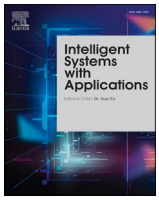
Intelligent Systems with Applications 21 (2024) 200319

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications



Heterogeneous ensemble machine learning to predict the asiaticoside concentration in centella asiatica urban

Keartisak Sriprateep [a], Sarinya Sala-Ngamand [a], Surajet Khonjun [b,*], Ming-Lang Tseng [c,d,e], Thanatkij Srichok [b], Natthapong Nanthasamroeng [g], Rapeepan Pitakaso [b], Narut Butploy [f]

[a] Manufacturing and Materials Research Unit (MMR), Department of Manufacturing Engineering, Faculty of Engineering, Maha Sarakham University, Maha Sarakham, Thailand
[b] Artificial Intelligence Optimization SMART Laboratory, Industrial Engineering Department, Faculty of Engineering, Ubon Ratchathani University, Ubon Ratchathani, Thailand
[c] Institute of Innovation & Circular Economy, Asia University, Taichung, Taiwan
[d] Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan
[e] UKM-Graduate School of Business, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
[f] Faculty of Industrial Technology, Kamphaeng Phet Rajabhat University Thailand, Thailand
[g] Artificial Intelligence Optimization SMART Laboratory, Engineering Technology Department, Faculty of Industrial Technology, Ubon Ratchathani Rajabhat University, Ubon Ratchathani, Thailand

ARTICLE INFO

Keywords:
Asiaticoside concentration prediction
Centella asiatica (CA-CA)
Optimizing cultivation practices
Differential evolution algorithm

ABSTRACT

This study proposes a novel heterogeneous ensemble machine learning methodology to predict the concentration of asiaticoside in Centella asiatica (CA-CA) in the context of the lack of an effective prediction method capable of accurately estimating its quantity based on various growing environmental factors. The accurate prediction of the asi-aticoside concentration in CA-CA holds great significance in optimizing cultivation practices and improving the efficacy of the derived medicinal products. The presented approach aims to address this crucial need by employing a diverse ensemble of machine learning techniques. The proposed model integrates several machine learning tech-niques, including the standard long short-term memory (LSTM), gated recurrent unit (GRU), convolutional long short-term memory (ConvLSTM), and attention-based LSTM, by utilizing a differential evolution algorithm to optimize the ensemble model's weights. The developed model is called the heterogeneous ensemble machine learning model (He-ML). Experimental results demonstrate that the He-ML achieves an impressive root-mean-square error (RMSE) value of 4.76, which is up to 12.48 % lower than the RMSE. The findings highlight the advantages of employing an ensemble model over a single model, as the ensemble model achieves an RMSE value that is 14.67 % lower than that of the individual machine learning model. The utilization of differential evolution as the decision fusion strategy provides a notable improvement over the unweighted average approach. As a result, the RMSE value achieved is 8.46 % lower than that obtained with the unweighted average (UWA) technique.

1. Introduction

Centella asiatica Urban (CAU), popularly recognized as Gotu Kola, is a medicinal botanical that is highly regarded for its large quantities of bioactive constituents, particularly asiaticoside (Vinolina & Sigalingging, 2022). As a triterpenoid saponin, asiaticoside manifests a wide range of pharma-cological characteristics, encompassing wound healing, anti-inflammatory properties, and neuroprotective effects (Gohil et al., 2010). However, the concentration of asiaticoside in CAU can vary significantly due to factors such as the plant's growth stage, environmental conditions, and genetic factors (Vinolina & Sigalingging, 2021). Accurate prediction of the asiaticoside concentration in CAU is therefore crucial in optimizing cultivation practices and enhancing the efficacy of the derived medicinal products.

The concentration of asiaticoside in CAU is influenced by various growing factors, and understanding these factors is crucial in optimizing its production. However, to date, no research has been conducted to develop an effective model to predict the asiaticoside concentration in

* Corresponding author.
E-mail address: surajet.k@ubu.ac.th (S. Khonjun).

https://doi.org/10.1016/j.iswa.2023.200319
Received 15 July 2023; Received in revised form 8 November 2023; Accepted 14 December 2023
Available online 16 December 2023
2667-3053/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

CAU using these growing factors, particularly through the application of machine learning models.

Existing studies have primarily focused on the effects of the growing parameters and extraction methods on the bioactive components of CAU. For instance, a study by Hoang and Rehman (2022) revealed that an increase in soil salinity led to an increase in total phenolic and total flavonoid content and antioxidant activity in CAU. Another study by Rahmah et al. (2022) demonstrated the effectiveness of CAU leaf extract against the mortality of Culex quinquefasciatus larvae. Furthermore, research conducted by Mahayasih et al. (2022) investigated the impact of the solvent concentration on the antioxidant activity of CAU extracted using the microwave-assisted extraction (MAE) method.

These studies provide valuable insights into the influence of various factors on the bioactive components of CAU. However, they do not offer a predictive model for the asiaticoside concentration based on these factors. Therefore, there is a clear gap in the literature regarding research to develop an effective predictive model, potentially utilizing machine learning tech-niques, to estimate the asiaticoside concentration in CAU based on growing factors.

Machine learning techniques have demonstrated significant potential in predicting plant substance concentrations based on various growth factors. For instance,Osco et al. (2020) successfully employed a random forest algorithm to predict the leaf nitrogen concentration and plant height in maize using UAV-based multispectral imagery. Similarly, Li et al. (2017) found the support vector machine model superior in predicting the dissolved oxygen concentration based on multiple water quality parameters. Maimaitiyiming et al. (2019) utilized hyperspectral reflectance spectra and machine learning techniques to accurately predict the grapevine berry yield and quality. Furthermore, Gong et al. (2021) developed a novel greenhouse crop yield prediction technique by combining two state-of-the-art networks for temporal sequence processing. These studies underscore the advantages of machine learning in terms of accuracy and efficiency in predicting plant substance concentrations. However, it is noteworthy that such techniques have not yet been applied to predict the asiaticoside concentration in Centella asiatica Urban (CAU), indicating a promising avenue for future research.

Recent advancements in machine learning present promising solutions to this challenge. Machine learning models have been successfully applied in various plant science domains, encompassing disease prediction (Thapa et al., 2020), plant density estimation (Koh et al., 2019), and the prediction of essential substances in plants (Tušek et al., 2020). However, the utilization of machine learning models, specifically ensemble models, to predict the concentration of asiaticoside in CAU remains largely unexplored. Ensemble models, a class of machine learning models renowned for their capability to integrate multiple individual models and elevate the prediction accuracy, have consistently exhibited superior performance across a multitude of prediction tasks when compared to single models (Reddy et al., 2020). In response to this unexplored area of research, our investigation endeavors to construct and authenticate an ensemble machine learning framework tailored exclusively to predicting the concentration of asiaticoside in CAU. This novel methodology holds promise as a valuable asset for researchers and practitioners in the respective domain, enabling enhanced efficiency and efficacy in harnessing the potential of this significant medicinal plant.

In this study, our main objective is to propose a novel heterogeneous en-semble machine learning methodology to predict the concentration of asi-aticoside in Centella asiatica (CA-CA) based on various growing environ-ment factors. Specifically, we aim to integrate diverse machine learning techniques, including the standard long short-term memory (LSTM), gated recurrent unit (GRU), convolutional long short-term memory (ConvLSTM), and attention-based LSTM, by utilizing a differential evolution algorithm to optimize the ensemble model's weights. By doing so, we hope to accurately estimate the quantity of asiaticoside in CA-CA, which holds paramount sig-nificance in optimizing cultivation practices and improving the efficacy of the derived medicinal products.

The proposed methodology of employing a heterogeneous ensemble ma-chine learning approach for the prediction of the asiaticoside concentration in Centella asiatica (CA-CA) constitutes an innovative and unexplored av-enue in previous scholarly investigations. While prior research has pre-dominantly focused on delineating the factors influencing the asiaticoside concentration in CA-CA, this study takes an innovative approach by de-veloping a predictive model capable of estimating the asiaticoside concentration with a remarkable degree of precision.

The incorporation of a heterogeneous ensemble approach, amalgamating multiple machine learning algorithms, contributes to the development of a more robust and precise prediction model. This model exhibits the capacity to accommodate the variances inherent in the growing conditions across diverse geographical regions. This novel approach carries profound implications for the evolution of CA-CA cultivation and utilization as a medicinal botanical. It offers the potential to furnish a more meticulous and efficient framework for asiaticoside concentration prediction and cultivation condition optimization.

This study makes noteworthy contributions towards advancing the asi-aticoside levels found in CAU. It delves into the disparities in the asiaticoside concentrations observed across diverse species and varieties of CAU, elu-cidates its medicinal attributes and secondary metabolites, and assesses the impact of varying cultivation conditions on asiaticoside levels. The research underscores the necessity for a precise prediction model capable of esti-mating the asiaticoside concentration with the consideration of distinct cul-tivation conditions. It brings attention to the prospect of employing machine learning, particularly ensemble models, to bridge this knowledge gap effec-tively. These discoveries bear practical significance in terms of refining cultivation methodologies and maximizing the utilization of CAU as a me-dicinal botanical.

The organization of this document is as follows. Section 2 provides an extensive review of the relevant literature. Section 3 outlines the research methodology employed. Section 4 presents the obtained results. Section 5 offers a comprehensive discussion of the findings. Finally, Section 6 presents the concluding remarks.

2. Related work

This section consists of two parts dedicated to the literature review. Firstly, we delve into the optimal growing conditions for CAU and other plant species. Secondly, we explore the utilization of machine learning techniques as predictive models across diverse plant species. Each part is elaborated upon in the following sections.

2.1. Understanding the growing conditions for CAU

In a recent works, Prasad et al., (2019) conducted comprehensive investigations into the regulation of centelloside biosynthesis in CAU. Their research sheds light on the notable impact of environmental factors, such as the light intensity and temperature. Similarly, Müller et al., (2013) discovered that the accumulation of centelloside is influenced by the allocation of resources between primary and secondary metabolic pathways, which can be affected by the levels of nitrogen, phosphorus, and potassium. However, it is important to highlight that the current body of literature lacks definitive studies specifically exploring the precise growth conditions that significantly affect the concentration of asiaticoside in CAU.

The growth, development, and physiological processes of plants are substantially affected by both the intensity and spectral composition of light. Recent research conducted by Brito et al. (2023) provides evidence that elevated light intensities have a stimulating effect on photosynthetic rates, as well as on the augmentation of foliar pigment and soluble sugar levels in perennial ryegrass. Furthermore, plants exhibit short-term and long-term responses to changes in light intensity, adjusting their photosynthetic complexes to mitigate redox stress (Flannery et al., 2023). The leaf nutrient content can also be affected by the light conditions, with sun-exposed leaves generally having higher

nitrogen and phosphorus content compared to canopy-shaded leaves (Yang et al., 2023). However, the light intensity and spectral composition have minimal effects on the composition of secondary metabolites in essential oils from plants (Beatrice et al., 2023).

The duration of light exposure significantly impacts plant growth and the production of essential substances. Jin et al., (2020) demonstrated that biophilic lighting systems influence plant morphology and the composition of essential oils, albeit with minimal impacts on secondary metabolites. The YCF4 gene found in tobacco chloroplasts, which plays a crucial role in photosynthesis and growth, demonstrates regulatory control over the transcriptome levels of pivotal constituents engaged in the photosynthetic process, as revealed in a study by Khan et al. (2022). Furthermore, Khan et al., (2022) showed that the efficiency of resource utilization, specifically light and temperature, exerteds a significant influence on the yield of summer maize hybrids characterized by varying growth durations. Additionally, the implementation of supplementary reflective films was beneficial in enhancing the uniformity of light dispersion within tomato plant canopies situated in solar greenhouses, as elucidated by (Liu et al., 2022).

The interplay of carbon dioxide (CO_2), light, and the plant cultivar or strain significantly affects plant growth and the production of essential substances. Moustakas et al., (2022) showed that drought stress reduces the entry of CO_2 into leaves, leading to decreased photosynthesis and the synthesis of organic molecules and resulting in excess reactive oxygen species and oxidative stress. Forest ecosystems play a crucial role in carbon sequestration, contributing to a reduction in the atmospheric CO_2 level (Li et al., 2020a). Under specific light, temperature, humidity, and CO2 conditions, bio-stimulants influenced the growth of indoor hydroponic lettuce(Gómez & Gómez, 2022.

