite element model for the interrelated transport of water and stevia during mango OD; (ii) evaluate the impact of stevia concentration, temperature, and time on water loss and solid gain; (iii) describe the spatial and temporal concentration distributions within the mango matrix; (iv) develop machine learning models from the simulated data for swift prediction of OD kinetics; and (v) establish an integrated mechanistic-data-driven strategy for predictive analysis of stevia-assisted mango dehydration.

2. Mathematical Formulation and Numerical Methodology

2.1. Model Configuration

A three-dimensional finite-element model for mango was used to represent the transient mass transfer during Stevia assisted OD. Four steps comprised the modeling strategy: (i) the construction of the three-dimensional mango osmotic-solution domain; (ii) the formulation of coupled water and solute transport; (iii) a transient finite-element solution for solute gain (SG) and water loss (WL) in the domain; and (iv) post-processing of mass-transfer responses, moisture content, moisture ratio, water loss (WL), and solute gain (SG).

In OD treatment, two simultaneous and opposing transport processes occur: water is moving from the mango tissue towards the osmotic medium, while soluble matter is moving from the osmotic solution towards the mango tissue. Mechanistic and 3D numerical models of OD have been developed including mango specific FEM studies[4][19] that have included coupled water and solute transport.

The computational geometry was divided into two domains: (i) the mango domain, where the transient transport of the solution in the mango's porous tissue, and (ii) the domain of finite osmotic-solution, which

was next to the mango domain and allowed the solution's concentration to fluctuate while the mango dried. The mango solution interface served as the primary mass transfer boundary. The areas shown in Fig. 1 MSI(Mango solution interface), SSR(Subsurface region), IR(Intermediate region), and CR(Core region) were analyzed to look at the spatial shift in water and solute concentration from the mango's surface to its core.

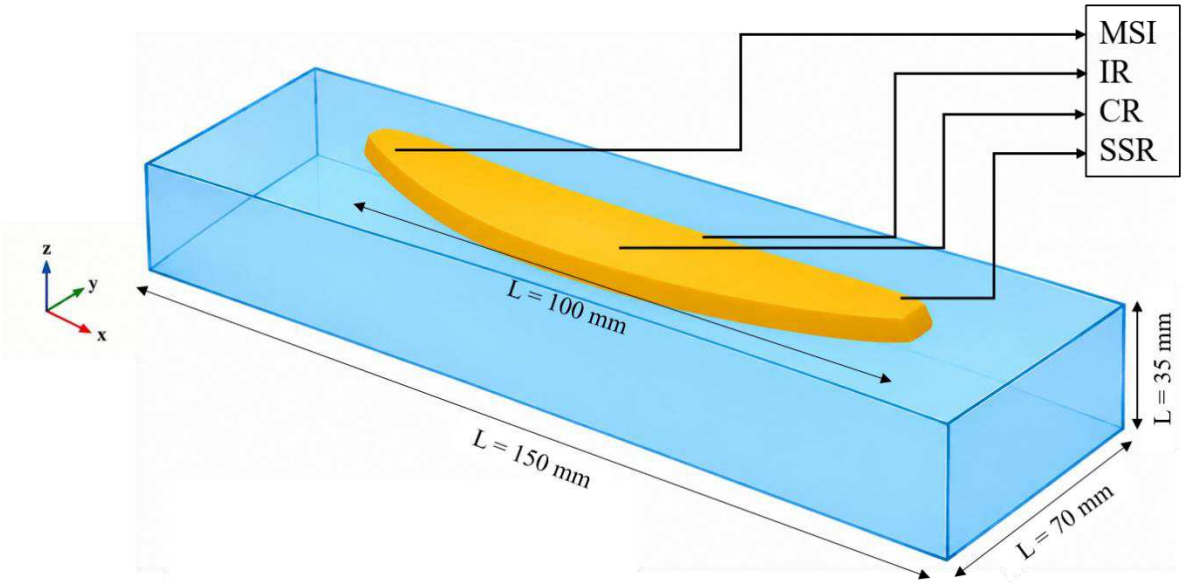


Fig. 1. 3D computational domain of the curved mango slice immersed in osmotic-solution bath.

2.2. Governing Equations

2.2.1. Transient Mass Transport

The transport of water and stevia solute was formulated using the species conservation equation for a non-reactive species *i*:

$$\frac{\partial C_i}{\partial t} + \nabla \cdot \mathbf{J}_i = 0, (i = w, s) \quad (1)$$

Where C_i is the molar concentration and $\mathbf{J}_i$ is the diffusive flux. The subscripts w and s denote water and stevia solute, respectively.

The diffusive flux was described by Fick's first law:

$$\mathbf{J}_i = -D_i \nabla C_i \quad (2)$$

The negative sign indicates transport down the concentration gradient. Fickian diffusion and separate water/solute transport formulations are widely used in mechanistic OD models, including three-dimensional mango modelling[12] [4] .

The governing equations for the mango domain reduce to

$$\frac{\partial C_w}{\partial t} = \nabla \cdot \left(D_{w,\text{eff}} \nabla C_w\right) \tag{3}$$

And

$$\frac{\partial C_s}{\partial t} = \nabla \cdot \left(D_{s,\text{eff}} \nabla C_s\right) \tag{4}$$

Here, $D_{w,\text{eff}}$ and $D_{s,\text{eff}}$ are the effective diffusivities of water and stevia solute within the mango tissue. This formulation represents the diffusion-controlled transport actually implemented in the numerical model.

2.2.2. Effective Diffusivity of Mango Tissue

The porous structure of mango tissue was represented through an effective diffusivity.

$$D_{i,\text{eff}} = D_{i,m} \varepsilon^{1.5} \tag{5}$$

where $D_{i,m}$ is the intrinsic diffusivity and ε is the effective porosity.

The $\varepsilon^{1.5}$ is the Bruggeman approximation for effective diffusion in porous media[20]. In the present study, Eq. (5) was adopted as a porous-medium closure relation rather than assumed to be a universal mango-specific relationship.

2.2.3. Temperature Dependence of Diffusivity

The temperature dependence of the intrinsic diffusivities was represented using an Arrhenius relationship:

$$D_{i,m} = D_{i,\text{ref}} \exp\left(-\frac{E_{a,i}}{R_g T}\right) \tag{6}$$

where $D_{i,\text{ref}}$ is the reference diffusion coefficient, $E_{a,i}$ is the apparent activation energy, R_g is the universal gas constant, and T is absolute temperature.

Arrhenius-type temperature dependence has been applied in mango OD modelling to represent changes in effective mass-transfer parameters with temperature. Arias et al. also fitted an Arrhenius relationship for the effective diffusion coefficient of mango during OD[21] [22] .

2.3. Mango Solution Interfacial Mass Transfer

Mass exchange between the mango and osmotic solution was represented through an interfacial flux condition:

$$-\mathbf{n} \cdot \mathbf{J}_i = k_i\left(C_{i,m} - C_{i,b}\right) \tag{7}$$

where $\mathbf{n}$ is the outward normal at the mango surface, $k_i$is the interfacial mass-transfer coefficient, $C_{i,m}$ is the mango-surface concentration, and $C_{i,b}$is the corresponding bulk-solution concentration.

For water,

$$-\mathbf{n} \cdot \mathbf{J}_w = k_w\left(C_{w,m} - C_{w,b}\right) \tag{8}$$

where the concentration difference drives water transfer from the mango toward the osmotic medium.

For stevia solute,

$$-\mathbf{n} \cdot \mathbf{J}_s = k_s\left(C_{s,m} - C_{s,b}\right) \tag{9}$$

Thus, when $C_{s,b} > C_{s,m}$, the resulting flux is directed into the mango. This treatment represents the simultaneous opposing water and solute fluxes characteristic of OD. Mechanistic OD models have similarly [23] incorporated diffusion within the product together with interfacial mass-transfer resistance[24] .

2.4. Initial Conditions

At the beginning of the OD process $t = 0$, the mango and surrounding osmotic solution were assigned their respective initial thermal and concentration states. The initial conditions were defined as,

$$T(\mathbf{x}, 0) = T_0,\ C_w(\mathbf{x}, 0) = C_{w_0},\ C_s(\mathbf{x}, 0) = C_{s_0}, \tag{10}$$

where T_0 is the initial temperature, while C_{w_0} and C_{s_0}denote the initial water and stevia-solute concentrations, respectively for the mango domain, the initial water concentration in mango was calculated from the initial water activity:

$$C_{w_m0} = a_w C_{w,\mathrm{ref}} \tag{11}$$

Where, $C_{w,\mathrm{ref}} = \frac{\rho_w}{M_w}$ (12)

Here, ρ_w is water density and M_w is the molar mass of water.

2.5. Calculation of Water Loss and Solute Gain

The water loss and solute gain determined using a mass balanced method[25].

$$WL(\%) = \frac{m_{w,0} - m_w(t)}{m_0} \times 100 \tag{13}$$

and

$$SG(\%) = \frac{m_s(t) - m_{s,0}}{m_0} \times 100 \tag{14}$$

The net weight reduction was calculated as

$$WR(\%) = WL(\%) - SG(\%) \tag{15}$$

assuming that water loss and solute gain are the dominant mass-transfer contributions.

2.6. Moisture Ratio

The moisture ratio(MR) of the sample was calculated according to following equation;

$$MR(t) = \frac{X(t) - X_e}{X_0 - X_e} \tag{16}$$

where X_0, $X(t)$and $X_e$represent the initial, instantaneous, and equilibrium moisture contents, respectively.

2.7. Numerical Implementation

An unstructured tetrahedral mesh was used to mesh the computational region, increasing the spatial resolution close to the mango solution interface. A time-dependent investigation was used to acquire the transient concentration distributions. The baseline simulation was performed at $T = 323.15$ K (50°C), with $B = 5°\text{Brix}$ and $t_{\text{end}} = 64800\ s\ = 18$ h

The resulting concentration fields were post-processed to obtain WL, SG, WR, moisture content and MR.

3. Computational Procedure

3.1. Computational Domain and Discretization

The computational domain consisted of the curved mango slice and the surrounding finite osmotic-solution region. The three-dimensional geometry was discretized using an unstructured tetrahedral finite-element mesh. Local mesh refinement was applied near the mango–solution interface to adequately resolve the concentration gradients associated with simultaneous water removal from and solute uptake by the mango tissue.

The transport domain included two coupled concentration fields, namely water and stevia solute. The transient concentration equations were solved throughout the mango and surrounding solution domains, while the finite-bath formulation allowed the solution concentration to evolve during OD rather than remaining fixed at its initial value.

The computational workflow consisted of geometry construction, material and transport-property assignment, definition of initial and boundary conditions, mesh generation, transient FEM solution, post-processing of water-loss (WL) and solute-gain (SG) responses, and subsequent generation of the dataset used for machine-learning (ML) surrogate modelling. The overall workflow is illustrated in Fig. 2(a).

3.2. Numerical Stability and Convergence

The evaluation of numerical stability included observing the transient reactions of water loss and solute growth over the simulation duration. The assessed concentration fields were analyzed to confirm their physically coherent behavior during osmotic dehydration, marked by a gradual reduction of water in the mango tissue and an increase in stevia solute. The numerical solution was regarded as adequately convergent when additional enhancement of the spatial discretization led to very minimal modifications in the key mass-transfer outputs. Moreover, the temporal concentration profiles were assessed for insignificant oscillations or drastic shifts that may imply numerical instability. A methodical grid-independence

assessment was therefore conducted using increasingly improved tetrahedral meshes. The ultimate WL and SG responses derived from each mesh were analyzed to assess the impact of spatial discretization on the computed solution.

3.3. Meshing and Grid-Independence Test

A grid-independence test (GIT) was performed with increasing resolution of the unstructured tetrahedral meshes up to six in order to investigate this, which is shown in Fig. 2(c). The number of elements was increased from around 3,700 to 58,000 while maintaining the same computational geometry, material characteristics, boundary conditions, and simulation parameters. The primary metrics used to assess mesh sensitivity were the final WL and SG.

The tetrahedral mesh was sufficient in capturing the curved geometry of the mango and the surrounding domain of osmotic solution, having higher resolution focusing on the mango solution interface as shown in Fig. 2(b).

