

Article

Combination of Multivariate Standard Addition Technique and Deep Kernel Learning Model for Determining Multi-Ion in Hydroponic Nutrient Solution

Vu Ngoc Tuan ^{1,2,3}, Abdul Mateen Khattak ^{2,4}, Hui Zhu ^{5,6}, Wanlin Gao ^{1,2}
and Minjuan Wang ^{1,2,*}

¹ Key Laboratory of Agricultural Informatization Standardization, Ministry of Agriculture and Rural Affairs, Beijing 100083, China; tuan.vn.nd@gmail.com (V.N.T.); gaowlin@cau.edu.cn (W.G.)

² College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; mateen@aup.edu.pk

³ Faculty of Electrical and Electronic Engineering, Nam Dinh University of Technology Education, Nam Dinh 420000, Vietnam

⁴ Departemnt of Horticulture, The University of Agriculture, Peshawar 25120, Pakistan

⁵ Key Laboratory of Liquor Making Biological Technology and Application, Zigong 643000, China; zhuhuiwn@outlook.com

⁶ School of Bioengineering, Sichuan University of Science and Engineering, Zigong 643000, China

* Correspondence: minjuan@cau.edu.cn

Received: 4 July 2020; Accepted: 24 August 2020; Published: 17 September 2020



Abstract: Ion-selective electrodes (ISEs) have recently become the most attractive tools for the development of efficient hydroponic systems. Nevertheless, some inherent shortcomings such as signal drifts, secondary ion interferences, and effected high ionic strength make them difficult to apply in a hydroponic system. To minimize these deficiencies, we combined the multivariate standard addition (MSAM) sampling technique with the deep kernel learning (DKL) model for a six ISEs array to increase the prediction accuracy and precision of eight ions, including NO_3^- , NH_4^+ , K^+ , Ca^{2+} , Na^+ , Cl^- , H_2PO_4^- , and Mg^{2+} . The enhanced data feature based on feature enrichment (FE) of the MSAM technique provided more useful information to DKL for improving the prediction reliability of the available ISE ions and enhanced the detection of unavailable ISE ions (phosphate and magnesium). The results showed that the combined MSAM–feature enrichment (FE)–DKL sensing structure for validating ten real hydroponic samples achieved low root mean square errors (RMSE) of 63.8, 8.3, 29.2, 18.5, 11.8, and 8.8 $\text{mg}\cdot\text{L}^{-1}$ with below 8% coefficients of variation (CVs) for predicting nitrate, ammonium, potassium, calcium, sodium, and chloride, respectively. Moreover, the prediction of phosphate and magnesium in the ranges of 5–275 $\text{mg}\cdot\text{L}^{-1}$ and 10–80 $\text{mg}\cdot\text{L}^{-1}$ had RMSEs of 29.6 and 8.7 $\text{mg}\cdot\text{L}^{-1}$ respectively. The results prove that the proposed approach can be applied successfully to improve the accuracy and feasibility of ISEs in a closed hydroponic system.

Keywords: ion-selective electrode; multi-ion sensor array; artificial neural network; gaussian process; deep kernel learning; hydroponics

1. Introduction

Hydroponics is a modern cultivation system involving aqueous solutions of nutrient salts to grow plants in a soilless culture. This farming system has been deployed widely in modern agriculture, where growers manipulate plant growth by adjusting fertilizer doses to increase crop yield and improve the quality of produce [1,2]. Particularly, the closed hydroponic system has been popularly

utilized because of the reusability of the drained solution. The system achieves better efficiency through used water and fertilizer that contributes toward sustainable and environmentally friendly agriculture. However, the closed hydroponic system still poses many challenges such as the imbalance in nutrient ratios caused by the lack or excess of some ions. This is due to a radical flaw in the conventional hydroponic system, where the nutrient concentration is controlled by modulating only electrical conductivity (EC) and pH values [3,4]. To avert this problem, growers normally adopt temporary measures—for example, periodically replacing the nutrient solutions or periodic sampling and analyzing nutrient solutions in a lab. However, these measures are inefficient either due to the use of extra water and fertilizer or loss of nutrient control due to lack of timely feedback. Ion-selective electrodes (ISEs) are considered useful tools to resolve this problem [5]. For example, the array of ISEs determines multi-ions simultaneously in samples of complex hydroponic solutions [6–8]. Nevertheless, several difficulties persist with the use of ISEs in the actual hydroponic system due to the effect of interfering ions, ionic strength, temperature noise [9,10], and the drift of ISEs' output signals [11].

Various approaches have been proposed to resolve the technical challenges of applying ISEs in closed hydroponic systems, such as applying particular sampling [12,13], calibrating techniques [14,15], and compensating interferences supported by machine learning techniques (for example, artificial neural network (ANN) and deep neural network (DNN)) [16–20]. Apart from that, preprocessing techniques as such principle component analysis (PCA) [21], independent component analysis (ICA) [22–24], partial least squares (PLS) model [25,26], and the kernel model as Gaussian Process (GP) model, support vector machine model (SVM) [27], etc. have also been used. However, none of technical strategies has radically resolved the drawbacks of ISEs. Other investigations focused on predicting the ions that are not available through an ion-selective electrode (as such PO_4 , SO_4 , and Mg) were examined by fusing the datasets acquired from an array of available ISEs [20,27–29].

Previously, ANN was the most attractive approach because it possessed flexibility in processing non-linear problems similar to ISE issues. Good generalization made ANN an essential tool for prediction based on the prior chemical relationships [20]. Nevertheless, the drift of ISEs' signal was the biggest factor that limited the application of ANN. Recently, ANN has been combined with other models to prepare dataset procedure, such as using the two-point normalization [30]. However, apart from the drift and interferences, ionic strength is a significant factor affecting the accuracy of ISEs [31]. Additionally, the training dataset and calibration sampling procedure considerably affect the performance of the models. Thus, modelers prepared a sufficient calibrating dataset by mixing solutions having individual ions appropriate for the target ranges [17,20]. A hydroponic nutrient solution contains several essential ions with different concentrations ranging from zero to hundreds of $mg \cdot L^{-1}$ [14,19,32]. This makes the preparation of mixture sample more complicated and time-consuming for acquiring the dataset. Therefore, the ANN does not work satisfactorily with a small dataset of 27 samples [19]. In case of a limited number of samples, GP is an appropriate choice [29]. However, the basic GP model does not fit well in high-dimensional input-output spaces [33]. Recent studies based on the parametric model [22,34] achieved several promising results, for example, simple explicit model construction, minimizing the calibration samples. However, to operate efficiently, these models require other conditions that are not suitable for hydroponic systems. For example, the assumption that the sources are mutually independent [22], the calibration samples for modeling must be in hundreds, and high computational costs [34] making them difficult for application in multi-ion sensing. Wilson et al. [35] proposed the deep kernel learning (DKL) model, which was a combination of DNN and GP models. The model performance was outstanding for image processing [36] and soft sensor purposes [37]. However, DKL was never applied to resolve the problems of ISEs in hydroponics systems.

This report presents an approach to combine the multivariate standard addition sampling technique with DKL for solving issues with ISEs to quantify the concentration of eight ions (i.e., nitrate, ammonium, potassium, calcium, sodium, chloride, phosphate, and magnesium) simultaneously in a hydroponic nutrient solution (Figure 1). The performance of DKL was improved by combining ANN (having a good generalizing capability) and GP (having flexibility induced power). This enhanced

the accuracy of determining multi-ions in a hydroponic solution by reducing interferences and uncertainties. Moreover, the multivariate standard addition method (MSAM) of sampling [29] was also utilized to prepare the training dataset for DKL. This was to overcome drift, interferences, and ionic strength obstacles of detecting multi-ions in hydroponic solutions. In this approach, we used a deep feed-forward network to extract a high-level representation of the data and also took advantage of the non-parametric flexibility induced by the Gaussian process regression. This improved the predictability of the proposed model for determining ions that are unavailable through commercial ISEs (i.e., phosphate and magnesium ions).