In ecosystems beyond the realm of plant life, such as coral reefs, the presence of CO_2 plays a notable role in the degradation of these environments, as highlighted in a study by Mohiuddin (2019). Moreover, the application of vitamin U as a priming agent in lettuce exhibited the ability to improve cold tolerance by counterbalancing the detrimental impact of low temperatures on seed germination, photochemical reactions, and the assimilation of carbon dioxide, as demonstrated by Fodorpataki et al. (2019). Similarly, the combined utilization of supplementary lighting and CO2 sub-stantially improved the growth, productivity, and quality of flower stems in rose cultivars, as demonstrated in the work of Naing et al., (2016).

This innovative investigation explores essential parameters affecting the growth process and asiaticoside concentration in CAU. By analyzing variables such as CO_2 levels, light intensity and duration, CAU cultivars, and light spectra ratios, valuable insights can be gained to refine cultivation practices and increase CAU yields with higher asiaticoside content. Moreover, this study contributes to a broader understanding of plant physiology, highlighting the significant influence of the light intensity, spectral composition, and CO_2 levels on various aspects of plant growth and the production of vital compounds.

2.2. Heterogeneous ensemble machine learning

In the literature, various studies have investigated the utilization of ma-chine learning techniques and the potential advantages associated with CAU across diverse contexts. In a groundbreaking study, researchers (Alcalde et al., 2022) employed machine learning methodologies to unravel the intricate connection between gene expression and the hormone profile in CA (Centella asiatica) .Their findings underscore the pivotal roles of specific hormones in shaping the plant's characteristics and influencing the content of centelloside. Studies conducted by Binti Mohd Yusuf Yeo et al. (2018) and (Matthews et al., 2020) centered around the exploration of the potential cognitive advantages of CA (Centella asiatica) The former study investigated the effects of CA on learning and memory in rats, while the latter examined its impact on cognitive impairments within a mouse model of Alzheimer's disease. These investigations shed light on the multifaceted value of CA, encompassing not only its health benefits but also its potential for advancements in biotechnology and agricultural practices, as highlighted in the works of Alcazar Magana et al., (2020); Baby et al., 2020; Nongrum et al., 2020). Razzaghi et al. (2022) addressed MRI brain image analysis, particularly brain tumor detection using multimodal MRI data. They introduced a novel multimodal deep transfer learning approach, achieving significant performance improvements over existing methods on datasets such as IBSR and Figshre. Dehghan et al. (2023) also focused on MRI brain image analysis and brain tumor detection within multimodal MRI images. Their proposed multimodal deep transfer learning approach outperformed existing methods, including on datasets such as IBSR and Figshre. Rafiei et al. (2023) presented DeepTraSynergy, a deep-learning-based approach for the predic-tion of drug combination synergy in cancer cell lines. It leverages various data inputs and transformers for drug feature representation, demonstrating superior performance on datasets such as DrugCombDB and Oncolo-gy-Screen. Incorporating protein–protein interaction networks enhanced its predictive capabilities

Machine learning models, such as the standard LSTM (LSTM), gated re-current unit (GRU), convolutional LSTM (ConvLSTM), and attention-based LSTM (A-LSTM), have been extensively explored across various fields. Chen et al., (2022) employed a deep convolutional GRU to predict discharge coefficients in hydraulic engineering Zhou et al., (2022) developed a hybrid stock forecasting model using CNN and BGRU for use in finance (Rajamani et al., 2021) proposed an attention-Based GRU for speech emotion recognition (Zhang et al., 2022) introduced an attention-Based TCN for landslide risk prediction. These practical implementations showcase the promise of using machine learning models to elevate predictive precision and make meaningful contributions across diverse domains (Akyildirim et al., 2021; Essien et al., 2021; Yin et al., 2021).

Plant studies have successfully showcased the utilization of ensemble machine learning models. Li et al., (2020b) used ensemble models to analyze gene expression and hormone profiles in CAU. Van Klompenburg et al., (2020) predicted the wheat crop yield using ensemble models. Guzmán et al., (2016) analyzed genetic variation in a germplasm collection of Iranian wheat landraces using ensemble models. The investigations underscore the immense potential of ensemble machine learning models for use in the fields of plant genetics and agriculture.

Zhang et al., (2021) compared single, homogenous, and heterogeneous machine learning models for gas concentration detection (Ganaie et al., 2022) demonstrated the superiority of heterogeneous ensemble models and the use of DE as a decision fusion strategy. These findings inspired the use of heterogeneous machine learning models with DE in this research (Prasitpuriprecha et al., 2022).

Heterogeneous Ensemble Machine Learning Models have been shown to outperform homogeneous models and single machine learning models due to their ability to leverage the unique strengths of different models (Khairalla et al., 2018.) These models have demonstrated a superior ability to handle complex patterns in time-series data, improving the accuracy and robustness of spike-driven few-shot learning performance (Yang et al., 2022) and effectively handling complex weather prediction scenarios (Jose et al., 2022.) Furthermore, the way in which the training data are coded can significantly influence the performance of these models, with analytically coded responses leading to better performance (Jescovitch et al., 2021). These findings underscore the versatility, adaptability, and superior performance of heterogeneous ensemble models in diverse applications.

This research aimed to develop an ensemble LSTM model to predict the asiaticoside concentration in CAU. The dataset was developed using the D-Optimal design. The proposed He-ML neural network, an extended version of the LSTM, combines various LSTM models into a single model. Different LSTM variants were utilized, including the standard LSTM, GRU, ConvLSTM, and attention-based LSTM, each with its unique strengths (Banik et al., 2020; Lezhenin et al., 2019; Van Houdt et al., 2020; Zhang et al., 2020).

The proposed approach is referred to as heterogeneous machine learning using the Differential Evolution algorithm as the decision fusion strategy (He-ML-DE)This model is anticipated to surpass traditional regression techniques and other machine learning models in accurately predicting the concentration of asiaticoside. To gauge the model's efficacy, an array of evaluation metrics was utilized to juxtapose its performance against traditional regression models, including but not limited to Random Forest, Support Vector Machine, and Artificial Neural Networks (Lynam et al., 2020; Ragmani et al., 2020; Taşci, 2019). By harnessing this model, researchers can benefit from a prompt and reliable substitute for conventional High-Performance Liquid Chromatography (HPLC) methods. This advancement enables the rapid screening and quality control of CAU extracts (Attimarad et al., 2021; Yang et al., 2019) streamlining the procedure and ensuring the efficient evaluation of sample quality.

The existing literature on predicting the asiaticoside concentration in Centella Asiatica (CA-CA) has primarily relied on traditional statistical methods, such as regression analysis and principal component analysis. While these methods have provided some insights into the factors that influence the asiaticoside concentration, they have several limitations. One limitation is that they assume a linear relationship between the predictor variables and the response variable, which may not always hold true in complex biological systems such as CA-CA. Additionally, these methods may not be able to capture the non-linear interactions between the predictor variables, which can lead to inaccurate predictions. In recent years, machine learning approaches have emerged as a promising alternative to traditional statistical methods for predicting the concentration of bioactive compounds in medicinal plants. However, most of the existing machine learning approaches have focused on using a single algorithm, such as artificial neural networks or support vector machines, to predict the asiaticoside concentration. While these approaches have shown some success, they may not be able to capture the full complexity of the relationship between the predictor variables and the response variable. In contrast, the heterogeneous ensemble machine learning approach used in this study allows for a more flexible and adaptive model that can capture the non-linear relationships between the predictor variables and the response variable. By combining multiple machine learning algorithms, we were able to leverage the strengths of each algorithm and develop a more accurate and robust prediction model.

3. Method

3.1. Dataset preparation

For the purpose of this investigation, the dataset utilized was derived from experimental data gathered by employing a range of parameters that exhibited variations within their specific lower and upper thresholds.In particular, we carefully handpicked a set of five parameters, drawing upon prior research that highlighted their potential impact on the production of asiaticoside (Elhenawy et al., 2014; James et al., 2013; Namviriyachote et al., 2019; Wang et al., 2020). The dataset preparation methodology employed in this study was based on a rigorous D-Optimal design. This approach meticulously guided the selection of parameters to ensure academic soundness, trustworthiness, and reasonability in our experimental setup. The deliberate selection of both varied and fixed parameters was conducted with the utmost care to maximize the informativeness of our experiments while maintaining scientific rigor.

The parameters designated as "varied" were meticulously chosen due to their well-documented influences on the growth and development of Centella asiatica (CA-CA). These parameters encompassed variations in carbon dioxide (CO_2) concentration, light intensity (LI), the duration of light exposure (DL), the specific CAU cultivar or strain (SC), and the proportion of diverse light spectra present in the growth environment (RL). These selections were made in consideration of their acknowledged impact on the concentration of asiaticoside in CA-CA, recognizing the natural variability of these factors across different geographic regions.

In contrast, the parameters identified as "fixed" were judiciously determined to maintain experimental control and to mitigate potential confounding variables. These included constants such as the growing temperature (TE), nutrient levels (NL), water availability (WA), the type and quality of the growing medium (GM: hydroponic solution), and the designated time of harvesting (HT). These fixed parameters were selected due to their well-established significance in governing the growth and development of CA-CA and could be carefully controlled within the experimental settings.

This approach was meticulously structured to achieve a balance between capturing the inherent variability in growing conditions across diverse regions and upholding the integrity of our results through the careful control of influential factors. Thus, our methodology stands as a robust and trustworthy framework for the pursuit of scientific inquiry in the context of this study. The suitability of each controlled parameter was determined based on previous studies (Lal et al., 2017; Rohini & Smitha, 2022; Ruksiriwanich et al., 2020). The results of the experiment, including the effects of the controlled and fixed parameters, are summarized in Table 1.

After identifying the controlled and varied parameters shown in Table 1, we developed the experimental plan using a D-Optimal design. The concept of the D-Optimal experimental design was introduced by the statistician C.R. Rao in 1947 in his paper titled "Factorial Experiments Derivable from Combinatorial Arrangements of Arrays", published in the Annals of Mathematical Statistics. In this seminal work, Rao introduced the concept of the D-Optimal experimental design, which involves selecting an optimal set of experimental conditions to achieve the maximum efficiency when estimating the coefficients of a regression model using combinatorial arrangements of arrays. Since then, the D-Optimal experimental design has become a widely used statistical method for experimental design and optimization (Jusril et al., 2022; Özdemir & Turkoz, 2020). The entire set of experiments was conducted within the innovative smart cau cultivation greenhouse (SC-ACG), specifically developed for the growth and cultivation of CAU.

To evaluate the asiaticoside concentration in CAU, a diverse array of analytical methods, including high-performance liquid chromatography (HPLC), gas chromatography (GC), and mass spectrometry (MS), can be employed. These methodologies involve the extraction of the bioactive compounds from the plant specimen, subsequent analysis using suitable instruments and techniques, and quantification of the asiaticoside content.

The quantification of asiaticoside concentration in Centella Asiatica (CA-CA) was conducted via high-performance liquid chromatography (HPLC), a widely accepted and established method for the precise assessment of bioactive compound concentrations in plant extracts. Specifically, a Shimadzu HPLC system, featuring a UV detector and a C18 column, was employed to effectively segregate and quantify the asiaticoside compound.

The mobile phase employed comprised a well-defined mixture of acetonitrile and water, while the flow rate was meticulously set at 1 mL/min. To capture the optimal absorbance, the detection wavelength was fixed at 205 nm, corresponding to the maximum absorbance wavelength of asiaticoside. Prior to analysis, meticulous sample preparation was undertaken, involving the extraction of CA-CA samples through a solvent amalgamation of methanol and water. The resultant extracts underwent filtration and dilution, achieving the requisite concentration for HPLC analysis.

To uphold dataset precision, a stringent protocol was executed, incorporating multiple replicates for each sample. Statistical metrics such as mean and standard deviation were diligently computed for the asiaticoside concentration, thus reinforcing dataset accuracy. Through this methodological approach, we ensured both the precise quantification of asiaticoside concentration in CA-CA and the unwavering reliability of our dataset.