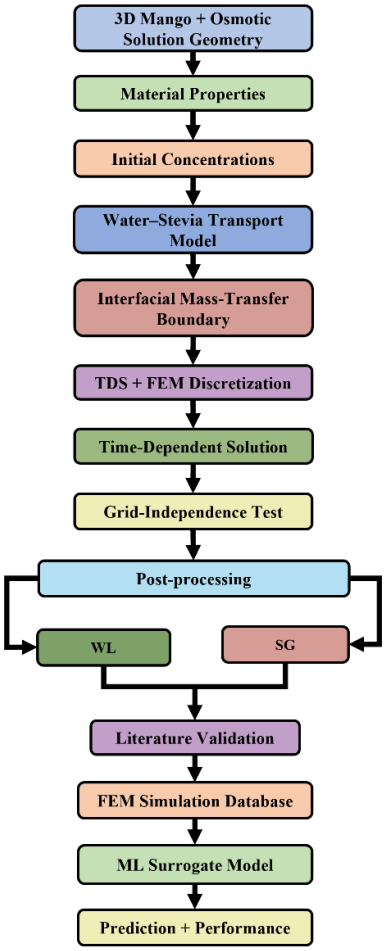
The relative change between two consecutive mesh levels was calculated as

$$\Delta Y_i(\%) = \frac{|Y_i - Y_{i-1}|}{Y_{i-1}} \times 100 \quad (17)$$

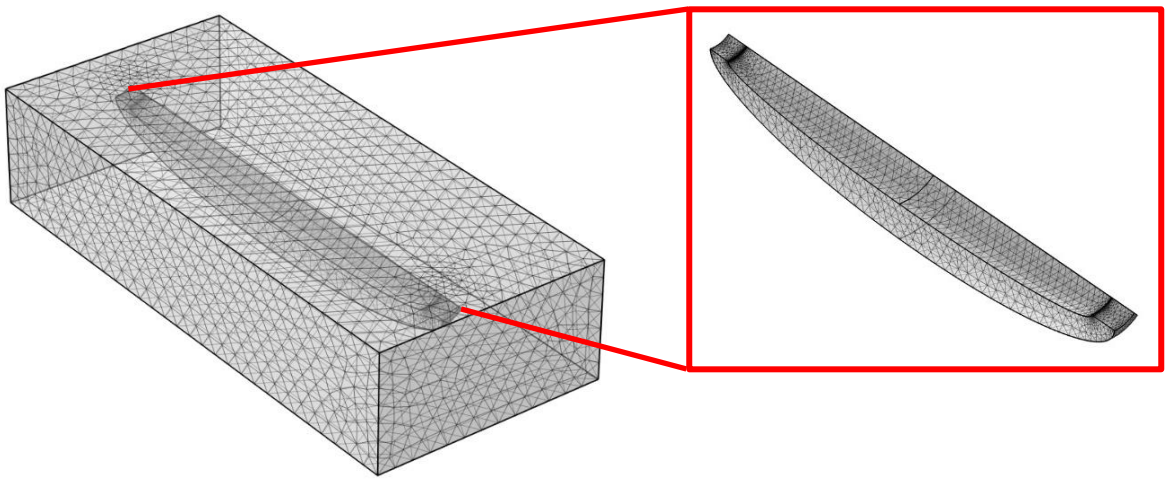
where $Y_i$represents the final WL or SG obtained using mesh i.

The results are summarized in **Table 1**. The transition from M4 to M5 produced changes of 1.08% in WL and 1.21% in SG. Further refinement from M5 to M6 reduced these changes to only 0.36% and 0.40%, respectively. Thus, the difference between the 38,000 and 58,000 element meshes was small relative to the changes observed between the finer meshes.

(a)


3D Mango + Osmotic Solution Geometry
Material Properties
Initial Concentrations
Water–Stevia Transport Model
Interfacial Mass-Transfer Boundary
TDS + FEM Discretization
Time-Dependent Solution
Grid-Independence Test
Post-processing
WL
SG
Literature Validation
FEM Simulation Database
ML Surrogate Model
Prediction + Performance

(b)



(c)

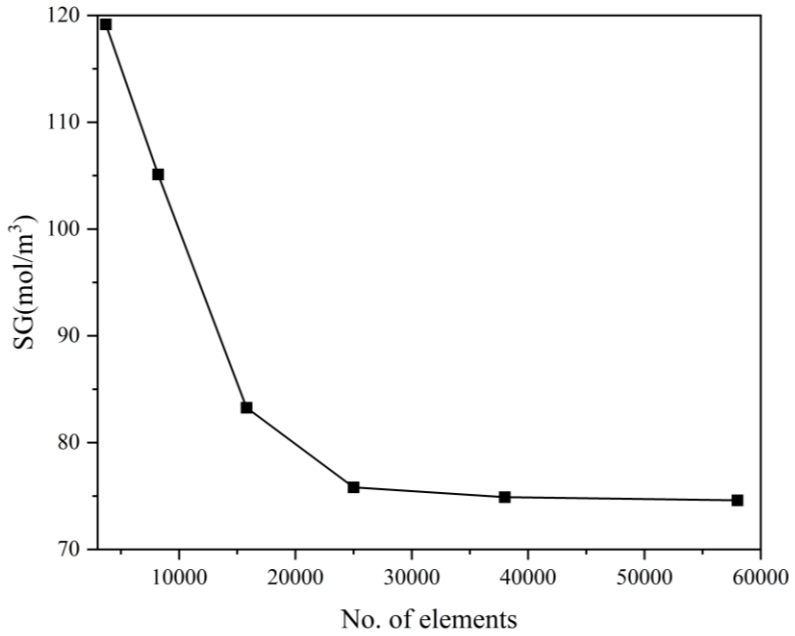


Fig. 2. Computational procedure and numerical verification: **(a)** computational workflow from FEM simulation, **(b)** three-dimensional tetrahedral mesh and **(c)** grid-independence test based on the predicted SG.

Table 1. Grid-independence analysis.

Mesh level	**No. of elements**	**Final WL (mol m^{-3})**	**Relative change WL (%)**	**Final SG (mol m^{-3})**	**Relative change SG (%)**
M1	3,700	22,786.4	—	119.17	—
M2	8,200	21,861.9	4.06	105.12	11.79
M3	15,800	19,946.2	8.76	83.27	20.78
M4	25,000	19,461.0	2.43	75.82	8.95
M5	38,000	19,250.0	1.08	74.90	1.21
M6	58,000	19,180.0	0.36	74.60	0.40

3.4. Machine-Learning Surrogate Modelling

A surrogate model was developed using a machine-learning technique to enhance computing efficiency in depicting the FEM simulations. The data were obtained directly from the COMSOL simulations and used to develop the surrogate model, which therefore generated an estimation of the true numerical relationship among the processing conditions, processing duration, and mass transfer responses.

Three process variables were used as model inputs: temperature (T), osmotic-solution concentration expressed as °Brix (B), and processing time (t). The input vector was therefore defined as

$$\boldsymbol{X} = [T, B, t] \tag{18}$$

Separate regression models were developed for water loss and solute gain:

$$\hat{y}_{WL} = y_{WL}(X) \tag{19}$$

$$\hat{y}_{SG} = y_{SG}(X) \tag{20}$$

where $\hat{y}_{WL}$and $\hat{y}_{SG}$ represent the predicted WL and SG, respectively.

Four regression algorithms CatBoost [26] , XGBoost [27] , Random Forest (RF) [28] , and Ridge regression [29] were selected to represent a diverse spectrum of machine learning paradigms. CatBoost, XGBoost, and RF were chosen as robust, non-linear ensemble methods capable of capturing the complex, multi-variable non-linear interactions inherent in transient mass-transfer processes [17][18] while effectively preventing overfitting [28]. In contrast, Ridge regression was incorporated as a regularized linear baseline to assess whether linear relationships are sufficient or if non-linear tree-based architectures are strictly required for accurate surrogate prediction [17] [29].The dataset was temporally segmented by Run_ID to ensure that observations from each distinct COMSOL simulation trajectory were contained inside the same data subset, hence reducing the risk of temporal data leakage. An autonomous test set was preserved, constituting about 20% of the overall simulation executions, and the other 80% were allocated for the creation of the simulation model. Subsequently, a 5-fold GroupKFold via Scitkit-learn [30] was used, with Run_ID as the grouping variable. To achieve this, it was confirmed that no temporally linked data from the same simulation trajectory were used concurrently for model training and validation. Key hyperparameters for each candidate model (such as maximum tree depth, learning rate, and number of estimators) were carefully tuned using grid search embedded inside each cross-validation fold to avoid overfitting and guaranty excellent predictive capacity.

Model performance was quantified using the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). These metrics were calculated as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (21)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i | \quad (22)$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (23)$$

where $y_i$is the FEM-generated value, $\hat{y}_i$is the corresponding ML prediction, $\bar{y}$ is the mean observed value, and N is the number of observations.

3.5. Model Validation

The outcomes of the current simulation were contrasted with the published findings on mango osmotic-dehydration by [21] and [15] . Arias et al. studied the mass-transfer kinetics of Tommy Atkins mango during OD at various sucrose concentrations (0-60%) and different processing temperatures (50 °C). Sulistyawati et al. investigated the loss of water, the increase of soluble solids, and the decrease of water during the OD of mangoes under regulated processing parameters [15] . The absolute solute concentration or mass gain could not be compared with previous investigations, as they used sucrose-based osmotic solutions, while the current model utilized a 5 °Brix stevia-based solution. In order to determine the link with it, the solute concentration found in mangos was represented in normalized form.

$$X^* = \frac{C(t) - C_0}{C_f - C_0} \quad (24)$$

Where $C(t)$ is the volume-averaged solute concentration in the mango at time t, while C_0 and C_f represent the initial and final solute concentrations, respectively. This normalization makes it possible to compare the temporal uptake behaviour without making any assumptions about the osmotic driving forces for stevia at a particular °Brix and for sucrose.

The present FEM prediction shows progressive uptake of solute with a relatively high uptake rate in the initial stage and a decreasing rate as the process progresses as indicated in Fig. 3. This behavior is consistent

with the general kinetic behavior of OD in mangos. The comparison was utilized to verify the physical feasibility and validity of the simulated kinetic behavior rather than to measure the agreement with the real solute. Multiple variables such as the concentration of osmotic agents, agent levels, mango cultivar, ripeness stage, sample shape, mango-to-solution ratio, and processing parameters were taken into account for the comparative study. It is crucial to understand that the juxtaposition of literature does not serve as empirical evidence of a direct correlation but instead functions as a dynamic metric of literary comparison. To quantitatively verify the current 5°Brix stevia model, experimental data must be collected under identical formulation and processing parameters.

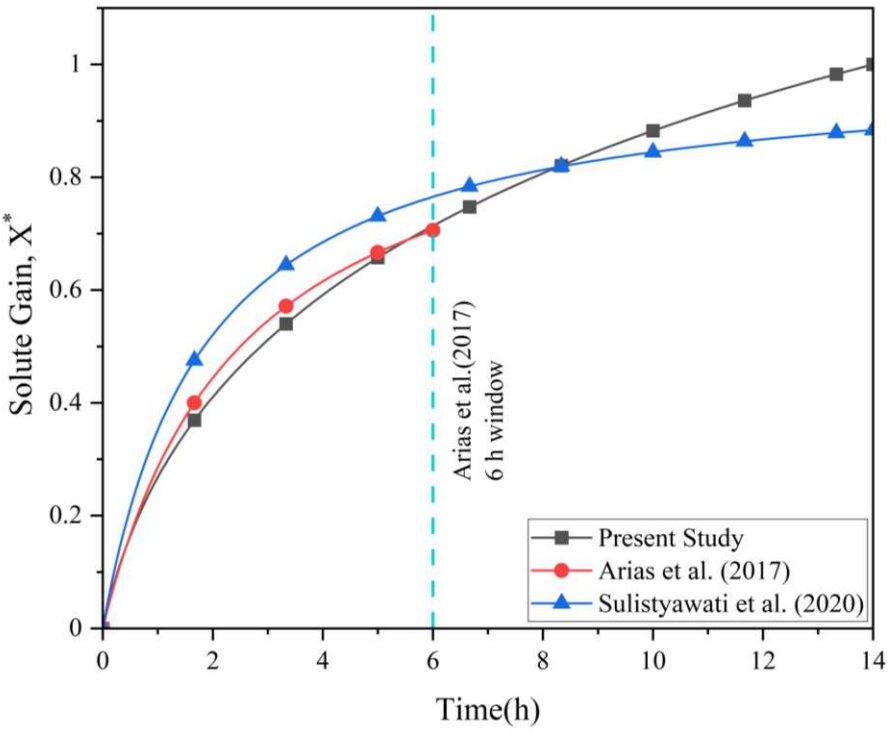


Fig. 3. Validate present work against Arias et al. (2017) and Sulistyawati et al. (2020).

4. Result and Discussion

4.1 Spatial Evolution of SG and WL

The spatial evolution of water and solute concentrations was studied to understand the mass-transfer processes that cause the expected water loss (WL) and solute gain (SG) during OD. In contrast to lumped models, which only describe the temporal evolution of WL and SG, the present finite-element model resolves all concentration fields in mango tissue, revealing the spatial evolution of mass transfer. The 3D

concentration fields were measured at 0, 300, 840, and 1080 min, or 5, 14, and 18 h. These time periods represent the early, moderate, and late dehydration stages.