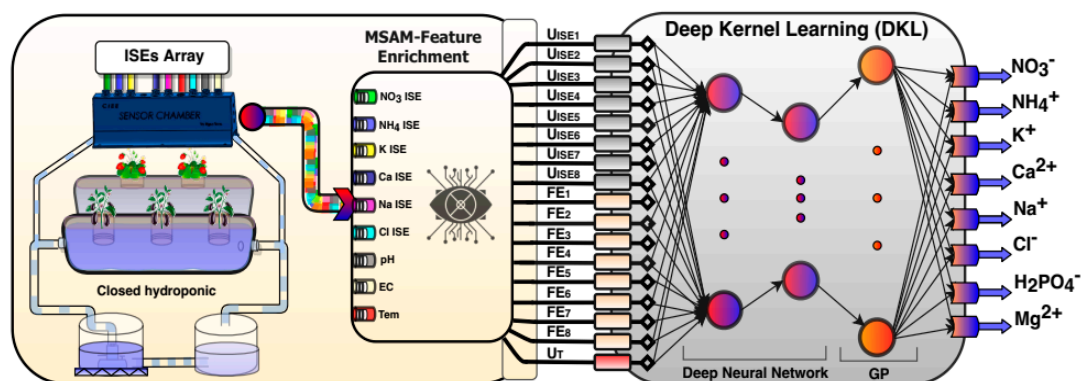


Figure 1. The novel combination of multivariate standard addition (MSAM)–feature enrichment (FE)–deep kernel learning (DKL) architecture for determining multi-ions in hydroponic solutions.

2. Materials and Methods

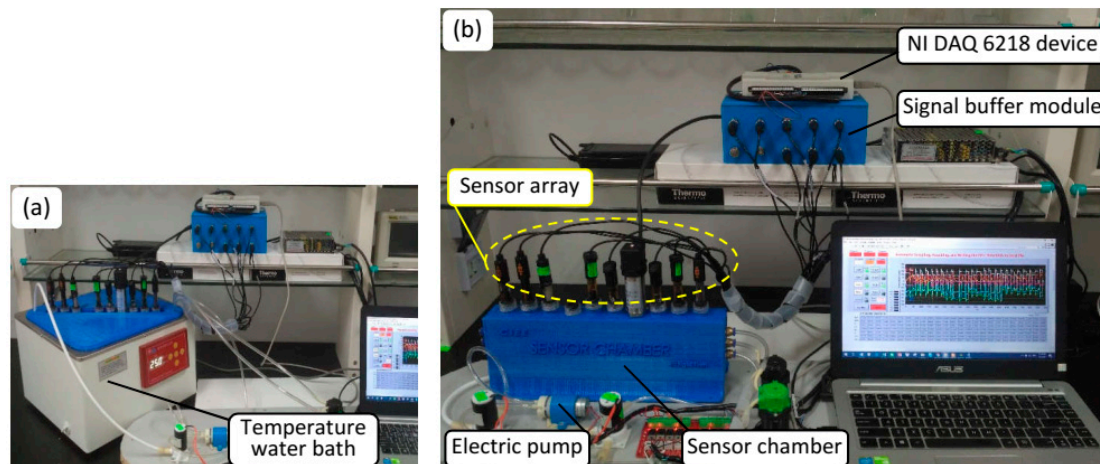
2.1. Experiment Preparation

2.1.1. Sensor Array and Apparatus

In this experiment, a sensor array composed of nine sensors including six commercial ISEs REX 972123, REX 972122, pNa 701, pCl 202 (Shanghai INESA, Shanghai, China), Orion 9719BNWP, and Orion 9320BNWP (Thermo Fisher Scientific, Waltham, MA, USA) for determining six ions (i.e., nitrate, ammonium, sodium, chloride, potassium, and calcium) and three transducers including an electrical conductivity probe (DJS-1C; Shanghai INESA, Shanghai, China), a pH electrode (E-201F; Shanghai INESA, Shanghai, China), and a temperature probe (Pt100; Yuace, Shanghai, China) for detecting conductivity, pH, and temperature of the samples was deployed. The basic specifications of the sensors are summarized in Table 1. The suite of sensors was plugged into a sensor chamber made of acrylonitrile butadiene styrene (ABS) plastic by a 3D printer connected with an electric pump (KLP05-6, Kamoer, Shanghai, China) to create a simple flow injection sampling structure. The sensor chamber was immersed into a temperature calibration water bath HH.S21-4 (Boxun, Shanghai, China) for calibrating interfered temperature, as illustrated in Figure 2a. Finally, the sensors were connected to a signal buffer module based on INA116 (Texas Instruments, Dallas, TX, USA) and a data acquisition device NI USB DAQ 6218 (National Instrument Corporation, Austin, TX, USA). The data were collected by lab program based on LabVIEW (National Instrument Corporation, Austin, TX, USA) and then processed by the proposed models to enhance the accuracy of ISEs. The connected system is depicted in Figure 2a,b.

Table 1. Characteristics of the ion-selective electrodes (ISEs) and other sensors used in the study.

Sensor	Measurement Range	Membrane Type	Response Time (s)	Manufacturer
Nitrate ISE: REX972123	0.6–60000 mg·L ⁻¹	PVC	~50	Shanghai INESA, China
Ammonium ISE: REX 972,122	0.02–14000 mg·L ⁻¹	PVC	~50	Shanghai INESA, China
Potassium ISE: Orion 9719BNWP	0.04–39000 mg·L ⁻¹	PVC	~50	Thermo Fisher, USA
Calcium ISE: Orion 9720BNWP	0.02–40000 mg·L ⁻¹	PVC	~50	Thermo Fisher, USA
Sodium ISE: pNa 701	0.03–23000 mg·L ⁻¹	Glass	~50	Shanghai INESA, China
Chloride ISE: pCl 202	0.35–3500 mg·L ⁻¹	PVC	~50	Shanghai INESA, China
pH electrode: E-201F	2–14	-	~50	Shanghai INESA, China
EC electrode: DJS-1C	0–10000 µS	-	-	Shanghai INESA, China
Temperature probe: Pt100	0–100 °C	-	-	Yuace, China

**Figure 2.** Temperature calibrating water bath (a) and the measurement system used in this study (b).

2.1.2. Sampling Preparation

In this study, data samples for developing the models were prepared from standard solutions in the lab for training and testing purposes. The samples were collected from real hydroponic systems for validating the model performance. To mimic the complex interaction among the ions of the actual hydroponic solution, one hundred training samples ($100 = 10^2$, for 10 levels of six factors i.e., six ions, as shown in Table 2) were prepared using the fractional factorial design technique [25,38]. Furthermore, the samples at different levels ($27 = 3^3$, $36 = 6^2$, and $64 = 8^2$ samples, corresponding with three, six, and eight levels of the six considered factors) were also used to estimate the efficiency of the number of levels for the performances of the models. Additionally, two elements, dihydrogen phosphate and magnesium, were also studied by randomly changing the concentration. A water bath was used to adjust the temperature of samples randomly from 15 °C to 35 °C, corresponding with the temperature of the actual hydroponic system, to supply information for neutralizing the temperature interference.

Table 2. The ranges concentration of considered ions prepared training samples.