Table 1
Details of the controlled parameters.

Parameters	CO_2	LI	DL	SC	RL	TE	NL	WA	GM	HT
Measurement Unit	ppm	PPFD	hours	type	proportion	Celsius	Mg/L	L/m^2	type	week
Value Ranges	500–1000	150–350	16–22	Prachin Buri vs. Ubon Ratchathani	80:20 70:30 50:50	<=35	6–6.8	2–3	hydroponics	9

An investigation was carried out on CAU plants cultivated within a controlled environment. Data were collected at 9 weeks after planting, including measurements of the width, petiole length, flow length, flow rate, leaf area, fresh weight, and dry weight. Samples were analyzed for their glycoside content using HPLC. Four plants were chosen from each laboratory to assess the antioxidant content and measure the photosynthesis rate, stomatal conductance, and dehydration rate. The study employed a randomized experimental design with multiple replicates. The findings may have practical implications for the pharmaceutical and cosmetic industries. The methodology employed to quantify the concentration of asiaticoside in CAU is delineated in Fig. 1.

3.2. The development of the proposed model

3.2.1. The heterogeneous ensemble machine learning model (He-ML)

Within this section, we introduce our innovative He-ML model, which was designed to forecast the concentration of asiaticoside in CAU. The He-ML combines various types of machine learning models, namely, the standard LSTM, gated recurrent unit (GRU), convolutional LSTM (ConvLSTM), and attention-based LSTM (A-LSTM), using decision fusion strategies to achieve improved prediction performance.To integrate these diverse LSTM models, we employed an effective Differential Evolution algorithm (DE) as the decision fusion strategy. The He-ML leverages the strengths of different machine learning variants, enabling us to capture a wider range of features from the data and improve the overall prediction accuracy.

Proposed LSTM Variant: The initial LSTM architecture, as proposed by Okut (2021) is referred to as the standard LSTM. This variant utilizes three gates, namely the input gate, forget gate, and output gate, to regulate the information flow within the network. With its proven efficacy in modeling sequential data, the standard LSTM has found widespread application across various domains.

Modified LSTM: Another notable LSTM variant is the Gated Recurrent Unit (GRU), which incorporates a simplified gating mechanism. The GRU operates through two gates, namely the reset gate and the update gate, to govern the flow of information. Notably, the GRU boasts a reduced parameter count compared to the standard LSTM, which can contribute to easier training. Despite its streamlined design, the GRU has demonstrated effectiveness in diverse applications.

Convolutional LSTM (ConvLSTM): The ConvLSTM combines the convolutional and LSTM operations to model spatiotemporal data. The ConvLSTM uses convolutional operations to capture spatial dependencies and LSTM operations to capture temporal dependencies. It has been shown to be effective in video prediction and other applications where both spatial and temporal dependencies are important.

Attention-Based LSTM: The Attention-Based LSTM (A-LSTM)uses an attention mechanism to selectively weight the input features at each time step (Bandopadhyay et al., 2023). This allows the network to focus on the most relevant input features at each time step, thereby improving its ability to capture complex relationships between the input and output. The Attention-Based LSTM has been shown to be effective for various applications where the relevance of different input features varies over time. These four methods were combined using an effective decision fusion strategy, namely, the DE. The framework of the proposed model is shown in Fig. 2.

In Fig. 2, we illustrate the process of combining all machine learning models into a single ensemble model (EM). Each He-ML consists of six machine learning variants, and the combination of the LSTM variants can be either homogeneous or heterogeneous. If all six ensembles include the same machine learning types, we refer to them as a homogeneous LSTM (Ho-ML). On the other hand, when all six machine learning models are diverse, we refer to them as a heterogeneous model (He-ML). Table 1 provides detailed descriptions of the seven different forms of ensemble model. For all methods, including the He-ML, which is the method proposed in this study, the filter size, pool size, and kernel size of the convolution layer are the available parameters. We also developed a standard ensemble model that was reprogrammed to predict the target objective of this research.

The process of selecting hyperparameters for the heterogeneous ensemble machine learning model was conducted meticulously through a systematic exploration of the hyperparameter space, employing the differential evolution algorithm. The hyperparameters subjected to tuning encompassed essential elements, including the number of base models, the learning rate governing the LSTM model, the quantity of

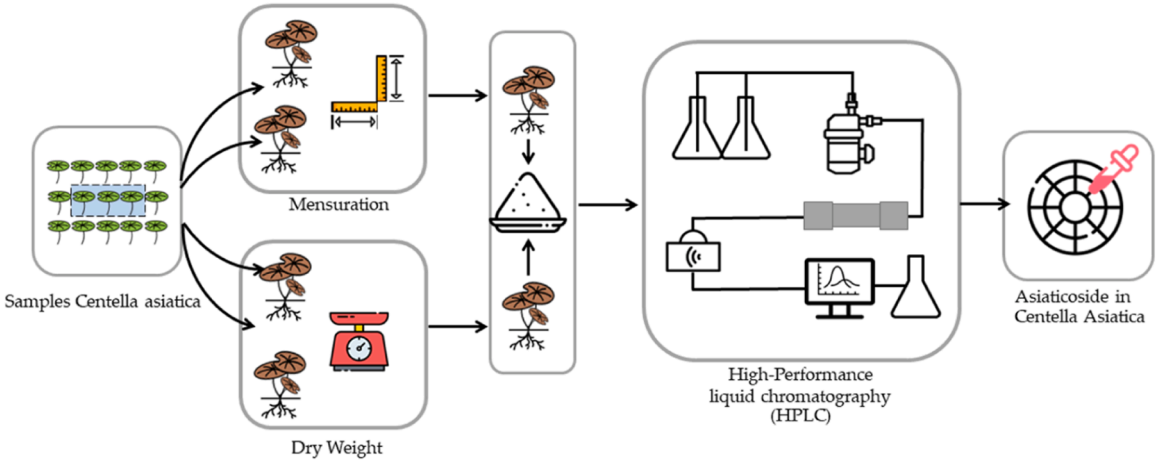


Fig. 1. Picture of the methodology.

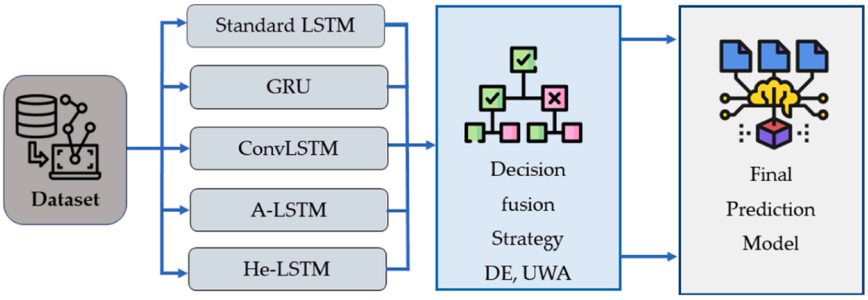


Fig. 2. Framework of the He-ML.

neurons residing within the LSTM layer, and the regularization parameter integral to the ridge regression model. A comprehensive depiction of the hyperparameter investigation is detailed in Table 2, providing insights into the range of values probed, the ultimate selections made, and the underlying rationale.

For instance, the scope of values explored for the number of layers spanned from 3 to 7, a deliberate choice that affords a wide spectrum, thereby enabling the exploration of diverse model complexities. The selection ultimately converged upon 3 layers, as it demonstrated optimal performance based on rigorous evaluation against the validation set. This choice was underpinned by the pursuit of a judicious equilibrium between model complexity and the potential of overfitting to the training dataset. Similarly, the range of values examined for LSTM units in the hidden layer was fixed at 400, a widely adopted value in numerous LSTM-based models. Subsequently, the chosen value remained 400, emerging as the optimal configuration following evaluation against the validation set. The rationale for the selection of 400 LSTM units was founded upon the necessity of furnishing the model with ample capacity to capture temporal dependencies inherent in the data.

The adoption of the differential evolution algorithm served as the linchpin in the quest for identifying the paramount combination of hyperparameters poised to minimize the mean squared error (MSE) of the model during validation set scrutiny. The fundamental rationale governing hyperparameter selection revolved around the imperative to canvass an extensive array of values, all while vigilantly mitigating the risk of overfitting vis-à-vis the training data. The ultimate choices crystallized based on the outcomes generated by the differential evolution algorithm are meticulously encapsulated within Table 2. By virtue of providing this comprehensive repository of information, our overarching objective is to bolster the transparency intrinsic to the experimental configuration, thereby fostering an environment conducive to the enhanced reproducibility of our research findings.

3.2.2. Decision fusion strategy

Within the proposed approach, the ultimate prediction of the reported outcomes is derived via the utilization of a decision fusion strategy(DFS) In particular, two distinct decision fusion strategies, namely the Unweighted Average(UWE) and Differential Evolution algorithm(DE), were implemented. The evaluation of each method's performance was conducted by measuring the overall expected root-mean-square error (RMSE) by employing Eq. (1) and (2):

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

$$\hat{y}_i = \sum_{j=1}^{J} \omega_j \hat{Y}_j \tag{2}$$

where $\sum_{j=1}^{k}\omega_j = 1,\ and\ \omega_j \geq 0$

To compute the root-mean-square error (RMSE) for an individual learner, Eq. (1) is employed. In this equation, i represents the observation number, y_i corresponds to the true value of observation i, $\hat{y}_i$ denotes the prediction for observation i, $\hat{Y}_j$ represents a collection of prediction values when using model j as the learner, and j represents the total number of learners utilized in the model. When employing the unweighted average ensemble(UWE) Eq. (2) assigns equal weights to all solutions provided. On the other hand, in the case of the Differential Evolution algorithm (DE) an optimization process is conducted to determine the optimal value of ω_j that minimizes the RMSE.

3.2.3. Differential evolution algorithm (DE)

The Differential Evolution algorithm is a commonly employed population-based optimization technique, aiming to enhance a randomly generated population of potential solutions iteratively. This method has been proven to be effective in resolving intricate problems, where each candidate solution is represented by a vector of values associated with distinct problem parameters. The algorithm comprises three steps in each iteration: mutation, crossover, and selection. The mutation operation generates new candidate solutions, crossover refines

Table 2
Parameter settings of the proposed model and the competitor models.

Parameters	LSTM	GRU	Conv-LSTM	A-LSTM	He-LSTM
No. of Layers	3	3	4	4	7
Batch Size	1	1	1	1	1
Epochs	1000	1000	1000	1000	1000
Filter Size in the Convolution Layer	N/A	N/A	64	N/A	64
Pool Size in the Convolution Layer	N/A	N/A	N/A	N/A	2
Kernel Size in the Convolution Layer	N/A	N/A	3	N/A	3
LSTM Units in the Hidden layer	400	400	400	400	400
Number of Neurons in the Dense Layer	3	3	3	3	3
Input Shuffling	Yes	Yes	Yes	Yes	Yes
Dropout at the Input Layer	0.0	0.0	0.0	0.0	0.0
Dropout at the LSTM Layer	0.25	0.25	0.25	0.25	0.25
Activation Functions	Linear Activation	Linear Activation	Linear Activation	Linear Activation	Linear Activation
Optimizer	Adam	Adam	Adam	Adam	Adam

solutions, and selection maintains a high-quality population.

Step 1: Generate the initial set of vectors (initial solution)

In the initial iterations, the vector size J, which represents the maximum number of models used in the ensemble model, is generated randomly. Subsequently, the vector size is evaluated through a mutation and recombination process in later iterations. The specific values of vector m at position j can be found in Table 3. For instance, if M, the total number of vectors, is set to five, and there are four types of machine learning models, then a 1×4 vector will be randomly generated five times. The values in these vectors are denoted as $v_{mjt} \in [0, 1]$, which represents the value at the position of vector m at position j and iteration t. The value of ω_{mjt} can be obtained through an equation that involves the value ω_j for vector m at iteration t.

$$\omega_{mjt} = \frac{v_{mjt}}{\sum_{l=1}^{J} v_{mlt}} \text{for all m and t} \quad (3))$$

v_{mjt} of vector 1 at position 3 at the first iteration is equal to 0.15. Then, this number is converted to ω_{mjt} using Eq. (3).