Fig. 4 illustrates the temporal variations in the expected concentration distribution of solutes inside the mango tissue over three dimensions. The solute concentration profile at the beginning time, is determined by the initial concentration distribution established inside the computational domain. As OD occurs, a concentration gradient develops at the boundary between the mango and the solution, providing the necessary push for solutes to migrate from the surrounding osmotic medium into the mango tissue. The concentration field increasingly redistributes within the mango domain at 300 minutes (5 hours). The formation of regions of moderate density on the mango surface indicates the gradual absorption of the osmotic fluid into the tissue. The concentration is distributed unevenly in space, indicating that the solute will not achieve equilibrium instantaneously inside the mango matrix. Conversely, mass transfer will occur progressively from the outside to the inside. After 840 minutes (14 hours), the solute concentration becomes more dispersed throughout the mango tissue, resulting in shallower concentration gradients in regions that have absorbed a significant amount of the solute. This occurrence signifies that the local concentration gradient is progressively diminishing as the internal solute concentration rises. The concentration field at 1080 minutes (18 hours) exhibits a more advanced spatial distribution, with the disparities between the states at 14 and 18 hours being less pronounced in several regions. This pattern aligns with the transition from a rapid diffusion-driven early phase to a more gradual and diffusion-restricted phase.

Fig. 4 shows concentration patterns that give a mechanistic description of the temporal development of SG. In the early stage, the concentration gradient between the osmotic solution and the mango tissue is relatively high, which prefers rapid uptake of solutes. The local concentration difference decreases in the presence of solute within the tissue, leading to a progressive decrease in the rate of solute penetration. The slowing down of SG with time should therefore be increasingly slower as the system moves towards a quasi-equilibrium state.

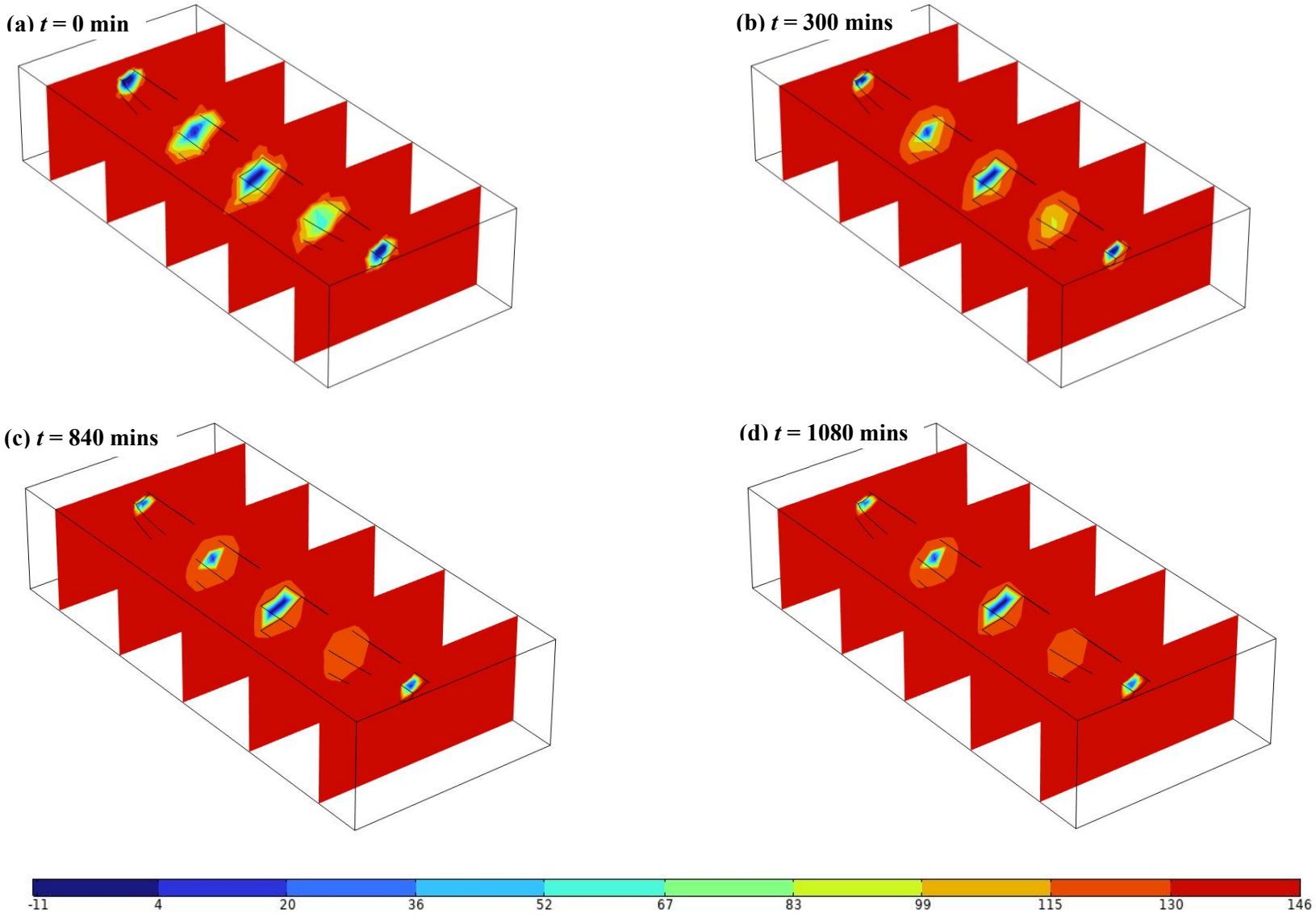


Fig.4. 3D temporal evolution of solute concentration distribution.

353

Fig. 5 depicts the significant development of the water concentration field. Unlike the migration of solutes, water during OD mostly flows into the mango tissue before transferring to the surrounding osmotic media. Therefore, the incremental change in water content in the mango provides the spatial framework for the expected WL.

The water concentration at t=0 min signifies the moisture level of the mango tissue at the outset. Following the initiation of OD, a concentration gradient develops at the interface between the mango and the solution. This gradient facilitates the migration of water in the tissue to the surrounding osmotic media. This culminates in a final spatial diversity in the water concentration landscape with the progression of dehydration.

At 300 minutes, the geographical differences in water content inside the mango area are observable. The concentration gradients indicate that the mass-transfer phenomenon is mostly focused in the areas adjacent to the mango-solution boundary. This is expected since the interface serves as the primary pathway via which water exits the tissue. This localized reduction of water creates concentration differences between the surface and subsurface regions. The concentration gradient of water throughout the procedure is then reallocated as the process advances to 840 minutes (14 hours). These concentration gradients penetrate farther inside the mango, indicating that dehydration occurs beyond the outside. Conversely, internal diffusion is assuming a more significant part in the process of water transfer. This spatial occurrence is significant since the macroscopic WL response relies on the diffusivity of water inside the mango matrix to reach the outer surface, in addition to the mass-transfer conditions at the interface. The distribution of water concentration at 1080 minutes (18 hours) has a highly established spatial configuration, but changes compared to the prior phase are less pronounced in certain regions. This behavior suggests that the primary impetus for further water extraction diminishes progressively. As the tissue progressively loses water, the concentration gradient that drives mass transfer diminishes, leading to a reduction in the immediate rate of dehydration.

The spatial evolution seen in Fig. 5 may thus provide a mechanical explanation for the characteristic form of the WL time curve. The rapid onset of WL is associated with steep concentration gradients at the mango-solution interface, whereas the subsequent deceleration in WL development over time results from the decreasing concentration gradient and the increasing resistance to internal water flow.

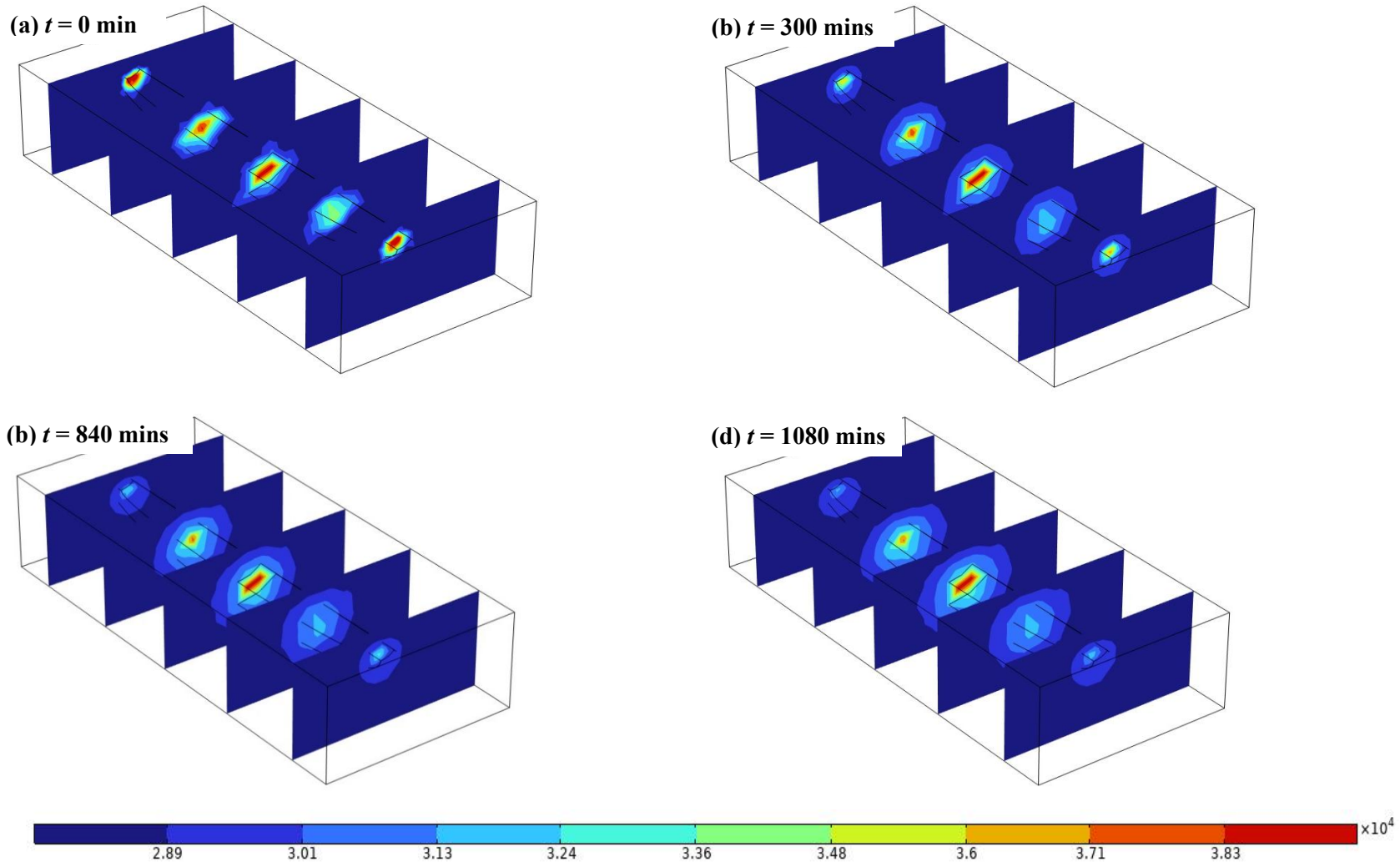


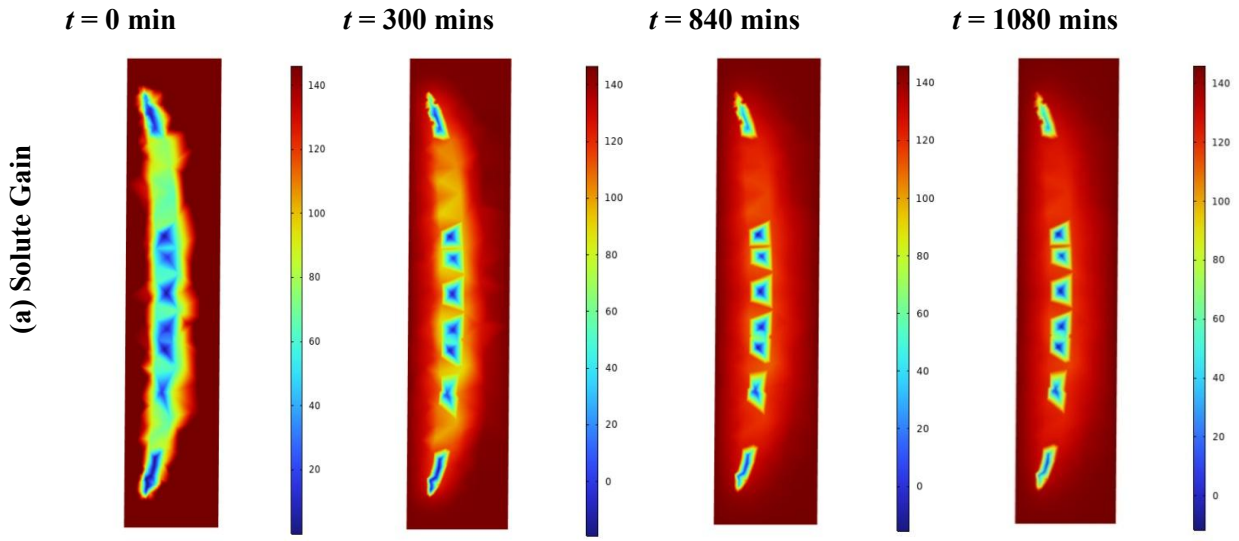
Fig. 5. 3D temporal evolution of Water concentration distribution

384

4.2. Spatial Evolution of Mass Transfer

Spatial dynamics of the water and solute transport was determined to explain the internal process of mass-transfers involved in the OD of the mango tissue. Although the total loss of water (WL) and the solutes gain (SG) provides macroscopic measurements of dehydration performance, spatially resolved concentration fields are needed to determine the rate at which the osmotic treatment diffuses across the mango-solution interface into the interior. The current FE model facilitates such a spatial behaviour to be represented at various points in the mango, mango-solution interface (MSI), subsurface region (SSR), intermediate region (IR), and core region (CR). It was predicted that at 0, 300, 840, and 1080 min (0, 5, 14 and 18 h), the distributions were as illustrated in Fig. 6 and the point-wise responses as marked in Fig. 7.