Ions	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10
Nitrate (mg·L ⁻¹)	44	88	177	221	332	442	553	769	1106	1328
Ammonium (mg·L ⁻¹)	6	10	15	20	25	35	45	55	75	120
Potassium (mg·L ⁻¹)	15	50	75	100	150	175	200	225	350	500
Calcium (mg·L ⁻¹)	10	25	50	75	100	125	175	225	250	350
Sodium (mg·L ⁻¹)	5	12	25	35	50	100	150	175	250	300
Chloride (mg·L ⁻¹)	5	15	35	50	80	125	175	200	300	350

The MSAM sampling method [29] was utilized to prepare the dataset for developing the models. Deionized water was mixed with a sufficient amount of stock solution of targeted ions to dilute the rinsed water (RW) in which the concentration of each ion species approximated to the lower limit value (C_0) of linearity range of the ISEs (roughly 1 mg·L⁻¹). In order to imitate the real conditions for

the training samples, a base solution (BS) was prepared by mixing the Hoagland standard solution [39] and tap water (1/1 “v/v”). A standard solution (SS) of 40 mL was prepared by mixing the appropriate amounts of the BS and the stock solutions of potassium nitrate, potassium chloride, potassium dihydrogen phosphate, magnesium sulfate, ammonium nitrate, ammonium dihydrogen phosphate, calcium nitrate, calcium chloride, sodium chloride, and sodium nitrate. This way, the concentration of the considered ions was set from the first level to the tenth level of 44–1328, 6–120, 15–500, 10–350, 5–300 and 5–350 mg·L⁻¹ for NO₃⁻, NH₄⁺, K⁺, Ca²⁺, Na⁺, and Cl⁻, respectively (Table 2). A random increase in range of 6–678 and 6–125 mg·L⁻¹ for H₂PO₄⁻ and Mg²⁺, respectively, was achieved. In the sampling procedure, first, 80 mL RW was injected into the sensor chamber. Then, the electric pump retained the solution flow to the ISEs’ surface until the stabilization of their electromotive force (EMF) values (U_0) (about 1 min). Subsequently, the SS was injected into the chamber and cycled for 3–5 min (depending upon the concentration of the sample) until the potentials (U_x) of the electrodes were stable. The stabilized EMF values were acquired by a lab program based on LabVIEW (LabVIEW 2017, National Instrument Corporation, Austin, TX, USA), and the measured values were exported to a comma-separated-values (CSV) file as the dataset to develop the models (Figure 2b). All of the models were developed by Python 3.6.2, Scikit-Learn library, SciPy, and several third-party libraries.

To evaluate the feasibility of the proposed models, 10 actual hydroponic nutrient solution samples were collected from various hydroponic systems developed for nine plant species (lettuce, perilla, purple bok choy, Chinese cabbage, strawberry, Gynura bicolor DC, amaranth, eggplant, and tomato) in the College of Information and Electrical Engineering, China Agricultural University, Beijing, China (CIEE, CAU), as shown in Table 3. The actual concentration values of the collected samples were analyzed at the Laboratory of Agricultural Informatization Standardization, Ministry of Agriculture and Rural Affairs, CAU, Beijing, China, and at the International Joint Research Center of Aerospace Biotechnology and Medical Engineering, Beihang University, Beijing, China. Furthermore, the direct calibration method (in which the array of sensors was directly immersed into the measured solutions, and the responding EMFs were recorded by sampling program) was also carried out for comparison with the proposed approach.

Table 3. The real hydroponic solution samples used to validate the models.

Sample	Grown Plant	Growing Period	Nutrient Standard	Sampling Sites
S1	Lettuce 1	Three weeks	Hoagland’s solution	Experimental Plant Factory, CIEE, CAU
S2	Perilla	Five weeks	Hoagland’s solution	Experimental Plant Factory, CIEE, CAU
S3	Lettuce 2	Four weeks	Hoagland’s solution	Experimental Plant Factory, CIEE, CAU
S4	Purple bok choy	Five weeks	Hoagland’s solution	Experimental Plant Factory, CIEE, CAU
S5	Chinese cabbage	Six weeks	Yamazaki’s solution	Experimental Plant Factory, CIEE, CAU
S6	Strawberry	Eight weeks	Hoagland’s solution	Experimental Plant Factory, CIEE, CAU
S7	Gynura bicolor DC	Five weeks	Hoagland’s solution	Experimental Plant Factory, CIEE, CAU
S8	Amaranth	Four weeks	Yamazaki’s solution	Experimental farm, CIEE, CAU
S9	Eggplant	Twelve weeks	Yamazaki’s solution	Experimental farm, CIEE, CAU
S10	Tomato	Six weeks	Yamazaki’s solution	Experimental farm, CIEE, CAU

2.2. Development of Models for Determining Multi-Ion

2.2.1. Neural Network Model

An artificial neural network (ANN) is a powerful and widely used algorithms [40]. In this study, a deep feed-forward network following the standard forward propagation and back-propagation was used to develop the regression model.

Figure 3 illustrates the simple structure of an ANN and the neuron [41]. An ANN has an input layer, which consists of x_i ($i = 1, 2, \dots, p$) independent variables (the signals from sensors), L layers of hidden k ($k = 1, 2, \dots, S$) number of neurons for each, and y_i outputs. In the feed-forward stage, the input x_i connects to each single input node, and that node transmits a weight value w_{ij} and a bias b_i to the hidden layer. Subsequently, the sum of these input-weight products is passed through an

activation function $f(z_i)$ to determine how to change the outcome. Equation (1) is the mathematical expression of the i^{th} neuron,

$$y_i = f(z_i) = f\left(\sum_i w_{ij}x_i + b_i\right) \quad (1)$$

where x_i is the input information of neuron i , w_i is the network connection weight, $f(z_i)$ is the activation function, b_i is the bias, and y_i is the output value.

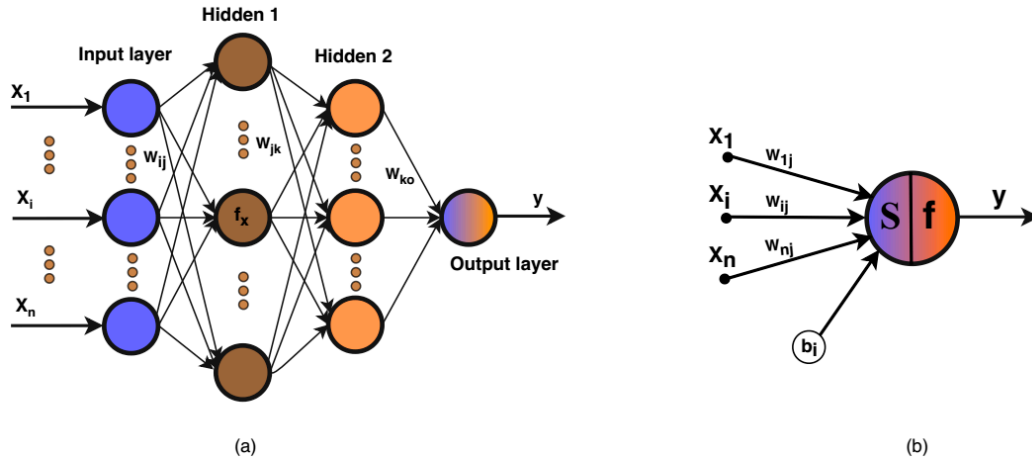


Figure 3. Neural network structure diagram (a) and the neuron (b).

The normally used activations include the sigmoid function $f(z) = \frac{1}{1+e^{-z}}$, the Tanh function $f(z) = \frac{2}{1+e^{-2z}} - 1$, and the ReLU function $f(z) = \begin{cases} 0, & z < 0 \\ z, & z \geq 0 \end{cases}$, etc.