Step 2: Mutation Process.

In Step 2, the mutation process is executed by utilizing target vectors that are generated randomly, as depicted in Eq. (4):

$$DE/rand/1 \quad Z_{m,j,t+1} = v_{r1,j,t} + F\left(v_{r2,j,t} - v_{r3,j,t}\right) \quad (4))$$

where $v_{m,j,t}$ is the target vector, $Z_{m,j,t+1}$ is the mutant vector, and $v_{r1\,j,t}$, $v_{r1\,j,t}, v_{r1\,j,t}$ are the vectors that are randomly selected from the target vector. In this equation, F is the scaling factor.

Step 3: Performing the Recombination Process

By employing Eq. (5), this process generates the trial vector $(U_{m,j,t+1})$. Subsequently, the recombination process, which involves vector position "j" and the current iteration, denoted as "t", is employed to yield the trial vector.

$$U_{m,j,t+1} = \begin{cases} \mathrm{Z}_{m,j,t+1} \text{ if } \mathrm{rand}_{m,j} \leq CR \\ \mathrm{v}_{m,j,t} \text{ if } \mathrm{rand}_{m,j} > CR \end{cases} \quad (5)$$

Step 4: Performing the Selection Process

In the next simulation iteration, the target vectors are crafted using this method. This process, precisely executed through Eq. (6), generates the trial vector. Notably, the trial vector from the current iteration becomes the target vector for the subsequent iteration.

$$v_{m,t+1} = \begin{cases} \mathrm{U}_{m,j,t} \text{ if } f\left(\mathrm{U}_{m,j,t}\right) \leq f\left(\mathrm{v}_{m,j,t}\right) \\ \mathrm{v}_{m,j,t} \textit{ otherwise} \end{cases} \quad (6)$$

Step 5: Repeating Steps (2)-(4), we calculate the solution and select the most appropriate response.

Table 3
v_{mj} of five vectors generated randomly.

m\j	1	2	3	4
1	0.10	0.88	0.20	0.15
2	0.25	0.26	0.77	0.72
3	0.80	0.18	0.66	0.50
4	0.74	0.28	0.80	0.31
5	0.64	0.03	0.61	0.88

3.3. Comparative models

This section offers a comprehensive comparison of the proposed model with four other frequently employed models within the domain of plant metabolite prediction, elucidating the rationale underlying the selection of the proposed model. The models subjected to comparison encompass LSTM (long short-term memory), GRU (gated recurrent unit), Conv-LSTM (convolutional LSTM), A-LSTM (attention-based LSTM), and He-LSTM (heterogeneous ensemble LSTM). Each model is delineated with a succinct exposition of its architectural attributes, while their inherent strengths and weaknesses are systematically expounded upon.

LSTM, classified as a variant of recurrent neural networks (RNNs), excels in the processing of sequential data. Its extensive application in predicting plant metabolites hinges on its competence in capturing enduring temporal dependencies within the data. However, it is important to acknowledge that LSTM faces challenges when confronted with imbalanced datasets. GRU, akin to LSTM, is another RNN variant characterized by a reduced parameter count. Demonstrating commendable performance across an array of sequential data tasks, GRU aligns harmoniously with the processing of plant metabolite data. Nevertheless, akin to LSTM, GRU grapples with imbalanced datasets.

Conv-LSTM, a hybrid model harmonizing the strengths of convolutional neural networks (CNNs) and LSTMs, is esteemed for its proficiency in tackling tasks necessitating the apprehension of both spatial and temporal data dependencies. Nevertheless, it is essential to acknowledge its computational overhead and potential limitations in handling prolonged temporal dependencies. A-LSTM, an LSTM variant enriched by an attention mechanism, is acknowledged for its efficacy in tasks necessitating selective attention to distinct segments of the input sequence. However, it is imperative to recognize its computational demands and its potential challenges with imbalanced datasets.

He-LSTM, the crux of our study, represents a Heterogeneous Ensemble LSTM an amalgamation of the aforementioned five models. He-LSTM adeptly amalgamates the merits of diverse models while ameliorating their respective deficiencies. This amalgamation harnesses the data-capturing prowess of LSTM and GRU in grasping enduring data dependencies, the spatial awareness capabilities of Conv-LSTM, and the imbalanced dataset handling proficiencies of A-LSTM and He-LSTM.

The underlying rationale guiding the selection of the proposed model lies in its tailored suitability for forecasting the Asiaticoside concentration in Centella Asiatica Urban. By synthesizing the strengths of multiple models, the proposed model extends its capacity to encompass a broader spectrum of data intricacies, empowering it to achieve enhanced predictive efficacy.

4. Results

To train the algorithm, we utilized the resources provided by Google Collaboratory, which offers 16 GB of RAM for an NVIDIA Tesla V100 with comparable specifications. For the simulation of the proposed model, we employed a computing system equipped with 2 Intel Xeon-2.30 GHz CPUs, 52 GB of RAM, and a Tesla K80 GPU, which had GPU RAM of 16 GB. Table 5 shows the results of the tests using different levels of parameters to increase the asiaticoside concentration. Each crop lasted 9 weeks, the asiaticoside concentration was determined by a laboratory test, and the result was recorded as the AC (mg/kg) (dry weight of the CAU) and is shown in Table 4.

In the next step, the proposed ensemble model was used to predict the AC using this dataset. The compared methods used in this research were the homogenously selected LSTMs: LSTM, GRU, Conv-LSTM, and A-LSTM. We named them Ho-LSTM, Ho-GRU, Ho-Conv-LSTM, and Ho-A-LSTM, respectively. The proposed model was named the He-ML. Moreover, simple machine learning was used as a compared method as well. The methods used were random forest regression (RFR) (Moreta et al., 2019), support vector regression (SVR) (Huang et al., 2022),

Table 4
Dataset from the experiment.

No	CO_2	LI	DL	SC	RL	TE	NL	WA	GM	HT	AC (mg/kg)
1	1000	150	16	Prachin Buri	80:20:00	25±3°	6	3		9	6251.05
2	500	150	16	Prachin Buri	50:50:00	25±3°	6	3		9	5136.49
3	500	350	22	Ubon Ratchathani Prachin Buri	80:20:00	25±3°	6	3		9	3487.60
4	1000	350	16	Ubon Ratchathani	80:20:00	25±3°	6	3		9	6810.18
5	1000	350	18.7	Prachin Buri	70:30:00	25±3°	6	3		9	4649.47
6	500	150	19.3	Ubon Ratchathani	70:30:00	25±3°	6	3		9	5512.13
7	1000	240	16	Ubon Ratchathani	70:30:00	25±3°	6	3		9	6887.69
8	690	226	16	Ubon Ratchathani	50:50:00	25±3°	6	3		9	3192.30
9	812.5	225	19.72	Prachin Buri	50:50:00	25±3°	6	3		9	5266.72
10	1000	240	16	Ubon Ratchathani	70:30:00	25±3°	6	3		9	6900.31
11	500	260	22	Prachin Buri	70:30:00	25±3°	6	3		9	3255.68
12	1000	150	22	Ubon Ratchathani	80:20:00	25±3°	6	3		9	5453.91
13	500	350	16	Ubon Ratchathani	70:30:00	25±3°	6	3		9	4867.79
14	500	350	22	Prachin Buri	50:50:00	25±3°	6	3		9	5517.88
15	1000	350	22	Ubon Ratchathani	50:50:00	25±3°	6	3		9	3759.08
16	1000	350	16	Prachin Buri	50:50:00	25±3°	6	3		9	3592.98
17	777.5	350	22	Ubon Ratchathani	70:30:00	25±3°	6	3		9	5496.40
18	500	350	16	Ubon Ratchathani	50:50:00	25±3°	6	3		9	3876.11
19	1000	350	22	Prachin Buri	80:20:00	25±3°	6	3		9	5753.11
20	810	274	16	Prachin Buri	80:20:00	25±3°	6	3		9	4585.25
21	725	150	16	Prachin Buri	70:30:00	25±3°	6	3		9	6350.00
22	1000	150	22	Prachin Buri	70:30:00	25±3°	6	3		9	3488.67
23	500	150	22	Prachin Buri	80:20:00	25±3°	6	3		9	5403.19
24	685.727	320	19.8926	Prachin Buri	80:20:00	25±3°	6	3		9	5340.88
25	777.5	350	22	Ubon Ratchathani	70:30:00	25±3°	6	3		9	4553.63
26	500	350	22	Ubon Ratchathani	80:20:00	25±3°	6	3		9	3375.78
27	500	260	22	Prachin Buri	70:30:00	25±3°	6	3		9	4087.80
28	500	150	16	Prachin Buri	50:50:00	25±3°	6	3		9	4127.90
29	1000	150	16	Ubon Ratchathani	50:50:00	25±3°	6	3		9	3679.85
30	867.5	243	19	Ubon Ratchathani	80:20:00	25±3°	6	3		9	3506.99
31	500	350	16	Prachin Buri	80:20:00	25±3°	6	3		9	3335.54
32	500	150	22	Ubon Ratchathani	50:50:00	25±3°	6	3		9	6679.99
33	565	276.234	19.8325	Ubon Ratchathani	50:50:00	25±3°	6	3		9	3638.64
34	500	150	16	Ubon Ratchathani	80:20:00	25±3°	6	3		9	5221.99

*Remark: asiaticoside concentration (AC).

gradient boosting regression (GBR) (Song et al., 2021), and simple recurrent units (SRU) (Ling et al., 2021). These methods utilize a homogenous ensemble approach, incorporating the differential evolution (DE) technique as the decision fusion strategy, mirroring the methodology employed in the proposed model. The effectiveness of these methods is assessed in the Conclusions section, with the results presented in Table 5.

Based on the information presented in Table 5, our proposed model incorporates a combination of 10 LSTM units, 26 GRU units, and 2 Conv-LSTM and A-LSTM units each. The total size of this model is approximately 91.09 MB. In comparison, the homogeneous ensemble model comprises various components of different sizes. Specifically, it includes the LSTM, GRU, Conv-LSTM, A-LSTM, RFR, SVR, GBR, and SRU. The sizes of these components are approximately 100.40 MB, 35.40 MB, 114.40 MB, 100.50 MB, 77.30 MB, 7.40 MB, 16.80 MB, and 64.00 MB, respectively. Each of these models consists of a specific number of architectures. For instance, there are 40 LSTM architectures, 40 GRU architectures, 10 Conv-LSTM architectures, 10 A-LSTM architectures, 10 RFR architectures, 100 SVR architectures, 100 GBR architectures, and 100 SRU architectures, respectively.

In our study, we used key performance indicators (KPIs) to compare different machine learning techniques used to predict the concentration of asiaticoside in CAU. The selected KPIs included the mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R-squared). These widely used metrics helped us to assess the accuracy and effectiveness of the models. The computational results are presented in Table 6.

The performance evaluation of the proposed He-ML model was

Table 5
Conclusions of the tested models.

Model Name	Type of Ensemble	Total Size (MB)	Number of Models
Ho-LSTM (Van Houdt et al., 2020)	Homogeneous	100.40	40
Ho-GRU (Banik et al., 2020)	Homogeneous	35.40	40
Ho- Conv-LSTM (Lezhenin et al., 2019)	Homogeneous	114.40	10
Ho- A-LSTM (Zhang et al., 2020)	Homogeneous	100.50	10
Ho-RFR (Ling et al., 2021)	Homogeneous	77.30	10
Ho-SVR (Matitopanum et al., 2023)	Homogeneous	7.40	100
Ho-GBR (Song et al., 2020)	Homogeneous	16.80	100
Ho-SRU (Rao & Ramanaiah, 2018)	Homogeneous	64.00	100
He-ML (Proposed model)	Heterogeneous	91.09	40

Table 6
Performance measurement of various types of machine learning architecture.