The time evolution of the simulated solute concentration field associated with SG is depicted in Fig. 6(a). The two domains' opposing spatial development clearly illustrates the counter-current nature of the OD process: water leaves the mango tissue while the osmotic solute enters. The local concentration gradients are altered continuously throughout the treatment as various processes transpire concurrently. Spatial distributions of Fig. 6 will provide the physical basis for macroscopic WL and SG kinetics determined by the model.



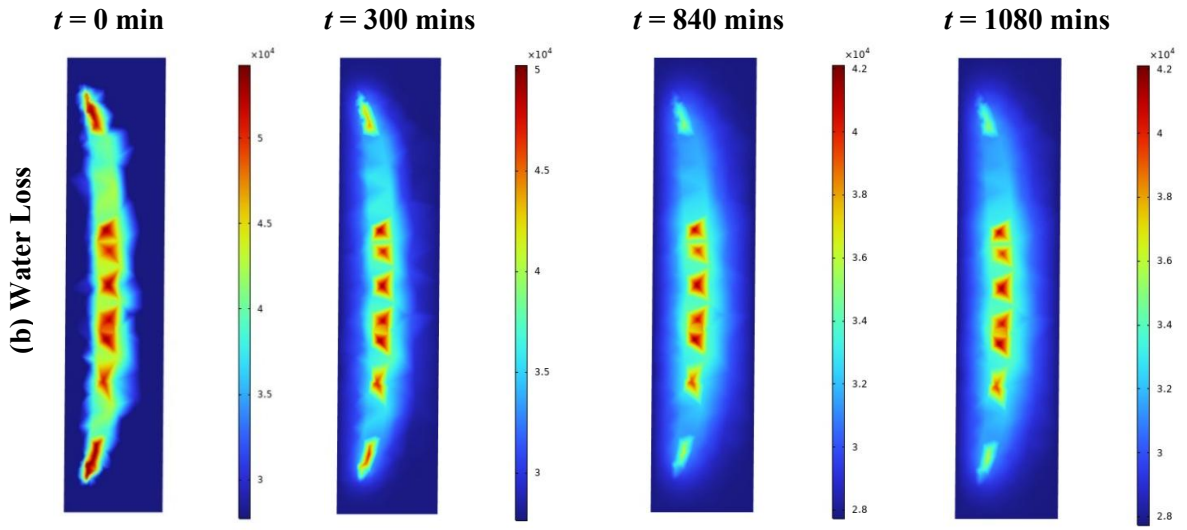


Fig.6. 2D spatial evolution of the simulated mass-transfer fields at (a) solute gain and (b) water loss for t = 0, 300, 840, and 1080 min.

This phenomenon of spatial location influencing solute absorption is also shown in Fig. 7(a). The temporal behavior of the SG responses at the CR, IR, SSR and MSI is distinctly different. The osmotic treatement is the most rapidly influential on the MSI because it is in direct contact with the external osmotic environment. The concentration at the SSR is following the interfacial response while the IR and CR are showing a delayed response, due to progressive diffusion of the solute in the surrounding tissues. At a certain treatment duration, the outside region may become considerably enriched with solute, while the inner region remains essentially unaltered. As the solute permeates further into the tissue, the differences in spatial concentration become less pronounced, as the location-dependent curves converge at longer times.

Fig. 7(b) depicts the geographic relationship of water movement. The concentration related to WL, in contrast to SG, is linearly decreasing with time across all investigated locations. The rapid reduction occurs in the MSI where the water at the outer surface is directly exposed to the osmotic driving force. The SSR and IR respond subsequently, but the CR retains a notably high. This diminishment in slope reflects a reduction in the velocity of instantaneous water movement as the internal concentration gradient develops. The methodology regarding the gradually altering values over extended periods aligns with the reduction of the concentration gradient and the escalation of resistance to internal moisture movement.

The spatially differentiated behavior provides a mechanical explanation for the general reduction in overall WL. The reduction in WL rate over prolonged treatment durations is not just due to diminished water

surface availability; it also signifies the gradual attenuation of the internal concentration gradient and the growing significance of diffusion inside the mango matrix.

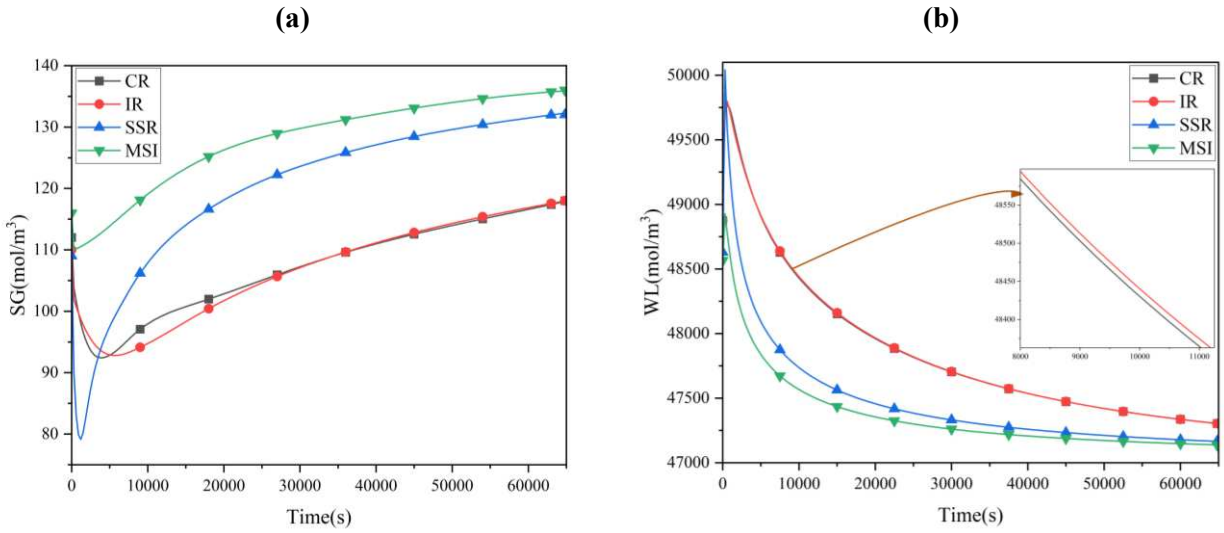


Fig.7. Location-dependent temporal evolution of (a) solute gain and (b) water loss.

The overall correlation between water extraction and solute absorption is seen in Fig. 8. The concentration of the predicted water-related diminishes consistently with time, but the concentration of the solute-related escalates gradually. The opposing tendencies demonstrate the concurrent dehydration and solute incorporation occurring inside the mango tissue.

The crucial element of Fig. 8 is the progressive alignment of the two responses toward one another in the latter stages of treatment. This intersection indicates that, two opposing mass-transfer processes are occurring concurrently rather than suggesting the water loss and solute gain will be quantitatively equal. Their varying rates are due to the different transport properties of water and solute in the mango matrix and to their concentrations gradients.

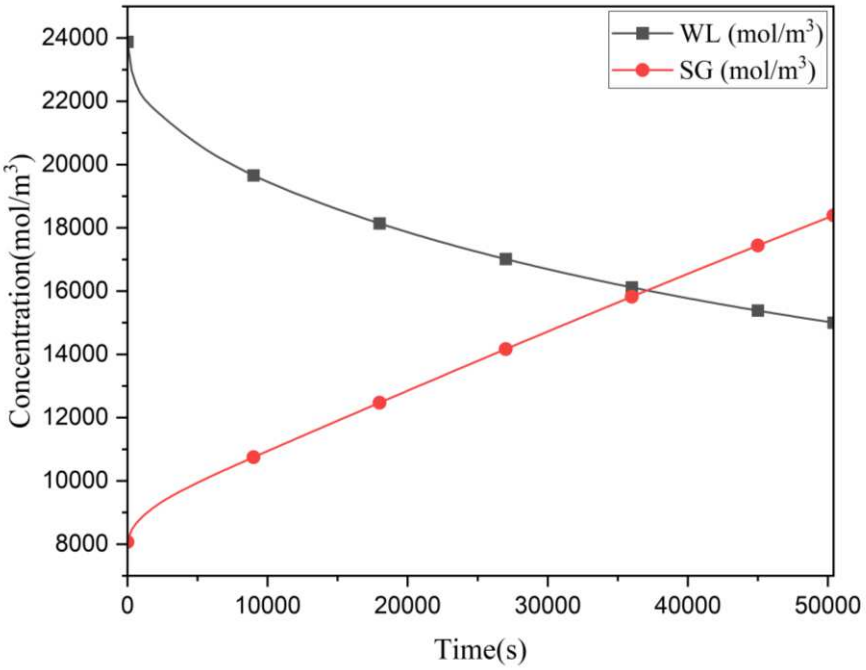


Fig.8. Evolution of the modeled water- and solute-related concentration.

4.3 Effects of Temperature and Solution Concentration

Two primary process factors that regulate mass transfer in OD are temperature and the concentration of the osmotic solution. Their performance was evaluated according to the projected temporal variations in solute concentration, water concentration, total solute accumulation, and total water loss. The simulations were performed for 64,800 seconds (18 hours) to see the first transient reaction and the later shift to a slower mass-transfer regime, with the influences of solution concentration and temperature shown in Figs. 9 and 10, respectively.

Fig.9(a) illustrates the predictions of the time-dependent solute concentrations for osmotic solution concentrations of 5, 7, and 10% Brix. The three circumstances demonstrate a continuous increase in solute concentration while maintaining a consistent treatment duration. The expansion is very substantial in the early stages and progressively diminishes with time. This temporal phase aligns with the ongoing reduction in concentration gradient as the solution permeates the mango matrix, resulting in the tissue being progressively saturated with the osmotic solute. The temporal dynamics of the simulated solute concentration are not fully controlled by the extra Brix degree that is supplied. The reaction is a product of the concentration gradient implemented, the concentration redistribution in the system, and the transport features dictated by the model. The gradual convergence of the solute-concentration curves with extended treatment duration is noteworthy. The outcome, however, is influenced by both the established

concentration gradient and the internal redistribution of concentration, together with the transport characteristics used in the model. The osmotic parameters established at the onset of dehydration significantly influence the concentration gradient. As the process progresses, the solute concentration inside the mango will approximate that of the external environment, resulting in a reduced rate of concentration change. Consequently, the curves converge toward a gradually changing state with extended treatment durations. Figure 9(b) illustrates the temporal variation of simulated water content for the three Brix values, revealing a decline in the water content of mango tissue with time, signifying the gradual extraction of moisture from the tissue. The extent of the reduction is significantly influenced by the osmotic concentration. It is anticipated that an increase in Brix correlates with a reduction in water concentration, since elevated Brix levels provide a stronger osmotic driving force due to the more concentrated external solution. An increased concentration of the osmotic medium results in a larger disparity in the water chemical potential between the mango tissue and the surrounding fluid, hence promoting the outward movement of water. The 10% Brix condition exhibits the most significant reduction in water, following by the 7% Brix and 5% Brix values. This distinction is prominently evident at the first development stage and remains robust throughout the simulation.