At the back-propagation stage, the weights and bias parameters of the networks are learned by incrementally adjusting the produced values of the network approaches to the expected values from the training data. The gradient-descent algorithm is a regular optimization method. Suppose a cost function C , expression of the weights (W), the biases (B), the single input (X_i), the single modeled output (\hat{Y}_i), actual output (Y_i), i.e., $C(W, B, X_i, \hat{Y}_i, Y_i)$, represents the difference between the model's predicted outputs and the actual outputs. The derivatives of C corresponding to each weight and bias value in each layer of the network could be determined using the chain rule, such as

$$\frac{\partial C}{\partial w_{ij}^{(L)}} = \frac{\partial C}{\partial z_j^{(L)}} \cdot \frac{\partial z_j^{(L)}}{\partial w_{ij}^{(L)}} \quad (2)$$

$$\frac{\partial C}{\partial b_j^{(L)}} = \frac{\partial C}{\partial z_j^{(L)}} \cdot \frac{\partial z_j^{(L)}}{\partial b_j^{(L)}} \quad (3)$$

Then, the weights and biases are updated by stochastic gradient descent optimization method,

$$w_{ij} \leftarrow w_{ij} - \gamma \frac{\partial C}{\partial w_{ij}^{(L)}} \quad (4)$$

$$b_j \leftarrow b_j - \gamma \frac{\partial C}{\partial b_j^{(L)}} \quad (5)$$

where γ is the learning rate.

2.2.2. Gaussian Process Model

A Gaussian process (GP) is a stochastic process $f(x)$ characterized by its mean function $\mu(x)$ and covariance function $k(x, x')$ [42]. In the Gaussian process regression task, consider a set of data D consisting of N input vectors $X = x_1, x_2, \dots, x_N$ (signals of dimension D from sensors) is a multivariate Gaussian distribution with corresponding continuous outputs $y = y_1, y_2, \dots, y_N$ (ion concentration of samples). In a Gaussian process regression (GPR) model, an output is assumed to be noisily observed from an underlying functional mapping $f(x)$. Then with the set of data $D = \{X^{[i]}, y^{[i]}\}$:

$$y^{[i]} = f(x^{[i]}) + \varepsilon^{[i]} \quad (6)$$

where $i = 1, 2, \dots, n$, $\varepsilon = N(0, \sigma_f^2)$ is the additive independent Gaussian noise with mean 0 and variance σ_f^2 . The collection of function values f will have a joint Gaussian distribution if $f(x)$ is a GP.

$$f(X) = [f(x_1), f(x_2), \dots, f(x_n)]^T \sim N(\mu, \Sigma) \quad (7)$$

where mean vector $\mu = \mu(x_i)$ and covariance matrix $\Sigma = k(x_i, x_j)$, determined by the mean function and kernel function of the GP.

Normally, the mean function of the GP is assumed to be zero. Thus, the relation from one to the other is only the covariance function $k(x, x')$. A popular kernel function is the radial basic function (RBF; also called squared exponential), given as follows:

$$k(x, \hat{x}) = \exp\left[\frac{-(x - x')^2}{2l^2}\right] \quad (8)$$

where l is length-scale, which quantifies the level of local smoothness level of the drawn Gaussian process distribution.

Consider another dataset $D_* = \{x_*^{[i]}, y_*^{[i]}\}, i = 1, 2, \dots, n_*$, which refers to the testing set, has the same distribution as the set D , then defined by the Gaussian process [42], we have

$$\begin{bmatrix} f(X) \\ f(X_*) \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \quad (9)$$

where $K(X, X)$, $K(X_*, X_*)$, and $K(X_*, X)$ are covariance matrices evaluated at training locations X , testing locations X_* , and locations X_* and X , respectively.. Then, given the additive Gaussian noises, and using the rules for conditioning Gaussian distributions, the posterior distribution of the GP is [42]

$$f(X_*)|y(X) \sim N(\mu_*, \Sigma_*) \quad (10)$$

where the posterior mean $\mu_* = m(X_*) + K(X_*, X)[K(X, X) + \sigma^2 I]^{-1}(y(X) - m(X))$ and the posterior variance $\Sigma_* = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma^2 I]^{-1}K(X, X_*)$, $m(X_*)$ and $m(X)$ are mean vectors calculated at X_* and X , I is the identity matrix.

The predictive distribution of the GP is based on Equation (10). The kernel learning algorithm is carried out by maximizing the log marginal likelihood of the targets y . The probability of the data conditioned only on kernel parameters θ is given as

$$\log p(Y|\theta, X) \propto -Y^T(K_\theta(X, X) + \sigma^2 I)^{-1}Y - \log|K_\theta(X, X) + \sigma^2 I| \quad (11)$$

2.2.3. Deep Kernel Learning Model

Recently, deep kernel learning (DKL), a combination of the deep learning structure and kernel methods, was proposed as an elegant and flexible algorithm [35]. In this study, the multi-ion

concentration of the hydroponic nutrient solution was determined by combining the MSAM technique [29] with DKL signal processing to enhance the accuracies and feasibilities of the ISEs, as shown in Figure 1. The collected signals from sensor array extracted more significant information by using the feature enrichment technique. In this manner, eight signals from six ISEs, pH, and EC probe were enriched to 16 data features composing of eight original features and eight feature enrichment (FE) data. The principle of the technique is illustrated by MSAM feature enrichment component in Figure 1 and Equation (12).

$$U_{FEij} = U_{xij} - U_{0ij} \quad (12)$$

where i is the number of samples (1–100), j is the number of ion-selective electrodes (1–8), and U_{0ij} , U_{xij} are the potential of corresponding ISE at concentration C_0 and C_x respectively. U_{FEij} , “enriched data value,” represents the values of differences between U_{0ij} and U_{xij} , which were used to improve the performance of the model. The details of this technique were reported by Tuan et al. [29]. The 17-dimensional input vector (16 enriched features plus one temperature sensor) was then introduced to the DKL, as shown in Figure 1.

The data of the DKL model were first propagated by the neural network in the forward stage. The high-dimensional MSAM-FE data (17 dimensions) were transformed into the lower-dimensional feature vector, which was suitable for the input arguments of Gaussian process regression. The expected values of the concerned ions were predicted by DKL, relying on the posterior distribution as a function of the input data. A Gaussian process is equivalent to a Bayesian neural network that has an infinite number of nodes [43]. Therefore, the ended Gaussian process layer of the DKL structure could be considered as an infinite number of nodes hidden in the deep neural network layer. This architecture greatly increases the expression ability of the network compared to a stand-alone deep neural network. Nevertheless, the Gaussian processes naturally do not fit well with the high-dimensional input–output spaces. The additive deep neural network acts as a feature extractor and dimensionality reduction method, which compensates more robustly for the Gaussian process regression.

The DKL could be viewed as a Gaussian process with a stand-alone deep kernel [35]. Therefore, the DKL could be constructed from a base kernel $k(x^{[i]}, x^{[j]} | \theta)$ with kernel parameters θ , as follows:

$$k(x^{[i]}, x^{[j]} | \theta) \rightarrow k(g(x^{[i]}, w), g(x^{[j]}, w) | \theta, w) \quad (13)$$

where $g(x, w)$ is a non-linear mapping performed by the neural network, w is weight parameter. The kernel is the core of DKL in this study. The DKL model was conducted with the RBF, Dotproduct, and the spectral mixture kernel [35], as per Wilson and Adams [44].

$$k_{SM}(x, x' | \theta) = \sum_{q=1}^Q a_q \frac{|\Sigma_q|^{\frac{1}{2}}}{(2\pi)^{\frac{D}{2}}} \exp\left(-\frac{1}{2} \left\| \sum_q^{\frac{1}{2}} (x - x') \right\|^2\right) \cos\langle x - x', 2\pi\mu_q \rangle \quad (14)$$

where the learnable kernel parameters $\theta = \{a_q, \Sigma_q, \mu_q\}$ include a weight, an inverse length scale, and a frequency vector for each of the Q spectral components.