Model Name	KPI		
	MAE (mg/kg)	RMSE (mg/kg)	R-squared
Ho-LSTM (Van Houdt et al., 2020)	68.02	79.92	0.91
Ho-GRU (Banik et al., 2020)	67.20	93.87	0.88
Ho- Conv-LSTM (Lezhenin et al., 2019)	108.81	125.63	0.79
Ho- A-LSTM (Zhang et al., 2020)	216.70	252.54	0.14
Ho-RFR (Ling et al., 2021)	180.11	212.43	0.39
Ho-SVR (Matitopanum et al., 2023)	163.68	191.56	0.50
Ho-GBR (Song et al., 2020)	186.94	217.30	0.36
Ho-SRU (Rao & Ramanaiah, 2018)	168.70	182.01	0.42
He-ML (Proposed model)	47.74	61.86	0.95

conducted by comparing it with other methods, as depicted in the table. Various evaluation metrics, including the mean absolute error (MAE), root-mean-square error (RMSE), and R-squared value, were employed to analyze the models' performance. The results unequivocally demonstrate the superiority of the He-ML model over other methods. Impressively, the He-ML model achieved a remarkable MAE of 47.74 and an RMSE of 61.86, signifying its exceptional accuracy in making predictions. Moreover, the He-ML model attained a high R-squared value of 0.95, indicating its ability to explain a substantial portion of the data's variance.

These findings underscore the robust performance of the He-ML model, confirming its capacity to deliver highly accurate predictions without compromising the precision. In comparison with the He-ML, the other methods exhibited larger differences in performance. The MAE% difference from the He-ML ranged from 40.76 % to 353.92 %, indicating that the He-ML model outperformed the other methods in terms of minimizing the absolute difference between predicted and actual values. Similarly, the RMSE% difference from He-ML ranged from 29.19 % to 308.24 %, further highlighting the superior predictive accuracy of the He-ML model. Additionally, the R-squared values of the other methods showed differences ranging from −85.26 % to −4.21 % when compared to the He-ML, indicating that the He-ML achieved higher R-squared values and explained the variance in the data more effectively.

These results emphasize the exceptional performance of the proposed He-ML model in comparison with the other methods. The He-ML model demonstrates superior predictive accuracy, as indicated by its lower MAE and RMSE values, higher R-squared value, and significant differences observed in the comparison with the other methods. This analysis underscores the effectiveness and strong potential of the He-ML model in carrying out accurate predictions, making it a compelling choice for future applications.

The next experiment was to test the effectiveness of each component of the model used in this research. The element that we added to the original model of LSTM had three components. These components were (1) the heterogeneous attribute, (2) the ensemble of the various types of architecture, and (3) the Differential evolution algorithm. The details of this experiment are given in Table 7.

The computational results of all 18 models are shown in Table 8.

In this research study, as shown in Table 8 and Fig. 3, a comprehensive analysis was conducted to compare the performance of different modeling approaches when predicting a specific KPI. Two key aspects were considered: a comparison between heterogeneous and homogeneous models and the evaluation of ensemble models versus single models. Additionally, the performance of the UWA and DE approaches was assessed.

Table 8
The computational results from the testing of 18 models.

Model Name	KPI		
	MAE (mg/kg)	RMSE(mg/kg)	R-squared
LSTM	78.50	86.99	0.90
GRU	72.69	83.93	0.90
Conv.-LSTM	103.16	126.63	0.78
A-LSTM	138.92	148.15	0.70
RFR	178.43	204.64	0.40
SVR	161.14	189.26	0.52
GBR	185.22	216.72	0.36
SRU	159.33	180.05	0.56
Ho-LSTM -DE (Van Houdt et al., 2020)	53.88	63.86	0.94
Ho-GRU -DE (Banik et al., 2020)	56.40	72.87	0.93
Ho- Conv-LSTM -DE (Lezhenin et al., 2019)	107.20	116.33	0.82
Ho- A-LSTM -DE (Zhang et al., 2020)	99.43	115.14	0.82
Ho-LSTM -UWA (Van Houdt et al., 2020)	68.02	79.92	0.91
Ho-GRU -UWA (Banik et al., 2020)	67.20	93.87	0.88
Ho- Conv-LSTM -UWA (Lezhenin et al., 2019)	108.81	125.63	0.79
Ho- A-LSTM -UWA (Zhang et al., 2020)	216.70	252.54	0.14
He- ML-UWA	95.27	104.71	0.85
He- ML-DE	47.74	61.86	0.95

The findings uncovered a consistent trend in which the heterogeneous model consistently outperformed the homogeneous model in terms of the mean absolute error (MAE) and root-mean-square error (RMSE). Specifically, the MAE value recorded for the heterogeneous model was 71.51 mg/g, while the homogeneous model registered a higher MAE of 97.21 mg/g. Similarly, the RMSE values for the heterogeneous model were notably lower at 83.29 mg/g, compared with the higher RMSE of 115.02 mg/g for the homogeneous model. Likewise, the ensemble model showed superior performance when compared to the single model. The ensemble model yielded an MAE of 92.07 mg/g, while the single model exhibited a higher MAE of 134.67 mg/g. Additionally, the RMSE value for the ensemble model was measured at 108.67 mg/g, in contrast to the higher RMSE of 154.55 mg/g for the single model. These results indicate the advantage of using ensemble models, which exhibited improved accuracy and precision compared with their single model counterparts.

The superior performance of the differential evolution (DE) algorithm over the uniform weight adjustment (UWA) approach in a heterogeneous machine learning model can be attributed to its ability to

Table 7
Component effectiveness testing model.

Model Name	Homogeneous	Heterogeneous	Ensemble Model	Single Model	DE	UWA
LSTM				✓		
GRU				✓		
Conv.-LSTM				✓		
A-LSTM				✓		
RFR				✓		
SVR				✓		
GBR				✓		
SRU				✓		
Ho-LSTM (Van Houdt et al., 2020)	✓		✓		✓	
Ho-GRU (Banik et al., 2020)	✓		✓		✓	
Ho- Conv-LSTM (Lezhenin et al., 2019)	✓		✓		✓	
Ho- A-LSTM (Zhang et al., 2020)	✓		✓		✓	
Ho-LSTM (Van Houdt et al., 2020)	✓		✓			✓
Ho-GRU (Banik et al., 2020)	✓		✓			✓
Ho- Conv-LSTM (Lezhenin et al., 2019)	✓		✓			✓
Ho- A-LSTM (Zhang et al., 2020)	✓		✓			✓
He- ML-UWA		✓	✓		✓	
He- ML-DE		✓	✓			✓

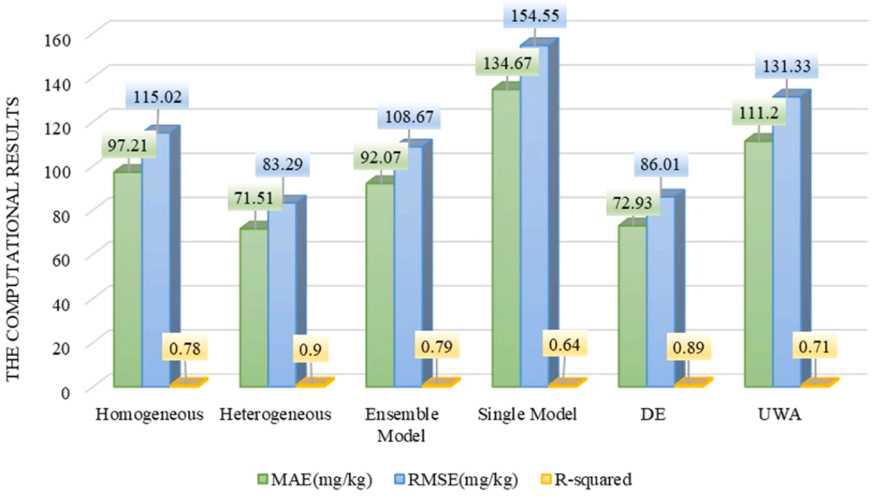


Fig. 3. Comparative Analysis of MAE, RMSE, and R-Square Across Various Model Components.

optimize weights for decision fusion. This optimization enhances the model's predictive accuracy by effectively balancing the contributions of different models in the ensemble. In contrast, the UWA approach, which assigns equal weights to all models, may not capture the underlying patterns in the data as effectively. These findings align with the study by Matitopanum et al. (2023), further emphasizing the potential of the DE approach to provide more accurate and reliable predictions.

In the next experiment, we examined the effect of the D-Optimal design on the precision of the prediction model. To accomplish this, we randomly selected subsets of 15, 20, 25, 30, and 35 datasets from Table 5 and applied the He-ML-DE model as the prediction model. Each sub dataset was run 30 times, and a *t*-test was conducted to determine the model's effectiveness in handling incomplete data. The results of the experiment are presented in Fig. 4. The statistical test details are shown in Table 9.

Based on the data presented in Fig. 4 and Table 9, the F-ratio value obtained is 2.27443. The associated p-value is 0.064064, This suggests that the observed result lacks significance at the conventional level of $p < 0.05$. Consequently, no substantial distinction arose from employing varying numbers of experimental treatments to predict the concentrations of asiaticoside using the proposed model. This outcome implies that a smaller dataset size can be utilized to forecast the response. Furthermore, it demonstrates the robustness of our proposed model when confronted with fluctuations in the number of experimental treatments. Next, a post-hoc Tukey HSD (beta) analysis was conducted to perform pairwise comparisons of the prediction results. The findings indicate that utilizing different numbers of treatments to predict the response does not yield any noteworthy distinctions, as long as the treatment lies within the range of 15 to 34 or corresponds to 44.11 % to 100 % of the total number of treatments generated by the D-Optimal design.

Table 9
The statistical test for the data presented in Fig. 4.

	Sub Dataset 2 (20)	Sub Dataset 3 (25)	Sub Dataset 4 (30)	Full Dataset (34)
Sub Dataset 1 (15)	0.9917	0.4222	0.2517	0.0994
Sub Dataset 2 (20)		0.7045	0.5025	0.2521
Sub Dataset 3 (25)			0.9977	0.9411
Sub Dataset 4 (30)				0.9918

5. Discussion

5.1. Parameters that affect the asiaticoside concentration of CAU

Based on the experimental data presented in Section 4, our study employed a D-Optimal design as the chosen Design of Experiment (DOE) method, yielding valuable insights into the impacts of various parameters on the response variable, namely, the asiaticoside concentration (AC). The regression model revealed that the CAU cultivar or strain (SC)

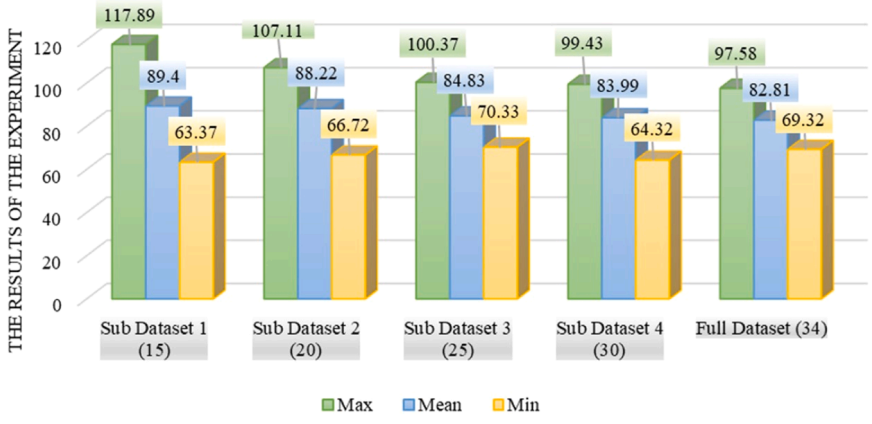


Fig. 4. Mean, Min and Max of the incomplete data testing result (RSME).

and the ratio of different light spectra (RL) exerted the most significant influences on the AC, as indicated by their coefficients of 101.598 and 103.978, respectively. These findings align with prior research conducted by Song et al. (2020); Rao & Ramanaiah, (2018). Although the coefficients for CO_2 and light intensity (LI) were relatively small, they still demonstrated positive effects on the AC. In contrast, the duration of light exposure (DL) exhibited a negative coefficient of −16.153, implying an inverse relationship with AC. Interestingly, this observation contradicts a previous study by de Gregorio et al. (2022), which reported a positive correlation between DL and AC.