All three concentrations exhibit a diminishing rate of water loss during extended treatment durations. This gradual leveling of the curves suggests that the initial impetus diminishes progressively as dehydration progresses. Thus, the effect of the rise in osmotic concentration is most accurately represented by the extent of water extraction, whereas the additional influence on the simulated internal solute concentration is comparatively less in the existing parameterization

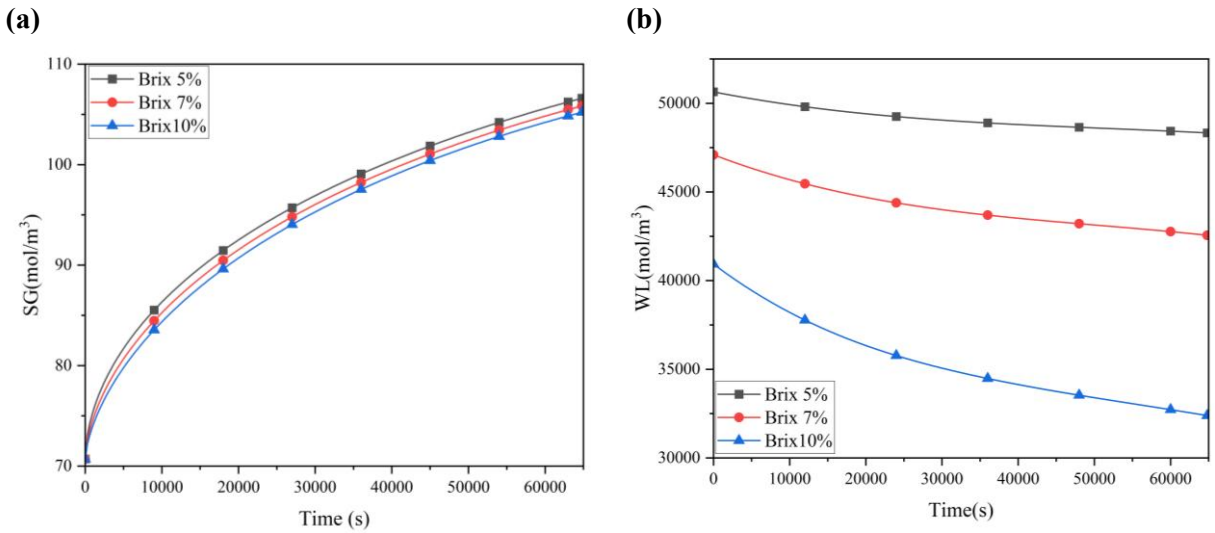


Fig.9. Effect of osmotic solution concentration on the temporal evolution of (a) solute concentration and (b) water concentration at 5, 7, and 10% Brix.

The influence of temperature on total solute acquisition and total water depletion is seen in Fig. 10 at 31°C, 41°C, and 50 °C, respectively. Simulations on temperature dependency indicate that the intake of solutes and the removal of water exhibit varied responses with each mass-transfer process reacting differently to temperature, under the model's present settings.

Fig. 10(a) illustrates the predicted cumulative solute gain (SG) over the simulation period. In the first phase, the rate of solute accumulation escalates rapidly with rising temperature, however it gradually diminishes in the subsequent phase at 304 K. The SG profiles at 314 K and 323 K are significantly diminished and exhibit a somewhat consistent trend throughout the experiment. This temperature-induced variation becomes more apparent during the first phase; after 18 hours of simulation, the cumulative SG for the 304 K scenario is considerably higher. The conduct in the early phases is also indicative. Cumulative SG curves indicate the highest modifications in the early stages followed by a progressively reducing rate of change. This behavior is analogous to the reduction in the solute concentration gradient owing to the rise in the solute concentration in the mango. Consequently, while the external temperature remains elevated, the reduction in the concentration gradient may restrict the following enhancement of solute absorption.

In contrast to the SG response, cumulative WL rises systematically with temperature is shown in Figure 10(b). The findings reveal that the most significant water loss transpires at 323 K, followed by 314 K and 304 K. As the duration of treatment extends, the disparities among the temperature-dependent curves become more pronounced, with the peak moisture extraction anticipated at the conclusion of the simulation achieved at the highest temperature applied.

The model parameters demonstrate that larger molecule mobilities and temperature-dependent effective transport parameters are linked to higher losses at higher temperatures. As the temperature rises, the simulated transport mechanism becomes increasingly advantageous for the movement of water inside the mango matrix. As a result, WL will accumulate more quickly at higher temperatures. On the other hand, the cumulative WL curves do exhibit a clear temperature-dependent pattern throughout the experiment. At 18 hours, when the temperature is at its highest, the most water is extracted overall. This is consistent with the previously described spatial concentration data, which indicated a relationship between the gradual loss of water and the build-up of concentration gradients in the mango tissue, followed by a drop in concentration.

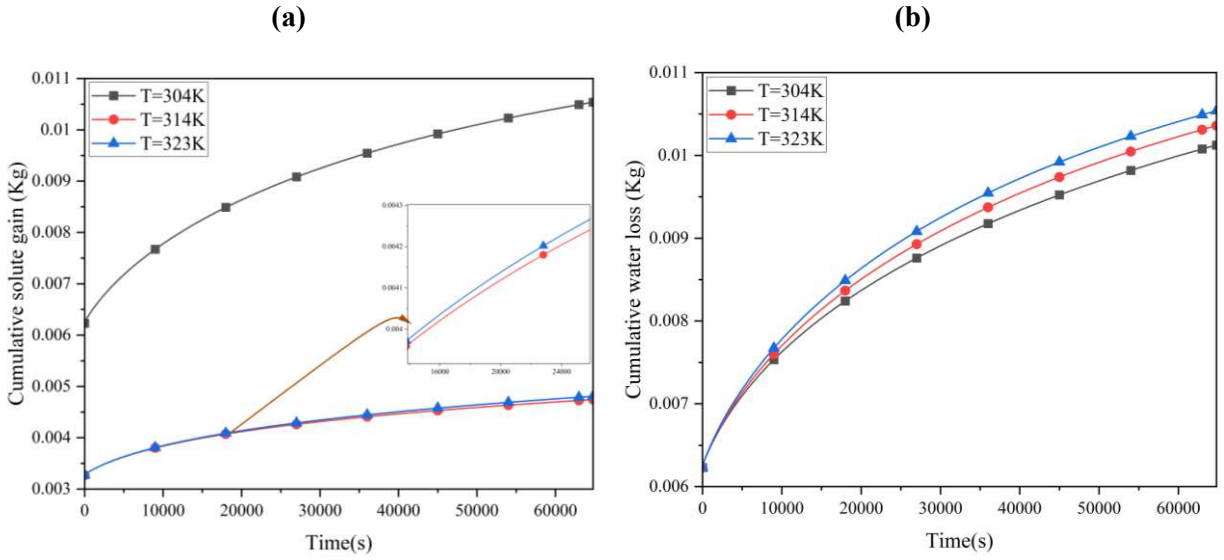


Fig.10. Effect of temperature on the temporal evolution of (a) cumulative solute gain and (b) cumulative water loss at 304, 314, and 323 K.

4.4. Evolution of Average Water and Solute Concentrations

Fig. 11 illustrates the temporal fluctuation of the geographic average concentrations of solute and water in mango tissue. The mean concentration of stevia consistently increased from about 188 mol m^{-3} to 250 mol m^{-3} over the simulated period, indicating that the solutes were persistently absorbed by the mango matrix. This escalation was rather fast initially and diminished progressively over time, as the concentration gradient for solute transport lessened. On the other hand, the average water concentration has steadily reduced from approximately 49700 mol/m^3 to 47300 mol/m^3. During the first phase of treatment, the rate of reduction was at its highest, and as the duration of treatment increased, the rate of reduction had a gradual decrease. It shows that the water migration is initially facilitated by a high concentration gradient between the mango tissue and osmotic medium. This gradient will decrease as the mango desiccates.

Opposing tendencies of the two average concentrations shows that OD is a coupled process. Water gradually permeates the mango tissue while concurrently, the osmotic solution infiltrates the mango matrix. Both variables exhibit a diminishing rate of change with time, suggesting a gradual progression toward a "quasi-equilibrium" condition. These responses are averaged and relate to the spatial concentration distribution outlined in the preceding section, providing a quantitative depiction of the overall internal mass transfer processes.

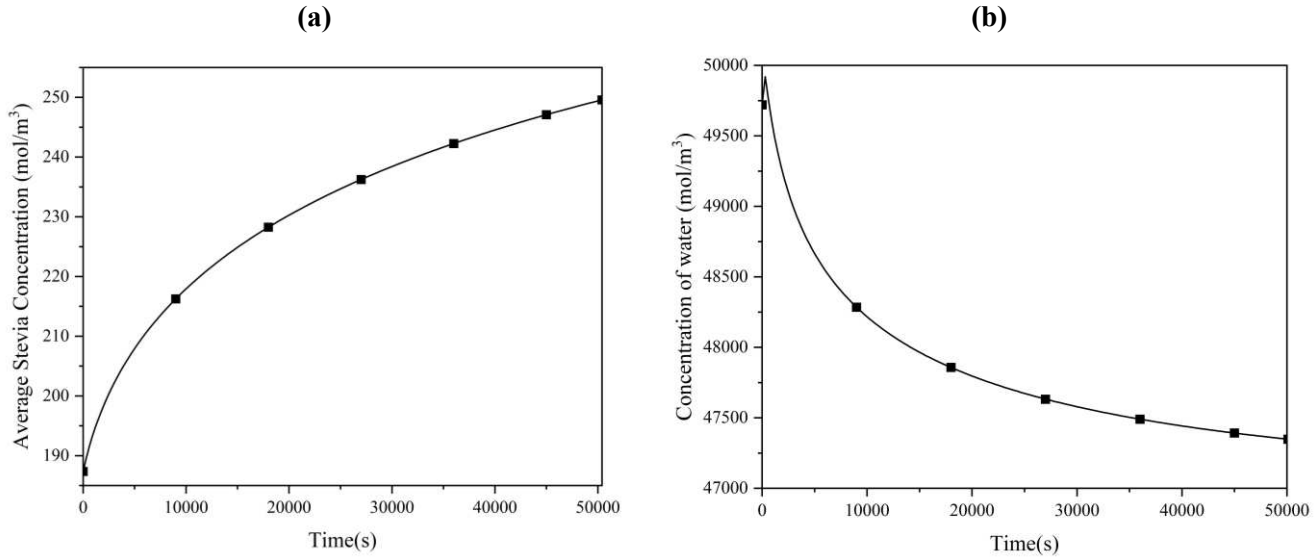


Fig.11. Temporal evolution of the spatially averaged (a) stevia concentration and (b) water concentration.

4.5. CFD-Integrated Machine-Learning Surrogate Modelling

To facilitate the use of the high-fidelity numerical model in a more efficient way, a machine-learning (ML) surrogate model was developed to approximate the water loss (WL) and solute gain (SG) responses of the simulated OD process. The development of the ML model was conceived as a data-centric substitute for the numerical model, rather than as a replacement for the mechanical structure of the model. Consequently, temperature, osmotic-solution concentration (measured in °Brix), and immersion duration were identified as predictor factors, whereas water loss (WL) and Solute Gain (SG) were represented as response variables. This approach exemplifies the surrogate-modeling framework, whereby the outcomes of a computationally intensive simulation are used to train a computationally efficient machine learning (ML) model, enabling rapid predictions within the parameter space defined by the simulations.

The run-wise validation method was used to evaluate the generalization capacity of the machine learning models. Given that individual CFD simulations yield multiple temporally correlated observations, conventional random row-wise partitioning may result in observations from the same simulation trajectory being included in both training and testing subsets, potentially leading to excessively optimistic assessments of predictive efficacy. To avert the problem of data leaking, we partitioned the simulations according to the Run_ID, executing complete simulation trajectories only on one of the two datasets: either model development or independent testing. Approximately 80 percent of all simulation executions were dedicated

to model creation, while the remaining simulation runs were held out as an independent test set for final model evaluation. Five-fold GroupkFold cross-validation was performed on the model-development dataset with Run_ID as the grouping variable. Since each COMSOL simulation resulted in many temporally linked data, classifying by Run_ID guaranteed that observations from the same simulation trajectory were not given to both the training and validation groups at the same time. Consequently, validation was performed at the simulation-trajectory level instead of via random observation-level segmentation, offering a more rigorous evaluation of the surrogate model's capacity to generalize to novel simulated trajectories.

Five forecasting methodologies were assessed: a mean predictor serving as a baseline, polynomial Ridge regression, Random Forest (RF), XGBoost, and CatBoost. Model efficacy was mostly evaluated by RMSE, in conjunction with MAE and R^2, to measure prediction inaccuracies and the quality of fit.

Table 2. Grouped five-fold cross-validation performance.

Model	Target	**RMSE**	**MAE**	**R^2**
CatBoost	Solute Gain (SG)	0.10318 ± 0.01642	0.07612 ± 0.00909	0.99245 ± 0.00141
	Water Loss (WL)	0.58411 ± 0.10322	0.43563 ± 0.06450	0.99243 ± 0.00142
XGBoost	Solute Gain (SG)	0.12723 ± 0.02128	0.09564 ± 0.01318	0.98782 ± 0.00482
	Water Loss (WL)	0.67693 ± 0.10311	0.49876 ± 0.05706	0.98981 ± 0.00165
Random_Forest	Solute Gain (SG)	0.12506 ± 0.02437	0.09128 ± 0.01831	0.98889 ± 0.00289
	Water Loss (WL)	0.69851 ± 0.13754	0.51415 ± 0.10260	0.98908 ± 0.00312
Ridge	Solute Gain (SG)	0.23780 ± 0.06015	0.16901 ± 0.03914	0.96092 ± 0.00976
	Water Loss (WL)	1.11275 ± 0.28383	0.80758 ± 0.18986	0.97246 ± 0.00895
Mean Baseline	Solute Gain (SG)	1.19806 ± 0.16788	1.00078 ± 0.14638	-0.00593 ± 0.00821
	Water Loss (WL)	6.71358 ± 0.64003	5.57651 ± 0.54244	-0.00376 ± 0.00526

Table 2 demonstrates a markedly enhanced predictive capability relative to the average baseline. WL's RMSE baseline score of 6.71358 and SG's RMSE baseline score of 1.19806 were reduced to 0.58411 and 0.10318 respectively with the use of CatBoost. CatBoost attained the minimal MAE for both target responses (WL=0.43563 and SG=0.07612) with corresponding R^2 values of 0.99243 and 0.99245, respectively.