Constructing the DKL involves optimizing learnable parameters including network weights and kernel parameters and also tuning hyper-parameters such as the learning rate, number of iterations, and number of nodes (neurons) in each layer of the neural network. Before training could be carried out, suitable values of the hyper-parameters were specified. We executed this based on cross-validation over a small hyper-parameter search-space and thus employed a systematic optimization-based procedure. In this manner, we determined the major hyper-parameters, such as the number of nodes in each hidden layer, the prior white-noise level of the Gaussian process, and the number of epochs, i.e., training iterations and the learning rate. The parameters that were included in the model are listed in Table 4.

The parameters of the deep kernel learning (including neural network parameters w , and Gaussian process kernel parameter θ) were learned by maximizing the log posterior marginal likelihood applying

Equation (11), with respect to $\gamma = \{w, \theta\}$. The derivatives concerning the weight variables $\frac{\partial g(x, w)}{\partial w}$ were computed using the standard back-propagation algorithm. To avoid local minima and overfitting, the dropout regularization algorithm was used for training the network [45] with the adjusted dropout rate of 0.5 to 0.99.

Table 4. The structure of the DKL for predicting multi-ion concentration.

Parameters	Values
Number of hidden layers	1, 2, 3, 4, 5, 6
Hidden layer size	1 to 1000
Hidden layer transfer function f(x)	tansig, logsig, linear, ReLU
Output layer transfer function	ReLU
Optimization algorithm	Stochastic gradient descent (SGD), Broyden–Fletcher–Goldfarb–Shanno (BFGS), Adam
Dropout rate	0.5 to 0.99
Learning rate	0.001 to 0.1
Max number of epochs	1000
Prior whitenoise level	0.001 to 1
Kernel	Radial basic function (RBF), Dotproduct, Spectral mixture (SM)
Training goal	10^{-6}

2.2.4. Model Performance Metrics

To evaluate the efficiency of the proposed models for determining multi-ion concentration in hydroponic solution, three performance indices were estimated, i.e., the performance coefficient (R^2), the root mean square error (RMSE), and the coefficient of variation (CV). The smaller the values of RMSE, the closer the predicted values are to the true values, which means better prediction accuracy. The closer the R^2 value to unity, the better the machine learning prediction is. Moreover, the CV represents the precision of the model performance. CV is negatively correlated with the accuracy of predicted values. The RMSE is given as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

where n is the total number of data in the training set or test set, y_i is the actual ion concentration value, and \hat{y}_i is the predicted ion concentration value.

R^2 is an index that measures the degree of agreement between the test data and the fitting function. It is presented as

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (\bar{y}_i - \hat{y}_i)^2} \quad (16)$$

where \bar{y}_i is the average of the test set.

The CV is evaluated by the following expression:

$$CV = \frac{SD}{\bar{y}_N} \cdot 100 = \frac{\sqrt{\frac{\sum_i^N (\hat{y}_i - \bar{y}_s)^2}{N-1}}}{\bar{y}_N} \cdot 100 \quad (17)$$

where SD is the standard deviation, the average concentration estimated by the model for each sample \bar{y}_s , the average concentration of N measurements \bar{y}_N , and N is the number of sample measurements.

3. Results

3.1. Responses of the Ion Selective Electrodes

As mentioned above, the ISEs were calibrated by two methods: the direct calibration method (DCM) [46], and the multivariate standard additional method (MSAM). The Nikolsky–Eisenman

calibration equation concerning the logarithmic relationship between the concentrations of the standard samples (x) and EMFs of electrodes (y) is summarized in Table 5. The results showed that the determination coefficient (R^2) of the ISEs using the DCM was 0.90 to 0.95. In this MSAM-based approach, the R^2 significantly improved from 0.92 to 0.97. However, these achieved improvements still did not fulfill the requirements of multi-ion measurement in the actual hydroponic solution. Therefore, the models were deployed to improve the efficiency of electrodes for quantifying ions (nitrate, ammonium, potassium, calcium, sodium, and chloride) simultaneously. Furthermore, the models also fused the data collected from available ISEs to predict two ions (phosphate and magnesium), which had no feasible ISEs.

Table 5. The responses of the ISEs calibrated by the direct calibration method (DCM) and MSAM technique.

ISEs	DCM		MSAM	
	R^2	Calibrating Equation	R^2	Calibrating Equation
Nitrate	0.93	$y = -22.04\ln(x) + 202.62$	0.95	$y = -22.86\ln(x) + 208.47$
Ammonium	0.90	$y = 22.856\ln(x) - 255.01$	0.92	$y = 22.92\ln(x) - 253.87$
Potassium	0.95	$y = 23.39\ln(x) - 240.07$	0.97	$y = 23.07\ln(x) - 237.03$
Calcium	0.94	$y = 11.419\ln(x) - 79.31$	0.96	$y = 11.06\ln(x) - 76.90$
Sodium	0.91	$y = 18.227\ln(x) - 178.31$	0.93	$y = 19.76\ln(x) - 186.22$
Chloride	0.94	$y = -23.28\ln(x) + 186.54$	0.96	$y = -23.02\ln(x) + 192.33$

3.2. Determination of the DKL Architecture

The structure of DKL developed with the main network parameters and hyper-parameters (Table 4) was subjected to various trials to determine the best fit. The best DKL model architecture consisted of a five-layer neural network and a Gaussian process with RBF kernel. The Tanh and ReLU activation functions were applied to construct the hidden layer of the ANN stage. To avoid local minima and overfitting problems, the dropout algorithm was used for training the network. We adopted the standard root mean squared error (RMSE) loss function and the Adam optimizer for tuning the model [47]. Furthermore, to evaluate the most effective model for determining the concentrations of multiple ion species in the hydroponic nutrient solution, the ANN (having the same architecture as the ANN stage of the DKL) and the GP models were tested as well. The fitted DKL architecture was configured as in Table 6.

Table 6. The structures of the fitted DKL model for predicting multi-ion concentration.

Layer 1			Layer 2			Layer 3			Layer 4			Layer 5			Opt	LR	N.o.E	KF
N.o.N	AF	DR	N.o.N	AF	DR	N.o.N	AF	DR	N.o.N	AF	DR	N.o.N	AF	DR				
580	Tanh	0.99	580	Tanh	0.99	100	ReLU	0.99	100	ReLU	0.99	8	ReLU	0.99	Adam	0.005	250	RBF

N.o.N: number of nodes, AF: activation function, DR: dropout rate, Opt: optimizer, LR: learning rate, N.o.E: number of epochs, KF: kernel function.

The prediction RMSEs of the DKL, according to the number of epochs, are illustrated in Figure 4a. The DKL was converged at 250 epochs, as most of the ion predictions achieved the lowest errors RMSEs. The number of nodes of the last hidden layer was determined by relying upon the number of considered targets (target ions) and the coefficient of principal component (PC) analyzed from the high dimensional data of the MSAM-FE. Choosing a suitable degree of PC from the raw data of ISEs eliminates the noise [48], and improves the efficiency of GP performance. The ratio of PCs and variance of data is shown in Figure 4b. The first eight PCs contained roughly 98% of data variances. As per a previous study, there are about 3% system errors and noise in the ISEs based measurement system [49]. Therefore, the number of PCs could be chosen as eight and the last three PCs removed whilst losing roughly 2% of data variances.