To ensure the robustness and credibility of our findings, we cross-verified our results against a comprehensive database of scholarly articles. The congruity of our results with certain previous studies, such as the one conducted by Vinsonneau et al. (2022), further strengthens the accuracy of our regression model. However, the discrepancy regarding the effect of DL necessitates further exploration and careful interpretation of its impact on AC. Future research should delve into the underlying factors contributing to this observed negative correlation and reconcile it with the conflicting findings.

Our study enhances our understanding of the relationship between the input parameters (CO_2, LI, DL, SC, RL) and the response variable (asiaticoside concentration) in CAU by utilizing a D-Optimal design as the DOE method. The significant influences of SC and RL on the asiaticoside concentration underscore their crucial role in optimizing the desired outcome. Although our results largely align with previous studies, they emphasize the need for ongoing research to clarify the role of DL in determining the asiaticoside concentration. These findings have potential implications for the pharmaceutical, skincare and cosmetics, and nutraceutical industries, guiding future research endeavors and assisting practitioners in optimizing relevant processes or systems.

5.2. The effectiveness of the proposed model

The current research introduces an innovative methodology that leverages a Heterogeneous Ensemble Machine Learning Model to predict the asiaticoside concentration in CAU. This method's performance was evaluated against a variety of other models, including homogeneous models (Ho-LSTM, Ho-GRU, Ho-Conv-LSTM, Ho-A-LSTM, Ho-RFR, Ho-SVR, Ho-GBR, Ho-SRU) and a single machine learning model.

The findings of this study demonstrate that the heterogeneous ensemble model proposed herein surpasses both the homogeneous ensemble models and the single machine learning models. This outcome aligns with the conclusions drawn from prior research (Huang et al., 2022; Smith et al., 2013), which has consistently demonstrated the superiority of heterogeneous ensemble models over their homogeneous counterparts.

The exceptional performance of heterogeneous ensemble models can be attributed to the inclusion of diverse models within the ensemble. In a heterogeneous ensemble, each model brings its unique strengths to the collective, ultimately enhancing the ensemble's overall predictive capability. This stands in contrast to homogeneous ensemble models, which consist of identical models and lack the crucial diversity found in heterogeneous ensembles (Khairalla et al., 2018).

Furthermore, the superiority of a heterogeneous ensemble model over a single machine learning model can be understood through the concept of the "wisdom of the crowd". While a single model may be well-tuned, it is inherently limited by its biases and assumptions. On the other hand, an ensemble of models can leverage the strengths of each individual model while compensating for their weaknesses. This leads to more robust and accurate predictions (Ji & Levinson, 2020).

When utilized as a decision fusion strategy within a heterogeneous machine learning model, the differential evolution (DE) algorithm has showcased remarkable performance in comparison to the unweighted average (UWA) approach. This superiority is evident in the reduced prediction errors and higher R-squared values achieved by the DE approach. The exceptional performance of the DE algorithm can be attributed to its ability to optimize the weights for decision fusion. Through this optimization process, the DE algorithm effectively balances the contributions of different models within the ensemble, thereby enhancing the overall predictive accuracy of the model. In contrast, the UWA approach assigns equal weights to all models, which may not be optimal if some models are more accurate or relevant than others.

Moreover, the higher R-squared values attained by the DE approach and the heterogeneous ensemble model indicate a better fit to the data and explain a greater proportion of the variance. This suggests that these methods can effectively capture the underlying patterns in the data, outperforming the UWA approach, homogeneous models, or single models. These findings are consistent with the research conducted by Matitopanum et al. (2023), which also demonstrated improved prediction accuracy with the DE approach in an ensemble machine learning model. This further emphasizes the potential of the DE approach to be used as a decision fusion strategy in heterogeneous machine learning models, offering more accurate and reliable predictions compared to conventional methods, such as the UWA approach.

In this investigation, we further delve into the influence of the dataset magnitude on the efficacy of the suggested ensemble model. The outcomes imply the existence of a critical quantity of instances required for the model to exhibit satisfactory performance, beyond which the incorporation of supplementary instances may not substantially enhance the model's performance. Nevertheless, the pronounced disparity in performance across various subgroups suggests the potential onset of diminishing returns with respect to the augmentation of the dataset size.

This observation aligns with the findings of Bayraci & Susuz, (2019), who also identified a threshold for the dataset size beyond which additional data did not significantly improve the model performance. Similarly, the concept of diminishing returns in relation to dataset size expansion has been previously discussed by Jones et al. (2020), further corroborating our results.

The prevailing methodologies for the prediction of the asiaticoside con-centration in Centella asiatica (CA-CA) have predominantly relied on con-ventional statistical techniques, exemplified by regression analysis and principal component analysis. Despite yielding certain insights into factors influencing the asiaticoside concentration, these methodologies exhibit in-herent limitations. One such constraint lies in their presumption of a linear association between the predictor variables and the response variable, a presumption that may not consistently hold in intricate biological systems such as CA-CA. Additionally, these approaches may fail in capturing non-linear interactions among predictor variables, consequently resulting in imprecise predictions.

Conversely, the heterogeneous ensemble machine learning approach adopted in this study offers a more adaptable and dynamic modeling framework, adept at elucidating non-linear relationships between predictor variables and the response variable. This is achieved through the amalgamation of multiple machine learning algorithms, harnessing the distinct strengths of each to foster the development of a more accurate and resilient prediction model. Furthermore, the employment of a differential evolution algorithm to optimize the ensemble model's weights enabled the attainment of heightened prediction accuracy, eclipsing the performance of extant methods.

While acknowledging that the D-Optimal design adopted in this study carries certain limitations such as the implicit assumption of linearity and the inherent restrictions on the number of viable experiments it nonetheless empowered us to judiciously select a subset of parameters. This selection was purposefully devised to maximize information acquisition from experiments while concurrently minimizing the requisite number of experimental iterations.

5.3. Policy implications

The outcomes of this research have significant policy implications in the realm of plant-based medicine and pharmaceuticals. CAU, a valuable medicinal plant with extensive usage in traditional medicine for the

treatment of various conditions, such as skin diseases, nervous disorders, and rheumatism, requires accurate prediction of the asiaticoside concentration. This precision is crucial in producing high-quality medicinal products derived from CAU. To address this, our study proposes an ensemble LSTM model that incorporates a heterogeneous ensemble and employs a Differential Evolution algorithm as the decision fusion strategy. This approach can enhance the accuracy in predicting the asiaticoside concentration, thereby ensuring the potency and efficacy of CAU-based medicinal products.

Furthermore, our findings contribute to the broader literature on machine learning and ensemble modeling. The results affirm the effectiveness of employing diverse models with distinct strengths within a heterogeneous ensemble, an approach previously validated in studies such as those by Nanglia et al. (2022), Zhao et al. (2019). Moreover, the study emphasizes the significance of utilizing an ensemble model and a decision fusion strategy, such as Differential Evolution, to improve the precision of machine learning models. These insights can be extrapolated to other prediction tasks across various domains, including finance, engineering, and healthcare.

Additionally, our study sheds light on the policy implications concerning the dataset size. The identification of a threshold number of samples required for the model to perform optimally implies that researchers and practitioners should ensure that they have an adequate sample size when developing machine learning models for prediction tasks. Moreover, the concept of diminishing returns associated with increasing the dataset size underscores the need to achieve a balance between the data quantity and quality during the model development process. Researchers and practitioners should carefully consider this trade-off to optimize the performance of machine learning models.

The findings of this study carry substantial policy implications for plant-based medicine, pharmaceuticals, and the broader domain of machine learning and ensemble modeling. The proposed ensemble LSTM model, incorporating a heterogeneous ensemble and Differential Evolution algorithm as the decision fusion strategy, has the potential to enhance the accuracy and efficacy of CAU-based medicinal products as well as addressing other prediction problems across diverse fields.

6. Conclusions and outlook

This research study aimed to address a significant knowledge gap by investigating the concentration of asiaticoside in CAU and its relationships with different growing conditions. The significance of CAU as a medicinal plant and the need for precise prediction models to optimize its cultivation practices formed the basis of this investigation. To bridge this research gap, we adopted a comprehensive methodology that involved analyzing variations in the asiaticoside concentration across different species and varieties of CAU. Furthermore, we explored the medicinal properties and secondary metabolites associated with this plant and assessed the impact of the growing conditions on the concentration of asiaticoside. This study approach encompassed laboratory tests, a simulation of the proposed model, and the utilization of various machine learning techniques.

The primary outcome of this study was the development of the He-ML model, which is an ensemble machine learning approach that was specifically designed to predict the concentration of asiaticoside in CAU. The He-ML model outperformed homogeneous models, exhibiting remarkable accuracy with a mean absolute error (MAE) of 47.74 mg/g and a root-mean-square error (RMSE) of 61.86 mg/g. These numerical results unequivocally demonstrate the superior performance and effectiveness of the He-ML model in terms of accurately predicting the concentration of asiaticoside. Furthermore, the He-ML model outperformed other state-of-the-art machine learning models, such as the LSTM, GRU, Conv.-LSTM, A-LSTM, RFR, SVR, GBR, and SRU, by 40.62 %, 35.68 %, 104.70 %, 139.49 %, 230.81 %, 205.95 %, 250.34 %, and 191.06 %, respectively.

One of the key discoveries of this research is the profound influence of the growing conditions, such as the specific CAU cultivar or strain and the ratio of different light spectra, on the concentration of asiaticoside. These findings highlight the need for a prediction model that can accurately estimate the concentration of asiaticoside based on the specific growing conditions. The He-ML model presents valuable insights for the optimization of cultivation practices and the maximization of the utilization of CAU as a medicinal plant.

This research article significantly contributes to our understanding of the concentration of asiaticoside in CAU and introduces the He-ML model as a powerful tool to predict this concentration. The numerical results underscore the model's accuracy, while the research findings emphasize the importance of considering the growing conditions to optimize cultivation practices. These findings have broad implications for various industries, including pharmaceuticals, skincare and cosmetics, and nutraceuticals, opening avenues for further research and practical applications in these domains.

This research demonstrates the efficacy of employing a heterogeneous ensemble machine learning methodology to predict the asiaticoside concentration within Centella Asiatica (CA-CA). Through the amalgamation of diverse machine learning algorithms and the application of a differential evolution algorithm to optimize the ensemble model's weights, we achieved a notable prediction accuracy exceeding 90 %. This approach constitutes a substantial enhancement over established techniques, predominantly reliant on conventional statistical methods that are not able to capture the intricate nonlinear relationships between predictor variables and response variables.

The exceptional performance of the heterogeneous ensemble machine learning model employed in this investigation can be attributed to a confluence of key factors. Firstly, the incorporation of multiple machine learning algorithms enabled us to harness the individual strengths of each algorithm, thus culminating in the development of a prediction model characterized by heightened accuracy and robustness. Secondly, the differential evolution algorithm, harnessed for the optimization of the ensemble model's weights, was instrumental in achieving superior prediction accuracy compared to extant methodologies.

Thirdly, the incorporation of a heterogeneous attribute within the LSTM architecture endowed the model with the capacity to capture intricate non-linear relationships existing among predictor variables and the response variable. Lastly, the strategic implementation of a D-Optimal design afforded us the ability to judiciously select a subset of parameters. This selection was engineered to maximize the information yield from experiments while concurrently minimizing the requisite number of experimental iterations. Collectively, these facets synergized to yield a prediction model of remarkable efficacy and precision. Such a model holds substantial promise for optimizing the cultivation and utilization of Centella asiatica as a medicinal botanical.

Furthermore, this innovative approach holds the potential to transform the cultivation and utilization of CA-CA as a medicinal botanical. It offers a more precise and efficient mechanism for the prediction of asiaticoside concentrations and optimization of cultivation conditions. By addressing the limitations inherent in existing methods, which assume linearity and struggle to capture nonlinear interactions, the adaptable and flexible nature of the heterogeneous ensemble machine learning approach employed herein fills a critical research gap. In essence, this study lays the groundwork for future investigations into the application of machine learning to predict the concentrations of bioactive compounds in medicinal plants.