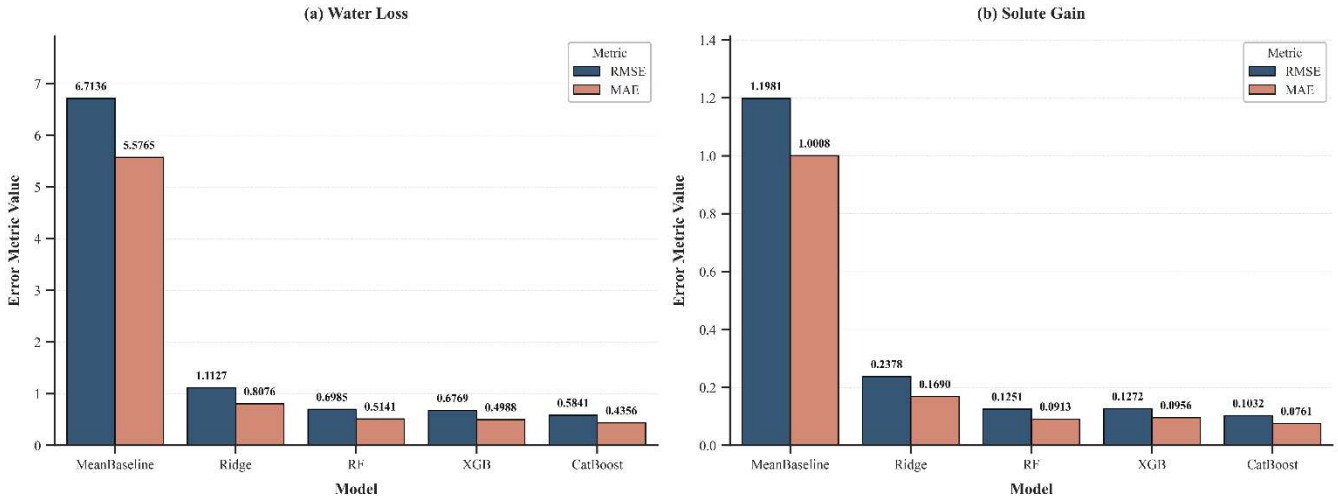


Fig. 12. Comparative cross-validation performance of the evaluated machine-learning surrogate architectures for (a) water loss (WL) and (b) solute gain (SG)

The alternative ensemble models also much surpassed the baseline. The RMSE achieved by XGBoost was 0.67693 for WL and 0.12723 for SG, while Random Forest recorded 0.69851 for WL and 0.12506 for SG. Ridge regression yielded much more errors, with an RMSE of 1.11275 for WL and 0.23780 for SG. In contrast, the average baseline had significantly elevated errors and R^2 values approaching zero. The findings indicate that the response dynamics generated by the numerical simulations contain nonlinear correlations that are inadequately represented by a basic baseline or polynomial regression model, but may be investigated using ensemble-learning methodologies.

Based on results CatBoost consistently outperformed the other architectures in grouped cross-validation, it was selected for subsequent independent testing on unseen simulation trajectories for both target variables. The chosen model was then retrained with the full model-development set and tested solely on independent, unseen simulation trajectories. The independent-test results showed that there was a high level of agreement between CatBoost predictions and the respective numerical-model responses.

4.5.1. Prediction accuracy on unseen CFD simulations

The independent-test results are presented in **Fig. 12**. For WL, the CatBoost surrogate achieved $R^2 = 0.9943$, RMSE = 0.4772, and MAE = 0.3620. For SG, the corresponding values were R^2=0.9913, RMSE = 0.1028, and MAE = 0.0748.

The prediction values are clustered tightly around the optimal y=x correlation for both responses, indicating strong concordance between the CFD-derived and ML-predicted values. The comparatively narrow spread

at the parity line indicates that the surrogate effectively captures the primary variance in the CFD responses. The crucial aspect to highlight is that the performance metrics were derived from simulation executions that were excluded from model training, and the data was obtained from trajectories that were not chosen at random. This demonstrates trajectory-level generalization rather than just interpolation among the training data.

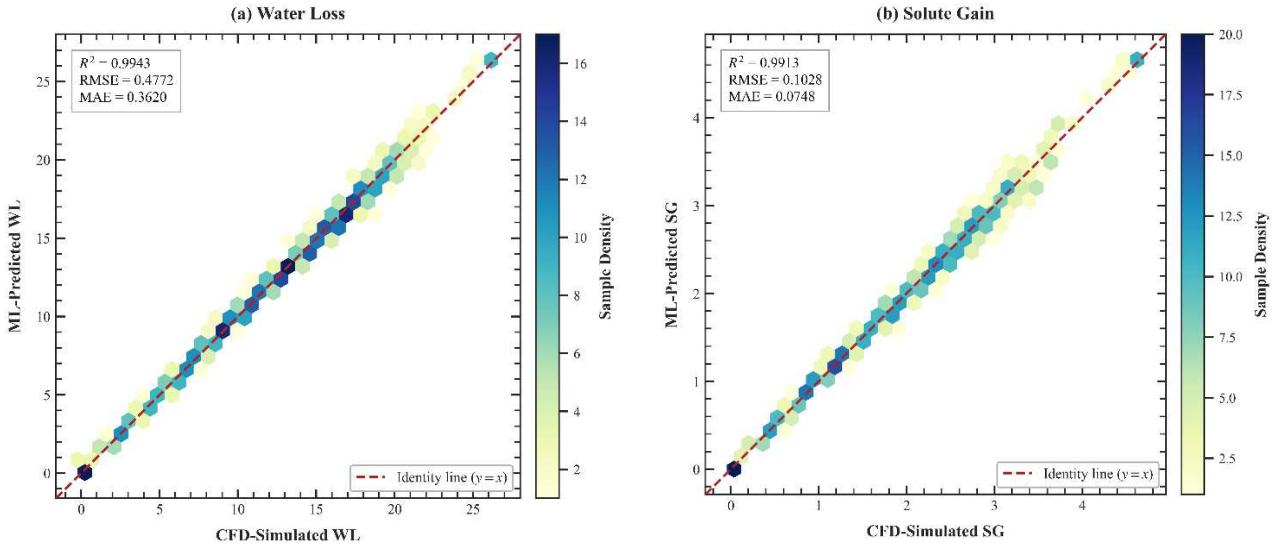


Fig. 13. Parity evaluation of the CatBoost surrogate model for (a) Water loss(WL) and (b) Solute Gain (SG)

The capacity to precisely forecast a specific outcome does not ensure that a continuous transient CFD process can be replicated by a machine learning surrogate model. Consequently, the temporal accuracy of the CatBoost model was evaluated using an additional unobserved simulation trajectory, as seen in **Fig. 13**.

The simulated trajectories of the machine learning predictions for both WL and SG closely correspond with the corresponding computational fluid dynamics results throughout the whole simulation duration. The surrogate demonstrates the standard nonlinear temporal response, characterized by a relatively fast alteration in the first phase of treatment, followed by a plateau for extended treatment durations. This agreement indicates that the model encapsulates the time progression of the mass-transfer process with specific CFD response values. This arrangement is particularly significant for OD, in which WL and SG are adaptive responses influenced by shifting concentration gradients. The surrogate's capacity to recreate the whole transient trajectory therefore offers further proof that the trained model maintains the critical temporal aspects of the high-fidelity CFD solution.

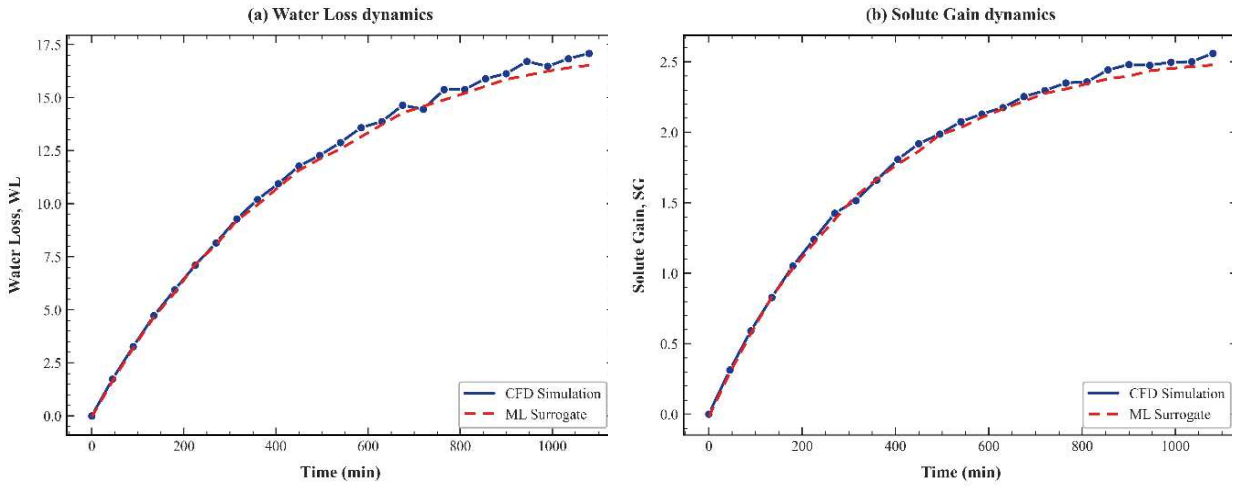


Fig. 14. Dynamic trajectory tracking and validation of the ML surrogate model against CFD numerical simulations for an unseen operational run at a constant condition of at $T = 51.96° \; C \; and \; 5 \; °Brix$ (a) Water Loss (WL) dynamics, and (b) Solute Gain (SG) dynamics over processing time.

590

591 **4.5.2. Comparative performance of machine-learning architectures**

592 The relative efficacy of the assessed surrogate designs is shown in Fig.14. The mean baseline had
593 significantly larger errors for both WL and SG, as a constant mean prediction was inadequate for the
594 nonlinear predicted mass transfer responses from the CFD simulations. The baseline was enhanced using
595 ridge regression; nevertheless, the precision did not reach the levels of tree-based ensemble techniques. The
596 numerical findings shown in Table 2 indicate that among the ensemble models, the CatBoost model has the
597 lowest RMSE and MAE for both WL and SG.

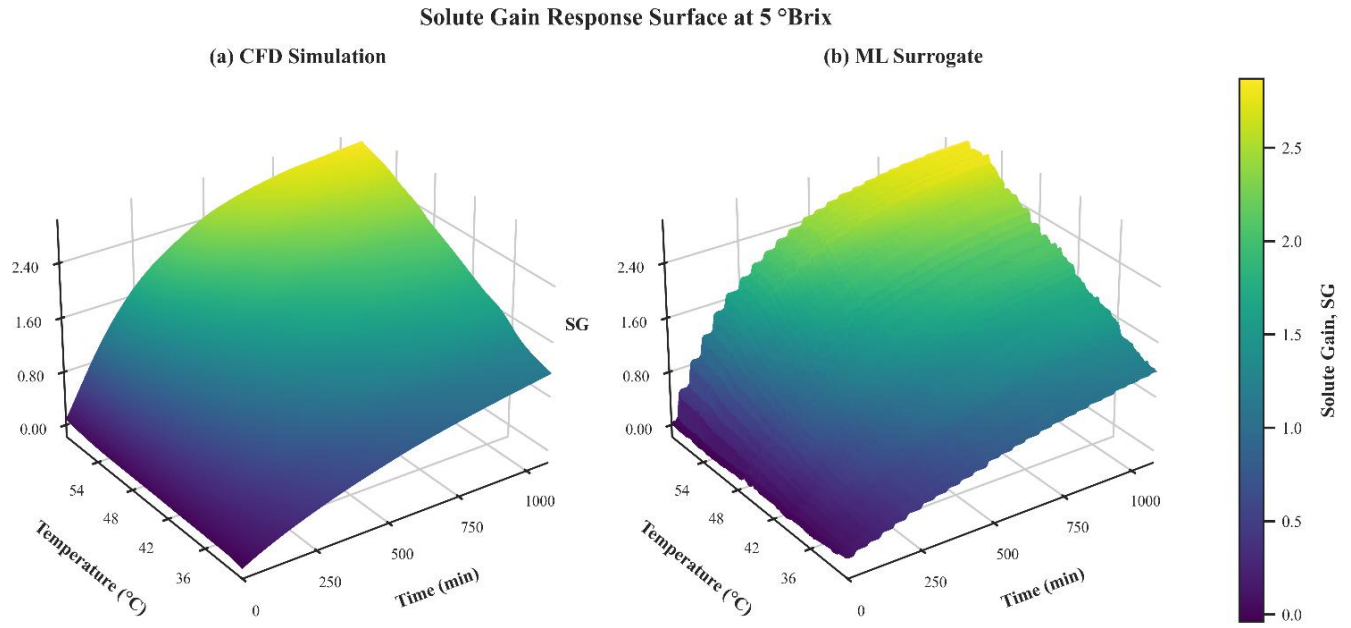
598 The findings indicate that the correlations among immersion duration, temperature, °Brix of the osmotic
599 solution, WL, and SG are fundamentally nonlinear, as seen by superior fitting with CatBoost, XGBoost,
600 and Random Forest compared to Ridge regression. This behavior is physically analogous to the transient
601 and coupled nature of osmotic dehydration, where the mass-transfer driving forces change continuously
602 with time as a result of changes in water and solute concentrations inside the mango tissue and outside the
603 osmotic solution. The attenuation of concentration gradients is responsible for the nonlinear temporal
604 evolution of water loss and solute gain. The gradual reduction of concentration gradients, along with
605 changes in internal transport resistance, accounts for the nonlinear progression of WL and SG during
606 processing time. Consequently, the findings of the machine learning study corroborate the mechanistic
607 computational fluid dynamics model and suggest that the response space produced by CFD is

multidimensional, characterized by nonlinear interactions among the process variables that may be proficiently captured by ensemble-based surrogate models.