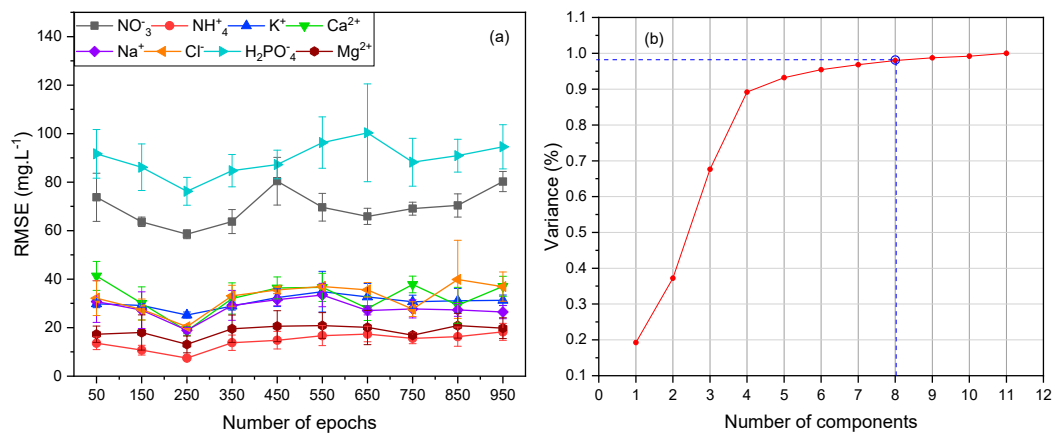


Figure 4. Relationship between the number of epochs and RMSEs (a) and PC components and variance of the dataset (b).

3.3. Evaluation of the Performance of Proposed Models

In this work, three models were conducted using Python language and several supported libraries. Normalized data allows the model to learn rapidly. Therefore, the raw dataset was preprocessed first in the range of -1 to 1 using MinMaxScaler of the Scikit Learn library. In the training stage, the k-fold cross-validation algorithm ($k = 10$) was used to evaluate RMSE and R^2 based on the metrics to find out the fitted model for determining the multi-ion. The performance results of the three models are presented in Figure 5. The predicted nitrate is shown in Figure 5a. The results of the ANN and the GP showed a relatively linear and accurate prediction with R^2 of 0.95 and 0.94, a slope of 0.95 and 0.94, and the RMSEs of $91.5 \text{ mg}\cdot\text{L}^{-1}$ and $102.7 \text{ mg}\cdot\text{L}^{-1}$, respectively. In the DKL case, the results were better than those of ANN and GP with RMSE of $58.5 \text{ mg}\cdot\text{L}^{-1}$. Specifically, the highly linear relationship (having an R^2 of 0.98 and a slope of 1.01) revealed that the training of the DKL was achieved well.

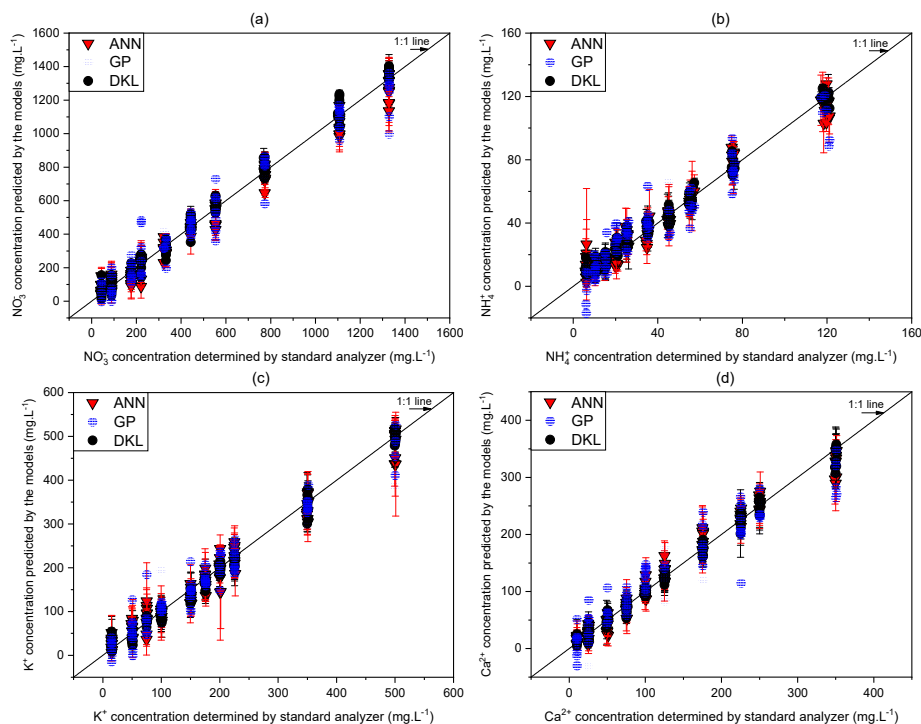


Figure 5. Cont.

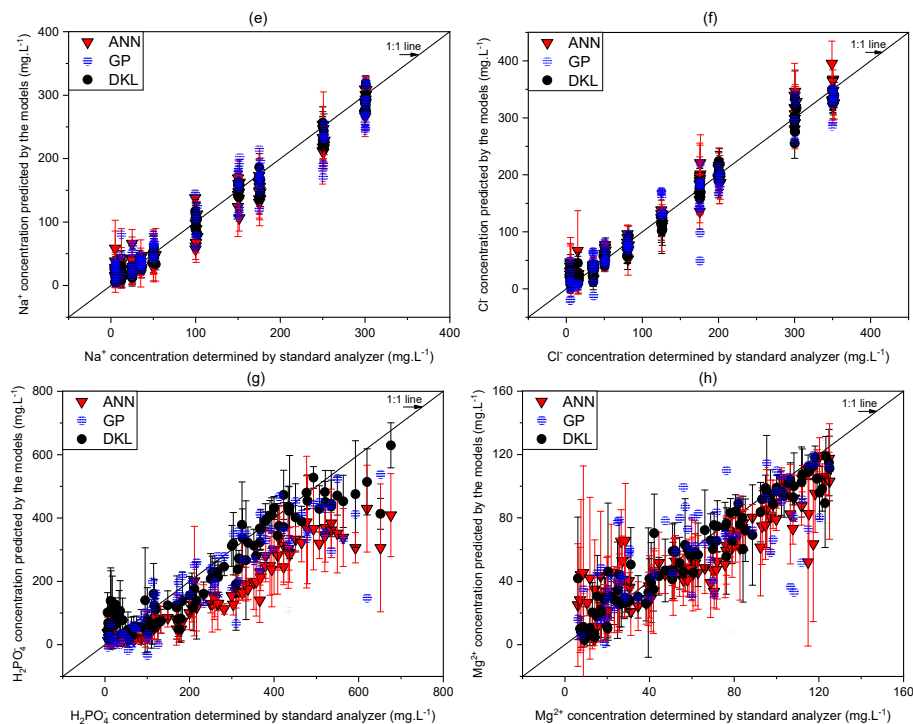


Figure 5. Relationships between predicted ion concentrations of the models and standard analyzers. (a) NO_3^- , (b) NH_4^+ , (c) K^+ , (d) Ca^{2+} , (e) Na^+ , (f) Cl^- , (g) H_2PO_4^- , and (h) Mg^{2+} . Error bars indicate standard deviations of three replicates.

For the ammonium prediction (Figure 5b), the performance of ANN and GP were slightly low (RMSE of $10.9 \text{ mg}\cdot\text{L}^{-1}$ and $13.1 \text{ mg}\cdot\text{L}^{-1}$, R^2 of 0.92 and 0.90, and slope of 0.92 and 0.90 for ANN and GP, respectively). Although the performance of DKL was better than those of the ANN and GP models, it just archived with an RMSE of $7.4 \text{ mg}\cdot\text{L}^{-1}$, an R^2 , and a slope of 0.95. The imperfect prediction results of ammonium might be caused by strong interferences of other ions in the hydroponic solution samples.

The GP model was effective for the prediction of potassium (Figure 5c) with an RMSE of $31.2 \text{ mg}\cdot\text{L}^{-1}$, an R^2 of 0.96, and a slope of 0.95, providing a better performance than that of the ANN model. The DKL model exhibited the best performance with an RMSE of $25.2 \text{ mg}\cdot\text{L}^{-1}$, a slope of 0.99, and an R^2 of 0.978.