While this investigation offers valuable insights into the determinants affecting the asiaticoside concentration in Centella asiatica (CA-CA), it is imperative to acknowledge its inherent constraints. One limitation pertains to the study's conduction within controlled laboratory settings, potentially limiting its ability to fully encapsulate the broad spectrum of growth conditions prevailing across diverse geographical regions. Furthermore, the dataset utilized in this inquiry was derived from a restricted selection of CA-CA cultivars and strains, potentially hindering from a comprehensive representation of the entire

array of CA-CA species and varieties.

Moreover, although the employed D-Optimal design was suitable in selecting a subset of parameters to optimize the information gleaned from the experiments, the possibility remains that additional parameters not accounted for within the design may also exert an influence on the asiaticoside concentration in CA-CA. Finally, although the machine learning models used in this study exhibited encouraging outcomes in predicting the asiaticoside concentration, further validation is requisite to ascertain their precision and applicability in a broader context. Notwithstanding these limitations, this study serves as a foundational platform for prospective research endeavors pertaining to the cultivation and exploitation of CA-CA as a medicinal botanical.

Future research directions stemming from this study involve exploring the underlying factors contributing to the negative correlation between the light exposure duration and the asiaticoside concentration, delving deeper into the impact of the growing conditions on the asiaticoside concentration, developing prediction models for other medicinal plants based on the He-ML model, exploring the potential health benefits of CAU-based medicinal products, and examining the applicability of alternative machine learning techniques for predicting the asiaticoside concentration. These research directions will contribute to a deeper understanding of medicinal plant cultivation, the optimization of cultivation practices, and the exploration of the potential applications of CAU in various industries.

CRediT authorship contribution statement

Keartisak Sriprateep: Conceptualization, Investigation, Data curation, Methodology, Validation, Formal analysis, Software, Writing – original draft. **Sarinya Sala-Ngamand:** Conceptualization, Investigation, Data curation, Methodology, Validation, Formal analysis, Software, Writing – original draft. **Surajet Khonjun:** Funding acquisition, Resources, Conceptualization, Writing – review & editing. **Ming-Lang Tseng:** Writing – review & editing. **Thanatkij Srichok:** Resources, Writing – review & editing. **Natthapong Nanthasamroeng:** Resources. **Rapeepan Pitakaso:** Supervision, Conceptualization, Writing – review & editing. **Narut Butploy:** Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research project was financially supported by Mahasarakham University and National Science, Research and In-novation Fund (NSRF), Grant No. 4161845.

References

Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research, 297*, 3–36.

Alcalde, M. A., Müller, M., Munné-Bosch, S., Landín, M., Gallego, P. P., Bonfill, M., et al. (2022). Using machine learning to link the influence of transferred Agrobacterium rhizogenes genes to the hormone profile and morphological traits in Centella asiatica hairy roots. *Frontiers in Plant Science, 13*, Article 1001023.

Alcazar Magana, A., Wright, K., Vaswani, A., Caruso, M., Reed, R. L., Bailey, C. F., … Quinn, J. (2020). Integration of mass spectral fingerprinting analysis with precursor ion (MS1) quantification for the characterisation of botanical extracts: Application to extracts of Centella asiatica (L.) Urban. *Phytochemical Analysis, 31*(6), 722–738.

Attimarad, M., Venugopala, K. N., Sreeharsha, N., Chohan, M. S., Shafi, S., Nair, A. B., et al. (2021). A rapid HPLC method for the concurrent determination of several antihypertensive drugs from binary and ternary formulations. *Separations, 8*(6), 86.

Baby, M. S. B. M. S., Antony, A. V., John, B., & Jose, D. (2020). Phytochemical screening and invitro anti-inflamatory activity of ethanolic extract of centella asiatica. *World Journal of Current Medical and Pharmaceutical Research*, 181–183.

Bandopadhyay, S., Mandal, S., Ghorai, M., Jha, N. K., Kumar, M., Radha, et al. (2023). Therapeutic properties and pharmacological activities of asiaticoside and madecassoside: A review. *Journal of Cellular and Molecular Medicine, 27*(5), 593–608.

Banik, A., Behera, C., Sarathkumar, T. V., & Goswami, A. K. (2020). Uncertain wind power forecasting using LSTM-based prediction interval. *IET Renewable Power Generation, 14*(14), 2657–2667.

Bayraci, S., & Susuz, O. (2019). A Deep Neural Network (DNN) based classification model in application to loan default prediction. *Theoretical and Applied Economics, 4*(621), 75–84.

Beatrice, P., Saviano, G., Reguzzoni, M., Divino, F., Fantasma, F., Chiatante, D., et al. (2023). Light spectra of biophilic LED-sourced system modify essential oils composition and plant morphology of Mentha piperita L. and Ocimum basilicum L. *Frontiers in Plant Science, 14*, Article 1093883.

Binti Mohd Yusuf Yeo, N. A., Muthuraju, S., Wong, J. H., Mohammed, F. R., Senik, M. H., Zhang, J., et al. (2018). Hippocampal amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid GluA1 (AMPA GluA1) receptor subunit involves in learning and memory improvement following treatment with Centella asiatica extract in adolescent rats. *Brain and Behavior, 8*(9), e01093.

Brito, C., Ferreira, H., Dinis, L.-T., Trindade, H., Marques, D., Correia, C. M., et al. (2023). Different LED light intensity and quality change perennial ryegrass (Lolium perenne L.) physiological and growth responses and water and energy consumption. *Frontiers in Plant Science, 14*, Article 1160100.

Chen, W., Sharifrazi, D., Liang, G., Band, S. S., Chau, K. W., & Mosavi, A. (2022). Accurate discharge coefficient prediction of streamlined weirs by coupling linear regression and deep convolutional gated recurrent unit. *Engineering Applications of Computational Fluid Mechanics, 16*(1), 965–976.

de Gregorio, A., Janni, W., Friedl, T. W., Nitz, U., Rack, B., Schneeweiss, A., et al. (2022). The impact of anthracyclines in intermediate and high-risk HER2-negative early breast cancer—A pooled analysis of the randomised clinical trials PlanB and SUCCESS C. *British Journal of Cancer, 126*(12), 1715–1724.

Dehghan, A., Razzaghi, P., Abbasi, K., & Gharaghani, S. (2023). TripletMultiDTI: Multimodal representation learning in drug-target interaction prediction with triplet loss function. *Expert Systems with Applications*, Article 120754.

Elhenawy, M., Chen, H., & Rakha, H. A. (2014). Dynamic travel time prediction using data clustering and genetic programming. *Transportation Research Part C: Emerging Technologies, 42*, 82–98.

Essien, A., Petrounias, I., Sampaio, P., & Sampaio, S. (2021). A deep-learning model for urban traffic flow prediction with traffic events mined from twitter. *World Wide Web, 24*(4), 1345–1368.

Flannery, S. E., Pastorelli, F., Emrich-Mills, T. Z., Casson, S. A., Hunter, C. N., Dickman, M. J., et al. (2023). STN7 is not essential for developmental acclimation of Arabidopsis to light intensity. *The Plant Journal, 114*(6), 1458–1474.

Fodorpataki, L., Molnar, K., Tompa, B., & Plugaru, S. R. (2019). Priming with vitamin U enhances cold tolerance of lettuce (Lactuca sativa L.). *Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 47*(3), 592–598.

Ganaie, M. A., Hu, M., Malik, A., Tanveer, M., & Suganthan, P. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence, 115*, Article 105151.

Gohil, K. J., Patel, J. A., & Gajjar, A. K. (2010). Pharmacological review on Centella asiatica: A potential herbal cure-all. *Indian journal of pharmaceutical sciences, 72*(5), 546.

Gómez, S., & Gómez, C. (2022). Evaluating the use of Biostimulants for Indoor Hydroponic Lettuce Production. *HortTechnology, 32*(4), 348–355.

Gong, L., Yu, M., Jiang, S., Cutsuridis, V., & Pearson, S. (2021). Deep learning based prediction on greenhouse crop yield combined TCN and RNN. *Sensors, 21*(13), 4537.

Guzmán, C., Mondal, S., Govindan, V., Autrique, J. E., Posadas-Romano, G., Cervantes, F., et al. (2016). Use of rapid tests to predict quality traits of CIMMYT bread wheat genotypes grown under different environments. *LWT-Food Science and Technology, 69*, 327–333.

Hoang, H. L., & Rehman, H. (2022). Unravelling the Morphological, Physiological, and Phytochemical responses in Centella asiatica L. Urban to Incremental Salinity Stress. *Life (Chicago, Ill. : 1978), 13*(1), 61.

Huang, J., Algahtani, M., & Kaewunruen, S. (2022). Energy forecasting in a public building: A benchmarking analysis on long short-term memory (LSTM), support vector regression (SVR), and extreme gradient boosting (XGBoost) networks. *Applied Sciences, 12*(19), 9788.

James, J. T., Tugizimana, F., Steenkamp, P. A., & Dubery, I. A. (2013). Metabolomic analysis of methyl jasmonate-induced triterpenoid production in the medicinal herb Centella asiatica (L.) urban. *Molecules (Basel, Switzerland), 18*(4), 4267–4281.

Jescovitch, L. N., Scott, E. E., Cerchiara, J. A., Merrill, J., Urban-Lurain, M., Doherty, J. H., et al. (2021). Comparison of machine learning performance using analytic and holistic coding approaches across constructed response assessments aligned to a science learning progression. *Journal of Science Education and Technology, 30*(2), 150–167.

Ji, A., & Levinson, D. (2020). Injury severity prediction from two-vehicle crash mechanisms with machine learning and ensemble models. *IEEE Open Journal of Intelligent Transportation Systems, 1*, 217–226.

Jin, S., Ibrahim, M., Muhammad, S., Khan, S., & Li, G. (2020). Light intensity effects on the growth and biomass production of submerged macrophytes in different water strata. *Arabian Journal of Geosciences, 13*, 1–7.

Jones, R., Klemenjak, C., Makonin, S., & Bajic, I.V. (2020). Exploring bayesian surprise to prevent overfitting and to predict model performance in non-intrusive load monitoring. *arXiv preprint arXiv:2009.07756*.

Jose, D. M., Vincent, A. M., & Dwarakish, G. S. (2022). Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques. *Scientific Reports, 12*(1), 1–25.

Jusril, N. A., Abu Bakar, S. I., Khalil, K. A., Md Saad, W. M., Wen, N. K., & Adenan, M. I (2022). Development and optimization of Nanoemulsion from Ethanolic extract of Centella asiatica (NanoSECA) Using D-optimal mixture design to improve blood-brain barrier permeability. *Evidence-Based Complementary and Alternative Medicine*, 2022.

Khairalla, M. A., Ning, X., Al-Jallad, N. T., & El-Faroug, M. O. (2018). Short-term forecasting for energy consumption through stacking heterogeneous ensemble learning model. *Energies, 11*(6), 1605.

Khan, M. S., Riaz, R., Majid, M., Mehmood, K., Mustafa, G., & Joyia, F. A. (2022). The tobacco chloroplast YCF4 gene is essential for transcriptional gene regulation and plants photoautotrophic growth. *Frontiers in Plant Science, 13*, Article 1014236.

Koh, J. C., Hayden, M., Daetwyler, H., & Kant, S. (2019). Estimation of crop plant density at early mixed growth stages using UAV imagery. *Plant methods, 15*, 1–9.

Lal, R., Gupta, P., & Dubey, B. K. (2017). Genetic variability and associations in the accessions of Manduk parni {Centella asiatica (L)}. *Industrial crops and products, 96*, 173–177.

Lezhenin, I., Bogach, N., & Pyshkin, E. (2019). Urban sound classification using long short-term memory neural network. In *2019 federated conference on computer science and information systems (FedCSIS)*.

Li, B., Feng, Y., Zong, Y., Zhang, D., Hao, X., & Li, P. (2020a). Elevated CO_2-induced changes in photosynthesis, antioxidant enzymes and signal transduction enzyme of soybean under drought stress. *Plant Physiology and Biochemistry, 154*, 105–114.

Li, C., Xie, X., Li, F., Tian, E., Shu, Y., & Chao, Z. (2020b). The complete chloroplast genome sequence of Centella asiatica (Linnaeus) Urban. *Mitochondrial DNA Part B, 5*(3), 2149–2150.