4.5.3. Prediction of temperature, time and Brix response surfaces

At various osmotic-solution concentrations, the chosen surrogate, CatBoost, was further assessed by comparing its multidimensional CFD responses to the temperature-time response surfaces. Both the SG and WL response surfaces at 5%, 7%, and 10% Brix are shown in Figures 15 and 16, respectively.

At each Brix level, SG increases with time and responds differently depending on temperature represented in Fig 15; the CFD-derived surfaces for SG correspond well with the ML-predicted surfaces for the examined time and temperature ranges. Results from both CFD and ML were in agreement across all three Brix conditions, indicating that the surrogate could not only imitate one operating state but also capture the combined influence of processing time and temperature.



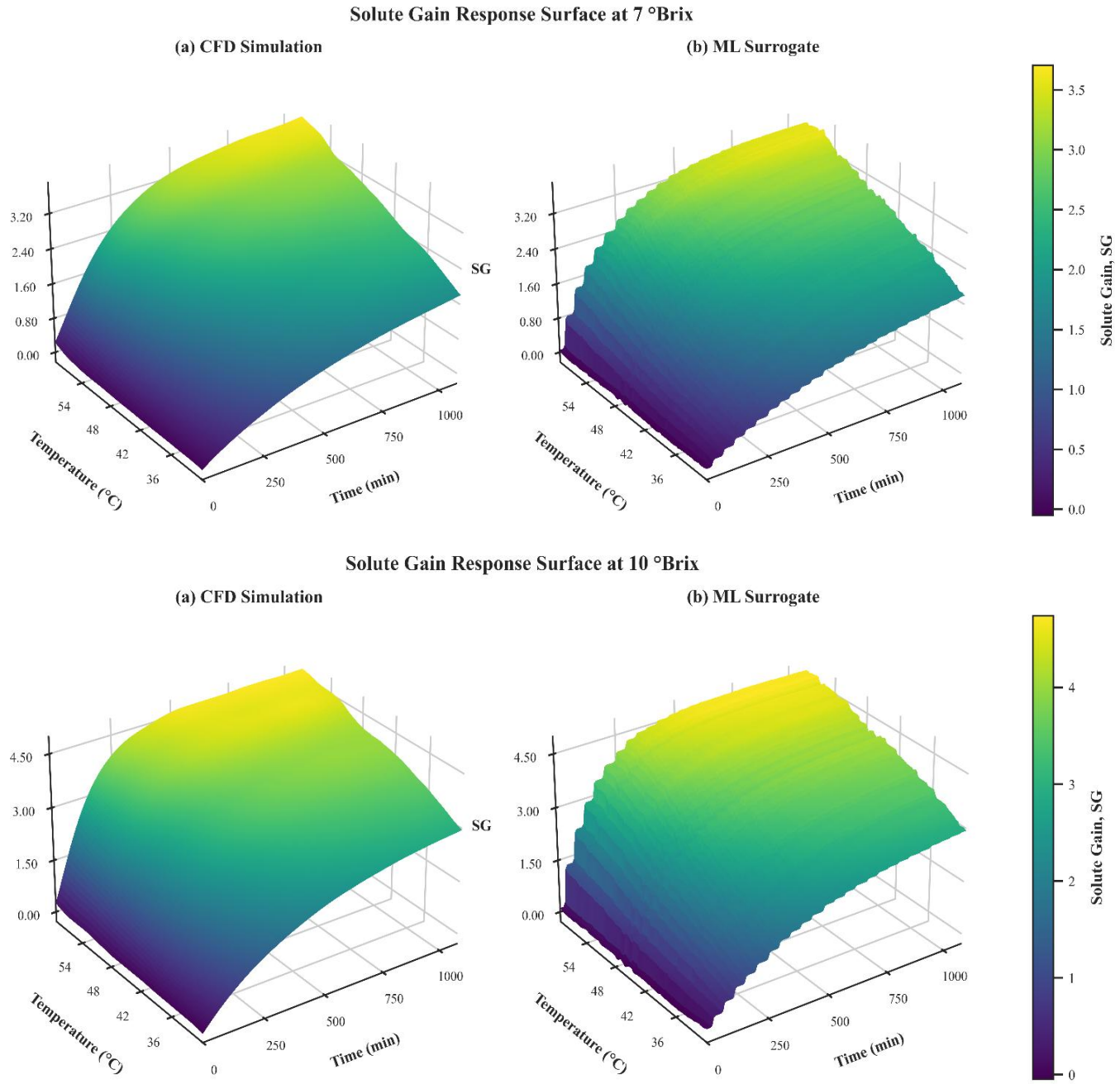
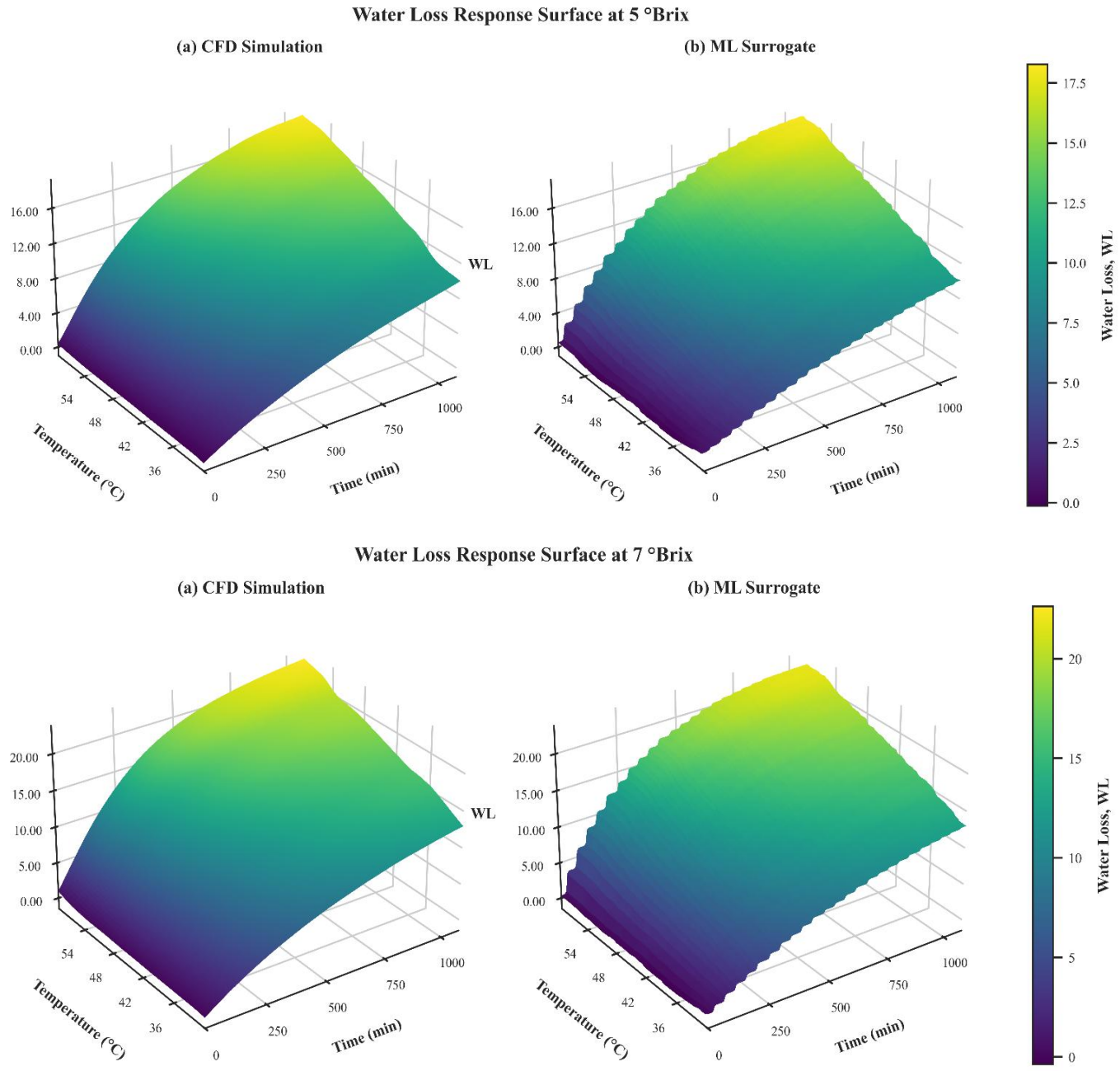


Fig. 15. Comparison of CFD-derived and CatBoost-predicted solute-gain (SG) response surfaces as functions of processing time and temperature at (a) 5% Brix, (b) 7% Brix, and (c) 10% Brix

The situation is same for WL in Fig. 16. The anticipated surfaces align well with the CFD-generated surfaces at 5, 7, and 10% Brix. The findings are shown as response surfaces that depict the nonlinear correlation between WL and processing time and temperature, which fluctuates with the concentration of the osmotic solution. The significant resemblance between the actual and anticipated surfaces indicates that the surrogate preserves the essential characteristics of the response surface produced by the CFD.


Water Loss Response Surface at 5 °Brix
(a) CFD Simulation
(b) ML Surrogate
WL
16.00
12.00
8.00
4.00
0.00
54
48
42
36
Temperature (°C)
0
250
500
750
1000
Time (min)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Water Loss, WL
Water Loss Response Surface at 7 °Brix
(a) CFD Simulation
(b) ML Surrogate
WL
20.00
15.00
10.00
5.00
0.00
54
48
42
36
Temperature (°C)
0
250
500
750
1000
Time (min)
20
15
10
5
0
Water Loss, WL

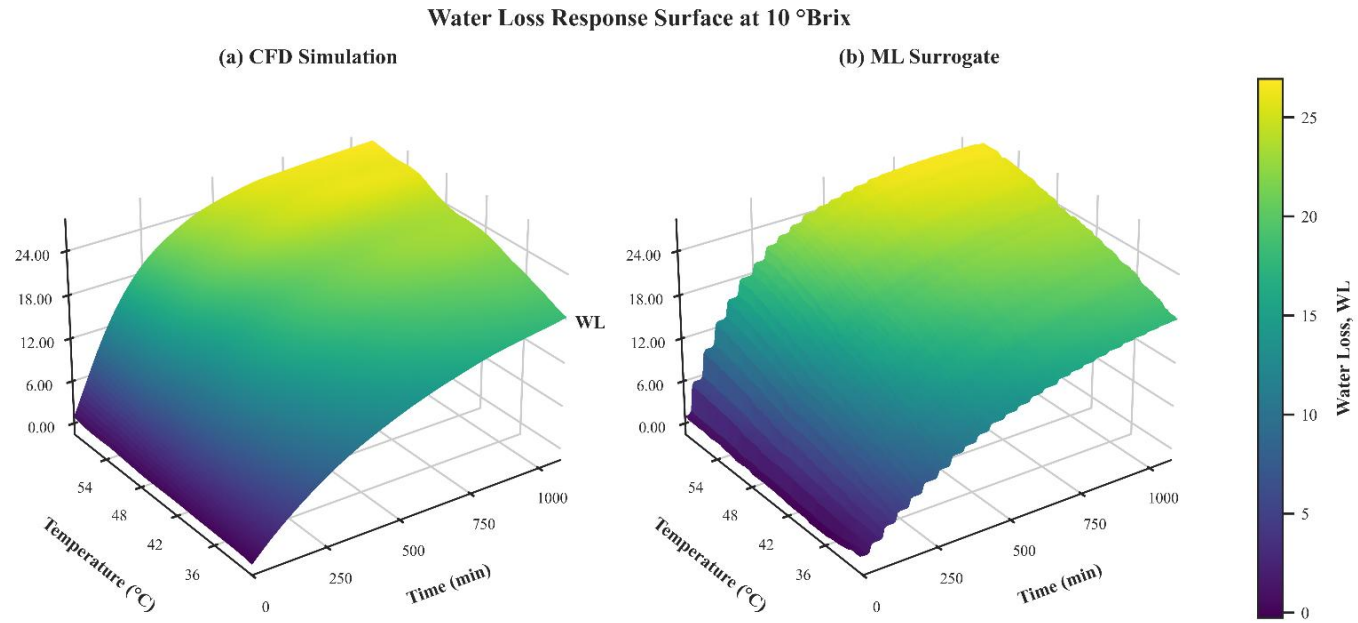


Fig.16. Comparison of CFD-derived and CatBoost-predicted water-loss (WL) response surfaces as functions of processing time and temperature at (a) 5% Brix, (b) 7% Brix, and (c) 10% Brix

624

625 Figs 15 and 16 together illustrate that the machine learning model is capable of replicating the interrelated
626 effects of time, temperature, and osmotic-solution concentration on the two primary mass-transfer reactions.
627 The examined operational range includes around (0 -1000) minutes, 36, 42, 46, and 54°C, along with 5, 7,
628 and 10% Brix.