Calcium prediction was more stable and linear for ANN than for GP (Figure 5d). Specifically, the GP showed a slightly low linear relationship (RMSE $35.3 \text{ mg}\cdot\text{L}^{-1}$, R^2 0.92, and slope 0.85), which was relatively lower than those of the ANN (RMSE $23.6 \text{ mg}\cdot\text{L}^{-1}$, R^2 0.96, and slope 0.98) and the DKL (RMSE $18.8 \text{ mg}\cdot\text{L}^{-1}$, R^2 0.97, and slope 0.99) models.

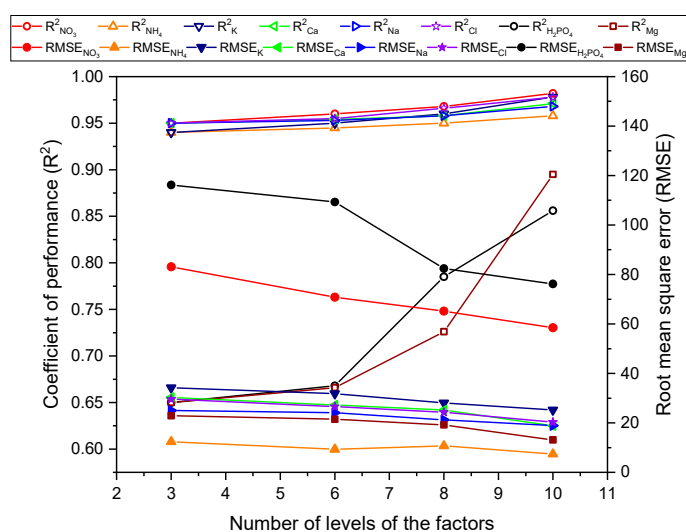
The sodium and chloride prediction results (Figure 5e,f) showed similar trends in terms of RMSEs, R^2 , and slopes. The performance of the DKL was the best, with the RMSE of 18.9 and $20.3 \text{ mg}\cdot\text{L}^{-1}$, the R^2 of 0.96 and 0.97, and the slope of 0.97 and 0.99 for sodium and chloride, respectively. In general, most of the R^2 and slopes of the DKL model were relatively high (0.95 to 0.99), which showed that the training stage of the models was at the acceptable levels [50].

The phosphate and magnesium prediction results (Figure 5g,h) showed that the ANN performance was slightly preferable to GP. However, neither of the models provided expected results. The RMSEs of ANN and GP models were $122.5 \text{ mg}\cdot\text{L}^{-1}$ and $135.8 \text{ mg}\cdot\text{L}^{-1}$, respectively, for phosphate and $21.3 \text{ mg}\cdot\text{L}^{-1}$ and $25.2 \text{ mg}\cdot\text{L}^{-1}$, respectively, for magnesium. Moreover, unstable results were observed for both models. Conversely, the DKL provided relatively satisfying result, exhibiting $76.2 \text{ mg}\cdot\text{L}^{-1}$ RMSE, $0.86 R^2$ and 0.85 slope for phosphate and $13.1 \text{ mg}\cdot\text{L}^{-1}$ RMSE, $0.89 R^2$, and 0.88 slope for magnesium. The details of correlation values of the actual concentrations versus the predicted concentrations of the models are summarized in Table 7.

Table 7. The correlation of the predicted concentrations (y) with the actual values (x).

Species	Models	Predicting Equation	RMSE (mg·L ⁻¹)	Coefficient of Performance (R ²)
Nitrate	ANN	$y = 0.95x + 18.11$	91.5	0.95
	GP	$y = 0.94x - 10.93$	102.7	0.94
	DKL	$y = 1.01x + 17.81$	58.5	0.98
Ammonium	ANN	$y = 0.92x + 4.54$	10.9	0.92
	GP	$y = 0.90x + 3.80$	13.1	0.90
	DKL	$y = 0.95x + 5.13$	7.4	0.95
Potassium	ANN	$y = 0.94x + 7.33$	33.5	0.95
	GP	$y = 0.95x + 11.50$	31.2	0.96
	DKL	$y = 0.99x + 5.59$	25.2	0.978
Calcium	ANN	$y = 0.98x + 4.20$	23.6	0.96
	GP	$y = 0.85x + 22.30$	35.3	0.92
	DKL	$y = 0.99x + 5.27$	18.8	0.97
Sodium	ANN	$y = 0.94x - 1.97$	22.5	0.94
	GP	$y = 0.86x + 14.11$	29.3	0.92
	DKL	$y = 0.97x + 1.08$	18.9	0.96
Chloride	ANN	$y = 0.97x + 1.22$	25.0	0.95
	GP	$y = 0.95x + 5.67$	27.2	0.94
	DKL	$y = 0.99x + 2.68$	20.3	0.97
Phosphate	ANN	$y = 0.71x + 31.59$	122.5	0.76
	GP	$y = 0.62x + 20.27$	135.8	0.61
	DKL	$y = 0.85x + 17.82$	76.2	0.86
Magnesium	ANN	$y = 0.82x + 11.43$	21.3	0.75
	GP	$y = 0.63x + 10.64$	25.2	0.62
	DKL	$y = 0.88x + 2.11$	13.1	0.89

Furthermore, the DKL was also trained with three smaller datasets (27, 36, and 64 samples corresponding to 3, 6, and 8 levels of six factors, respectively) to estimate which sample size was appropriate. The results (Figure 6) showed that the ISEs prediction of ions was slightly improved having the RMSEs reduced from 5 mg·L⁻¹ (minimum) to 25.6 mg·L⁻¹ (maximum), and the gain in R² was from 0.018 to 0.032 for ammonium and nitrate, respectively. Particularly, the predictions of two unavailable ISEs ions had considerable improvement at the tenth level. The RMSEs reduced to 40.1 mg·L⁻¹ and 9.8 mg·L⁻¹, and the R² increased to 0.205 and 0.245 for determining phosphate and magnesium, respectively.

**Figure 6.** The performance and efficiency of the DKL model with varying number of levels of factors.

3.4. Validation of the Proposed Models with Real Hydroponic Samples

After training and cross-validating with the laboratory-made samples, the applicability of the models for determining multi-ion was validated with real hydroponic samples. The relationship of the ion concentration predicted by the models and the standard analyzers is shown in Figure 7. In most sample measurements, the ANN predictions were more accurate than those of the GP. The ion concentrations predicted by the DKL model were closer to the actual concentrations than those of ANN and GP, which indicated that the DKL model improved the accuracy of the sensor array by processing the signals effectively.