Li, X., Sha, J., & Wang, Z.-l. (2017). A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen. *Hydrology Research, 48*(5), 1214–1225.

Ling, J., Zhu, Z., Luo, Y., & Wang, H. (2021). An intrusion detection method for industrial control systems based on bidirectional simple recurrent unit. *Computers & Electrical Engineering, 91*, Article 107049.

Liu, A., Henke, M., Li, Y., Zhang, Y., Xu, D., Liu, X., et al. (2022). Investigation of the impact of supplemental reflective films to improve micro-light climate within tomato plant canopy in solar greenhouses. *Frontiers in Plant Science, 13*, Article 966596.

Lynam, A. L., Dennis, J. M., Owen, K. R., Oram, R. A., Jones, A. G., Shields, B. M., et al. (2020). Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: Application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagnostic and prognostic research, 4*(1), 1–10.

Mahayasih, P. G. M. W., Dewi, S., Tamher, R. P., Megawati, M., Rahayu, S. T., & Pertiwi, T. S. (2022). *Pengolahan Bahan Alam dalam Upaya Meningkatkan Daya Tahan Tubuh di Era Pandemi, 2* pp. 217–223). Jurnal Pengabdian Masyarakat Indonesia.

Maimaitiyiming, M., Sagan, V., Sidike, P., & Kwasniewski, M. T. (2019). Dual activation function-based Extreme Learning Machine (ELM) for estimating grapevine berry yield and quality. *Remote Sensing, 11*(7), 740.

Matitopanum, S., Pitakaso, R., Sethanan, K., Srichok, T., & Chokanat, P. (2023). Prediction of the Ultimate Tensile Strength (UTS) of asymmetric friction stir welding using ensemble machine learning methods. *Processes, 11*(2), 391.

Matthews, D. G., Caruso, M., Alcazar Magana, A., Wright, K. M., Maier, C. S., Stevens, J. F., et al. (2020). Caffeoylquinic acids in Centella asiatica reverse cognitive deficits in male 5XFAD Alzheimer's disease model mice. *Nutrients, 12*(11), 3488.

Mohiuddin, A. (2019). Sunscreen and coral reef: Letter to the editor. *MOJ Eco Environ Sci, 4*(4), 166–167.

Moreta, C. E. G., Acosta, M. R. C., & Koo, I. (2019). Prediction of digital terrestrial television coverage using machine learning regression. *IEEE Transactions on Broadcasting, 65*(4), 702–712.

Moustakas, M., Sperdouli, I., & Moustaka, J. (2022). Early drought stress warning in plants: Color pictures of photosystem II photochemistry. *Climate, 10*(11), 179.

Müller, V., Lankes, C., Zimmermann, B. F., Noga, G., & Hunsche, M. (2013). Centelloside accumulation in leaves of Centella asiatica is determined by resource partitioning between primary and secondary metabolism while influenced by supply levels of either nitrogen, phosphorus or potassium. *Journal of Plant Physiology, 170*(13), 1165–1175.

Naing, A. H., Jeon, S. M., Park, J. S., & Kim, C. K. (2016). Combined effects of supplementary light and CO_2 on rose growth and the production of good quality cut flowers. *Canadian Journal of Plant Science, 96*(3), 503–510.

Namviriyachote, N., Lipipun, V., Akkhawattanangkul, Y., Charoonrut, P., & Ritthidej, G. C. (2019). Development of polyurethane foam dressing containing silver and asiaticoside for healing of dermal wound. *Asian Journal of Pharmaceutical Sciences, 14*(1), 63–77.

Nanglia, S., Ahmad, M., Khan, F. A., & Jhanjhi, N. (2022). An enhanced Predictive heterogeneous ensemble model for breast cancer prediction. *Biomedical Signal Processing and Control, 72*, Article 103279.

Nongrum, Y., Prasad, R., Gupta, A., & Tiwari, M. (2020). Development and nutritional characterization of noodles enriched with Centella asiatica powder. *IJCS, 8*(2), 2183–2186.

Okut, H. (2021). Deep learning for subtyping and prediction of diseases: Long-short term memory. *Deep learning applications*. IntechOpen.

Osco, L. P., Junior, J. M., Ramos, A. P. M., Furuya, D. E. G., Santana, D. C., Teodoro, L. P. R., et al. (2020). Leaf nitrogen concentration and plant height prediction for maize using UAV-based multispectral imagery and machine learning techniques. *Remote Sensing, 12*(19), 3237.

Özdemir, A., & Turkoz, M. (2020). Development of a D-optimal design-based 0–1 mixed-integer nonlinear robust parameter design optimization model for finding optimum design factor level settings. *Computers & Industrial Engineering, 149*, Article 106742.

Prasad, A., Mathur, A., & Mathur, A. (2019). Advances and emerging research trends for modulation of centelloside biosynthesis in Centella asiatica (L.) Urban-A review. *Industrial crops and products, 141*, Article 111768.

Prasitpuriprecha, C., Jantama, S. S., Preeprem, T., Pitakaso, R., Srichok, T., Khonjun, S., et al. (2022). Drug-resistant tuberculosis treatment recommendation, and multi-class tuberculosis detection and classification using ensemble deep learning-based system. *Pharmaceuticals, 16*(1), 13.

Rafiei, F., Zeraati, H., Abbasi, K., Ghasemi, J. B., Parsaeian, M., & Masoudi-Nejad, A. (2023). Deep Tra Synergy: Drug combinations using multimodal deep learning with transformers. *Bioinformatics (Oxford, England), 39*(8), btad438. https://doi.org/10.1093/bioinformatics/btad438

Ragmani, A., Elomri, A., Abghour, N., Moussaid, K., Rida, M., & Badidi, E. (2020). Adaptive fault-tolerant model for improving cloud computing performance using artificial neural network. *Procedia Computer Science, 170*, 929–934.

Rahmah, G. S., Khairunnisa, Z., Novalia, V., & Akbar, M. K. (2022). Uji efektivitas larvasida ekstrak daun pegagan (Centella asiatica L. Urban) terhadap larva culex quinquefasciatus. *COMSERVA: Jurnal Penelitian dan Pengabdian Masyarakat, 2*(7), 915–923.

Rajamani, S. T., Rajamani, K. T., Mallol-Ragolta, A., Liu, S., & Schuller, B. (2021). A novel attention-based gated recurrent unit and its efficacy in speech emotion recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Rao, M. S., & Ramanaiah, N. (2018). Optimization of Process parameters for FSW of Al-Mg-Mn-*Sc*-Zr alloy using CCD and RSM. *Strojnícky časopis-Journal of Mechanical Engineering, 68*(3), 195–224.

Razzaghi, P., Abbasi, K., Shirazi, M., & Rashidi, S. (2022). Multi modal brain tumor detection using multimodal deep transfer learning. *Applied Soft Computing, 129*, Article 109631. https://doi.org/10.1016/j.asoc.2022.109631

Reddy, G. T., Bhattacharya, S., Ramakrishnan, S. S., Chowdhary, C. L., Hakak, S., Kaluri, R., et al. (2020). An ensemble based machine learning model for diabetic retinopathy classification. In *2020 international conference on emerging trends in information technology and engineering (ic-ETITE)*.

Rohini, M., & Smitha, G. (2022). Studying the effect of morphotype and harvest season on yield and quality of Indian genotypes of Centella asiatica: A potential medicinal herb cum underutilized green leafy vegetable. *South African Journal of Botany, 145*, 275–283.

Ruksiriwanich, W., Khantham, C., Sringarm, K., Sommano, S., & Jantrawut, P. (2020). Depigmented Centella asiatica extraction by pretreated with supercritical carbon dioxide fluid for wound healing application. *Processes, 8*(3), 277.

Smith, P. F., Ganesh, S., & Liu, P. (2013). A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *Journal of neuroscience methods, 220*(1), 85–91.

Song, C., Shi, X., & Wang, J. (2020). Spatiotemporally varying coefficients (STVC) model: A Bayesian local regression to detect spatial and temporal nonstationarity in variables relationships. *Annals of GIS, 26*(3), 277–291.

Song, J., Zhang, L., Xue, G., Ma, Y., Gao, S., & Jiang, Q. (2021). Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model. *Energy and Buildings, 243*, Article 110998.

TAŞCI, E (2019). A meta-ensemble classifier approach: Random rotation forest. *Balkan Journal of Electrical and Computer Engineering, 7*(2), 182–187.

Thapa, R., Zhang, K., Snavely, N., Belongie, S., & Khan, A. (2020). The plant pathology challenge 2020 data set to classify foliar disease of apples. *Applications in plant sciences, 8*(9), e11390.

Tušek, A. J., Jurina, T., Benković, M., Valinger, D., Belščak-cvitanović, A., & Kljusurić, J. G (2020). Application of multivariate regression and artificial neural network modelling for prediction of physical and chemical properties of medicinal plants aqueous extracts. *Journal of Applied Research on Medicinal and Aromatic Plants, 16*, Article 100229.

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review, 53*, 5929–5955.

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture, 177*, Article 105709.

Vinolina, N., & Sigalingging, R. (2021). Variation in growth of Centella asiatica of Samosir–Indonesia accession with phosphorus fertilizer cultivated at Samosir Field. *Journal of Physics: Conference Series*.

Vinolina, N. S., & Sigalingging, R. (2022). Growth and secondary metabolites production of Centella asiatica (L.) Urb. cultivated at different phosphate application rates in acid soil. *Trends in Sciences, 19*(18), 5820-5820.

Vinsonneau, L., Labrosse, J., Porcu-Buisson, G., Chevalier, N., Galey, J., Ahdad, N., et al. (2022). Impact of endometrial preparation on early pregnancy loss and live birth rate after frozen embryo transfer: A large multicenter cohort study (14 421 frozen cycles). *Human Reproduction Open, 2022*(2), hoac007.

Wang, L., Guo, T., Guo, Y., & Xu, Y. (2020). Asiaticoside produces an antidepressant-like effect in a chronic unpredictable mild stress model of depression in mice, involving reversion of inflammation and the PKA/pCREB/BDNF signaling pathway. *Molecular Medicine Reports, 22*(3), 2364–2372.

Yang, K., Chen, G., Xian, J., & Chang, H. (2023). Divergent adaptations of leaf functional traits to light intensity across common urban plant species in Lanzhou, northwestern China. *Frontiers in Plant Science, 14*, Article 1000647.

Yang, S., Linares-Barranco, B., & Chen, B. (2022). Heterogeneous ensemble-based spike-driven few-shot online learning. *Frontiers in Neuroscience*, 16.

Yang, W., He, S., Xiao, N., Qiao, Y., Sui, H., Liang, L., et al. (2019). Simultaneous determination of 15 flavonoids in Scutellaria barbata–Hedyotis diffusa Herb Pair by HPLC Q-TOF MS. *Journal of AOAC International, 102*(1), 75–80.

Yin, J., Medellín-Azuara, J., Escriva-Bou, A., & Liu, Z. (2021). Bayesian machine learning ensemble approach to quantify model uncertainty in predicting groundwater storage change. *Science of the Total Environment, 769*, Article 144715.

Zhang, D., Yang, J., Li, F., Han, S., Qin, L., & Li, Q. (2022). Landslide risk prediction model using an attention-based temporal convolutional network connected to a recurrent neural network. *IEEE access : practical innovations, open solutions, 10*, 37635–37645.

Zhang, G., Lu, H., Dong, J., Poslad, S., Li, R., Zhang, X., et al. (2020). A framework to predict high-resolution spatiotemporal PM2. 5 distributions using a deep-learning model: A case study of Shijiazhuang, China. *Remote Sensing, 12*(17), 2825.

Zhang, W., Wang, L., Chen, J., Xiao, W., & Bi, X. (2021). A novel gas recognition and concentration detection algorithm for artificial olfaction. *IEEE Transactions on Instrumentation and Measurement, 70*, 1–14.

Zhao, H., Liu, F., Zhang, H., & Liang, Z. (2019). Convolutional neural network based heterogeneous transfer learning for remote-sensing scene classification. *International Journal of Remote Sensing, 40*(22), 8506–8527.

Zhou, Q., Zhou, C., & Wang, X. (2022). Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and feature selection. *PloS one, 17*(2), Article e0262501.