629 Beyond prediction precision, the primary practical advantage of the surrogate is the significant reduction in
630 computational expenses. The comprehensive CFD assessment required around 120 seconds for each
631 evaluation (500 response evaluations), in contrast to the trained surrogate (WL 0.00224 seconds and SG
632 0.00206 seconds). This corresponds to an approximate 53,571-fold speedup for WL and a 58,252-fold
633 speedup for SG in computational prediction time compared to the numerical simulation. The CFD model
634 serves as the precise, mechanistic benchmark, while the surrogate provides an expedited method for
635 parameter screening, sensitivity assessment, iterative process evaluation, and optimization-focused
636 computations.

637

638

5. Conclusions

This study developed and validated a coupled finite element–machine learning (FE–ML) code to implement the OD technique on mango slices with stevia, integrating transport physics within a realistic three-dimensional framework alongside a data-driven surrogate model for expedited process prediction. The model was resolved for a treatment duration of 18 hours, with solution concentrations of 5, 7, and 10 °Brix, and temperatures of 304, 314, and 323 K.

The simulations demonstrated that the concentration gradients occurred from the interface of the mango solution to the core, reducing as the OD moved closer to quasi-equilibrium, and that the mass transfer was both geographically and temporally heterogeneous during the OD. Concurrently, a decline in water concentration (WL) was observed, with the mean water concentration dropping from about 49,700 to 47,300 mol m^{-3}, while the average solute concentration surged from about 188 to 250 mol m^{-3}. As the concentration of the osmotic solution increases, the volume of water elimination also increases, and higher temperature levels correspond to an increase in water loss and the initial rate of solute gain.

The simulation-generated dataset was used to train machine-learning surrogates using temperature, solution concentration, and time as predictors. CatBoost achieved grouped cross-validation RMSE values of 0.58411 for WL and 0.10318 for SG, and independent-test values of 0.4772 and 0.1028, respectively. It also reproduced the transient responses and response surfaces while reducing computational time by approximately 53,571-fold for WL and 58,252-fold for SG compared with the finite-element simulations. Overall, the integrated FE–ML framework offers spatially resolved insight and rapid prediction for process screening and optimization within the investigated operating range.

The main flaw in this research is that no tests were carried out to confirm the findings under the identical stevia-mango settings. As a consequence, the results reported here are merely predictive and numerically consistent when compared to the simulation data. Subsequent investigations should thus concentrate on matched-condition trials within the temperature and concentration spectrum examined, including the calibration of parameters and independent experimental verification. The combined finite-element and machine-learning architecture provides a basis for geographically detailed analysis, rapid evaluation of processes, and subsequent enhancement of stevia-assisted OD of mango, contingent upon additional validation.

Acknowledgements

The authors also sincerely acknowledge to The Simulation Lab Department of Mathematics, CUET, Bangladesh and The Centre of Excellence in Mathematics, Mahidol University, Bangkok, Thailand, for providing COMSOL Multiphysics® software support. The authors gratefully acknowledge Ms. Mim Saha for her valuable assistance in the preparation and refinement of the graphs and figures presented in this manuscript.

Funding

The authors received no specific funding or financial support for this research, authorship, or publication of this manuscript.

Authors Contribution

Ripa Muhury: Conceptualization, investigation, methodology, resources, simulations and model validation, visualization, writing – original draft. **Usama Ibn Aziz:** Formal analysis, visualization, writing – original draft. **Md. Sakib Ahmmed:** Data curation and analysis, visualization, writing and editing. **Ferdusee Akter:** Supervision, writing – review and editing. **Ujjwal Kumar Deb:** Supervision, writing – review and editing.

Data Availability Statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] A. Asghari, P. A. Zongo, E. F. Osse, S. Aghajanzadeh, V. Raghavan, and S. Khalloufi, "Review of osmotic dehydration: Promising technologies for enhancing products' attributes, opportunities, and challenges for the food industries," May 01, 2024, *John Wiley and Sons Inc*. doi: 10.1111/1541-4337.13346.

[2] F. R. Assis, R. M. S. C. Morais, and A. M. M. B. Morais, "Mass Transfer in Osmotic Dehydration of Food Products: Comparison Between Mathematical Models," Jun. 01, 2016, *Springer New York LLC*. doi: 10.1007/s12393-015-9123-1.

[3] M. Maldonado and J. González Pacheco, "Mathematical modelling of mass transfer phenomena for sucrose and lactitol molecules during osmotic dehydration of cherries," *Heliyon*, vol. 8, no. 1, Jan. 2022, doi: 10.1016/j.heliyon.2022.e08788.

[4] M. Aguirre-García, O. Cortés-Zavaleta, H. Ruiz-Espinosa, C. E. Ochoa-Velasco, and I. I. Ruiz-López, "The role of coupled water and solute diffusion and product shrinkage during osmotic dehydration," *J. Food Eng.*, vol. 331, Oct. 2022, doi: 10.1016/j.jfoodeng.2022.111121.

[5] A. Mari, D. N. Parisouli, and M. Krokida, “Exploring Osmotic Dehydration for Food Preservation: Methods, Modelling, and Modern Applications,” Sep. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/foods13172783.

[6] J. R. De Jesus Junqueira, T. F. Rangel, R. P. Campos, T. C. V. Balbinotti, L. Miyagusku, and J. L. G. Corrêa, “Integrated analysis of osmotic dehydration of bocaiuva (Acrocomia aculeata) slices,” *Open Life Sci.*, vol. 21, no. 1, Jan. 2026, doi: 10.1515/biol-2025-1271.

[7] V. Ramya and N. K. Jain, “A Review on Osmotic Dehydration of Fruits and Vegetables: An Integrated Approach,” Jun. 01, 2017, *Blackwell Publishing Inc.* doi: 10.1111/jfpe.12440.

[8] I. Ahmed, I. M. Qazi, and S. Jamal, “Developments in osmotic dehydration technique for the preservation of fruits and vegetables,” Apr. 01, 2016, *Elsevier Ltd*. doi: 10.1016/j.ifset.2016.01.003.

[9] R. Yazidi *et al.*, “Osmotic Dehydration of Orange Fruits in Sucrose and Prickly Pear Molasses Solutions: Mass Transfer and Quality of Dehydrated Products,” *Pol. J. Food Nutr. Sci.*, vol. 74, no. 4, pp. 340–349, 2024, doi: 10.31883/pjfns/194785.

[10] S. Wu, J. Wang, L. Zhang, S. Liu, and C. Li, “Effects of Osmotic Dehydration on Mass Transfer of Tender Coconut Kernel,” *Foods*, vol. 13, no. 14, Jul. 2024, doi: 10.3390/foods13142188.

[11] S. Prajapati, M. Shah, and S. Srivastava, “Mass Transfer Behavior of Jamun (Indian Blackberry) Slices Using Convective Drying Technique,” *Journal of Biosystems Engineering*, vol. 50, no. 3, pp. 353–363, Aug. 2025, doi: 10.1007/s42853-025-00270-3.

[12] J. Floury, A. Le Bail, and Q. T. Pham, “A three-dimensional numerical simulation of the osmotic dehydration of mango and effect of freezing on the mass transfer rates,” *J. Food Eng.*, vol. 85, no. 1, pp. 1–11, Mar. 2008, doi: 10.1016/j.jfoodeng.2007.06.011.

[13] K. G. Angilelli, J. R. Orives, H. C. da Silva, R. L. Coppo, I. Moreira, and D. Borsato, “Multicomponent Diffusion during Osmotic Dehydration Process in Melon Pieces: Influence of Film Coefficient,” *J. Food Process. Preserv.*, vol. 39, no. 4, pp. 329–337, Aug. 2015, doi: 10.1111/jfpp.12236.

[14] J. E. González-Pérez, A. Romo-Hernández, N. Ramírez-Corona, and A. López-Malo, “Modeling mass transfer during osmodehydration of apple cubes with sucrose or apple juice concentrate solutions: Equilibrium estimation, diffusion model, and state observer-based approach,” *J. Food Process Eng.*, vol. 45, no. 10, Oct. 2022, doi: 10.1111/jfpe.14125.

[15] I. Sulistyawati, R. Verkerk, V. Fogliano, and M. Dekker, “Modelling the kinetics of osmotic dehydration of mango: Optimizing process conditions and pre-treatment for health aspects,” *J. Food Eng.*, vol. 280, Sep. 2020, doi: 10.1016/j.jfoodeng.2020.109985.

[16] D. Kaur, M. Singh, R. Zalpouri, and I. Singh, “Osmotic dehydration of fruits using unconventional natural sweeteners and non-thermal-assisted technologies: A review,” Dec. 01, 2022, *John Wiley and Sons Inc*. doi: 10.1111/jfpp.16890.

[17] D. Ghosh and A. Datta, “Deep learning enabled surrogate model of complex food processes for rapid prediction,” *Chem. Eng. Sci.*, vol. 270, p. 118515, Apr. 2023, doi: 10.1016/J.CES.2023.118515.

[18] S. Francik, B. Łapczyńska-Kordon, M. Hajos, G. Basista, A. Zawiślak, and R. Francik, "Modeling the Drying Process of Onion Slices Using Artificial Neural Networks," *Energies* , vol. 17, no. 13, Jul. 2024, doi: 10.3390/en17133199.

[19] C. Tsopwo Zena and Y. Jiokap Nono, "Investigating intermittent immersion during osmotic dehydration of mango (Mangifera indica L. Moench). Part A: Determination of optimal conditions for mango (Mangifera indica L. Moench) dehydration impregnation by immersion (D2I) and intermittent immersion (D3I)," *Heliyon*, vol. 10, no. 16, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35808.

[20] O. Cekmer, S. Um, and M. M. Mench, "A Novel Approach to Structure/Property Relations in Inhomogeneous Porous Media with High and Low Porosity," *Transp. Porous Media*, vol. 152, no. 7, Jul. 2025, doi: 10.1007/s11242-025-02180-6.

[21] L. Arias, Y. Perea, and J. E. Zapata, "Cinética de la Transferencia de Masa en la Deshidratación Osmótica de Mango (Mangifera indica L.) var. Tommy Atkins en Función de la Temperatura," *Informacion Tecnologica*, vol. 28, no. 3, pp. 47–58, Jun. 2017, doi: 10.4067/S0718-07642017000300006.

[22] M. A. M. Khan, L. Ahrné, J. C. Oliveira, and F. A. R. Oliveira, "Prediction of water and soluble solids concentration during osmotic dehydration of mango," *Food and Bioproducts Processing*, vol. 86, no. 1, pp. 7–13, Mar. 2008, doi: 10.1016/J.FBP.2007.10.012.

[23] H. Pacheco-Angulo, E. Herman-Lara, M. A. García-Alvarado, and I. I. Ruiz-López, "Mass transfer modeling in osmotic dehydration: Equilibrium characteristics and process dynamics under variable solution concentration and convective boundary," *Food and Bioproducts Processing*, vol. 97, pp. 88–99, Jan. 2016, doi: 10.1016/J.FBP.2015.11.002.

[24] S. Muñiz-Becerá, L. L. Méndez-Lagunas, and J. Rodríguez-Ramírez, "Solute Transfer in Osmotic Dehydration of Vegetable Foods: A Review," Oct. 01, 2017, *Blackwell Publishing Inc.* doi: 10.1111/1750-3841.13857.

[25] S. Zhelyazkov, S. Aleksandrov, M. Ruskova, T. Petrova, V. Gotcheva, and N. Penov, "Optimization of osmotic dehydration parameters for sweet cherries (Prunus avium) using response surface methodology," *Bulgarian Chemical Communications*, vol. 52, no. 4, pp. 500–505, 2020, doi: 10.34049/bcc.52.4.5304.

[26] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," Jan. 2019, [Online]. Available: http://arxiv.org/abs/1706.09516

[27] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[28] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, Apr. 2001, doi: https://doi.org/10.1023/A:1010933404324.

[29] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems."

[30] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX,

GRAMFORT ET AL. Matthieu Perrot," 2011. [Online]. Available: http://scikit-learn.sourceforge.net.