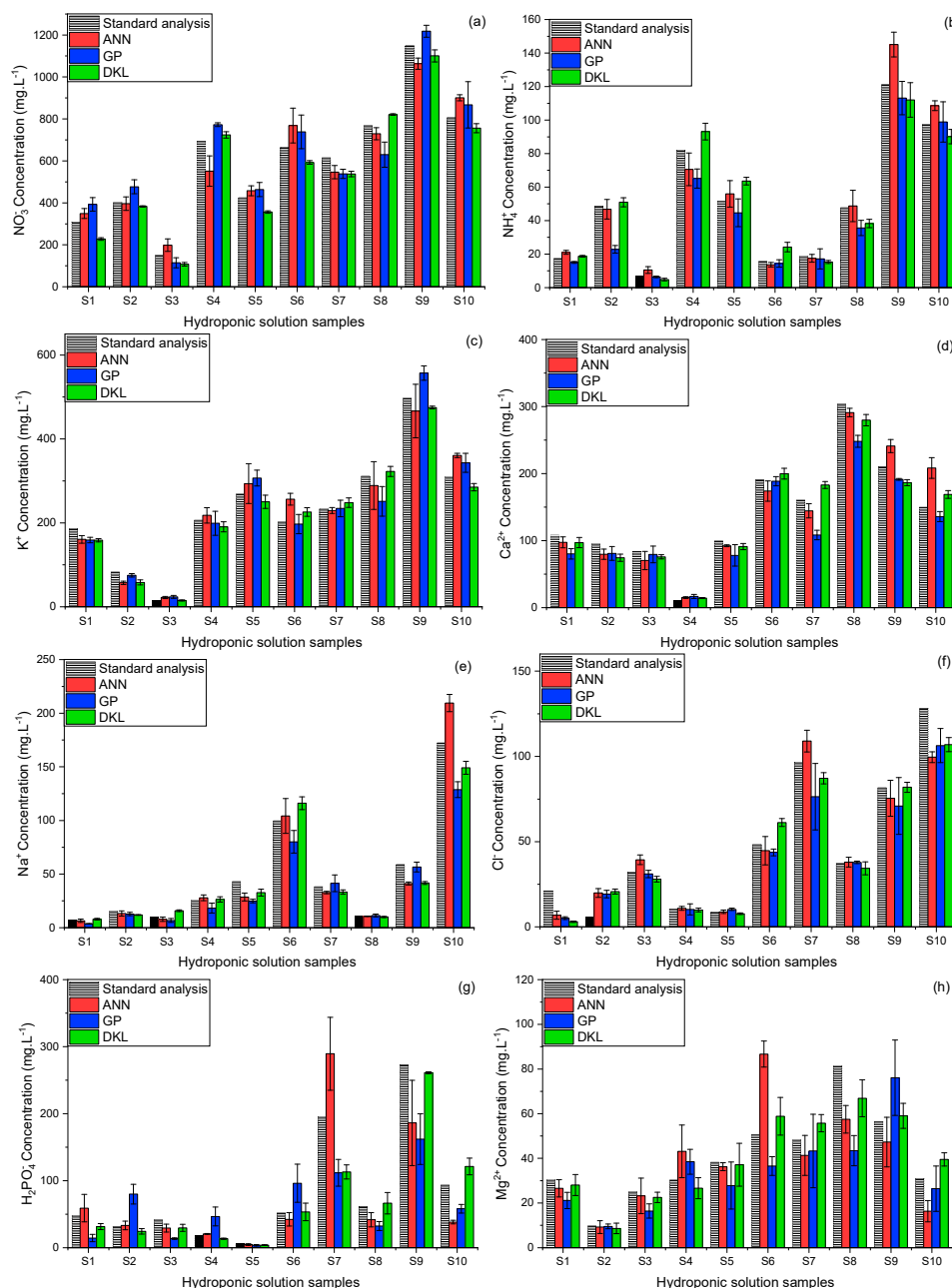


Figure 7. Comparison of the actual concentrations with the predicted concentrations determined by proposed models using ten different hydroponic samples (a) NO_3^- , (b) NH_4^+ , (c) K^+ , (d) Ca^{2+} , (e) Na^+ , (f) Cl^- , (g) H_2PO_4^- , and (h) Mg^{2+} . Error bars indicate standard deviations of three replicates.

Table 8 summarizes the RMSEs obtained from the three models. In potassium prediction, the RMSE of GP ($35.2 \text{ mg}\cdot\text{L}^{-1}$) was lower than that of the ANN ($36.3 \text{ mg}\cdot\text{L}^{-1}$). In most of the cases, the DKL model achieved the best predictability with the smallest RMSEs and CVs below 8%. In the prediction of phosphate and magnesium, even though the error bars showed relatively high CVs (13.9% and 14.8%) for DKL method, the prediction results were almost comparable to the actual values. This implies that the DKL model could be potentially deployed for developing multi-ion sensor for sensing phosphate and magnesium in hydroponic nutrient solutions.

Table 8. Comparison of the predicted quality of the proposed models with the real hydroponic solution tests.

Considered Ions	Range of Concentration ($\text{mg}\cdot\text{L}^{-1}$)	Models	Accuracy (RMSE, $\text{mg}\cdot\text{L}^{-1}$)	Precision (CV, %)
Nitrate	150–1150	ANN	83.8	7.2
		GP	86.1	8.4
		DKL	63.8	3.5
Ammonium	6–120	ANN	10.3	9.2
		GP	12.2	10.3
		DKL	8.3	7.0
Potassium	15–500	ANN	36.3	9.2
		GP	35.2	8.8
		DKL	29.2	5.4
Calcium	10–305	ANN	25.2	7.7
		GP	29.1	9.6
		DKL	18.5	5.5
Sodium	6–175	ANN	14.8	9.2
		GP	17.1	9.9
		DKL	11.8	6.8
Chloride	1.6–128	ANN	11.7	9.5
		GP	12.8	10.5
		DKL	8.8	6.9
Phosphate	5–275	ANN	50.5	22.3
		GP	55.8	23.6
		DKL	29.6	13.9
Magnesium	10–80	ANN	16.9	21.3
		GP	18.1	23.5
		DKL	8.7	14.8

4. Discussion

We proposed a combination of the multivariate standard addition sampling technique and three machine learning models to develop an architecture for effectively determining the concentrations of multiple ion species in a hydroponic nutrient solution. The aim was to compensate the potential drifts, interferences from secondary ions and temperature, and ionic strength problems of the ISEs.

The results of the training stage (Figure 5 and Table 7) showed that the GP-based structure resolved most of the fundamental problems of ISEs array with relatively high linear relationship ($R^2 = 0.96$) and low RMSE ($31.2 \text{ mg}\cdot\text{L}^{-1}$) for potassium prediction. The positive results with GP may be attributed to the exceptional processing ability of the GP in the scarce dataset (limited dataset) [51]. However, for predicting other ions (ammonium, calcium, and sodium), the GP responses were not adequate, although the MSAM technique had eliminated some issues of ISEs, as shown in Table 5 (Section 3.1). These unsatisfactory results may be due to the interferences of other ions that affected the Nernstian slope [16,52]. Another concern was the inherent weakness of GP [53] faced with the complex high dimensional input-output (i.e., 17-dimensional input signal, and eight outputs) of the multi-ion sensing in the hydroponic system. The ANN model was a suitable model to overcome interference problems. This was demonstrated by the predicted results of ANN, which were better than those of GP in case of several ions (Table 7). Having the ability of non-linear processing and high effective generalization, ANN may resolve the interferences [17,19]. Nevertheless, in coping with complex problems such as multi-ion interferences of the hydroponic solution, the ANN had some shortcomings that were not

completely rectified. The ANN predicted results of most ions (Figure 5 and Table 7) were comparable with those of previous studies [19]. However, the ANN responses were still slightly unstable and had moderate error bar levels. Therefore, the previous studies normally utilized ANN to overcome one pair of interference of ions [17,54] or some ions [19]. To process the high dimensional data efficiently, the models need to be fed with the dataset having hundreds of samples or more [55] to obtain accurate predictions. For determining multi-ion (eight ions) simultaneously, it is difficult to prepare the large dataset. However, small datasets with 27 samples, for example [19], may not be appropriate because the high number of dimensions and small size of the dataset compromises the robustness of ANN. This limitation motivated researchers to innovate by combining ANN with other techniques, e.g., PCA-ANN, ICA-ANN [48] to improve its efficiency. Thus, the DKL was acquired to solve the problems of the ISEs based sensing system. As shown in Figure 5 and Table 7, the results of DKL were better than those of ANN and GP models in all the cases. The model provided the highest relationship ($R^2 = 0.98$), slope (1.01), and the lowest RMSE of 58.5, 7.4, 25.2, 18.8, 18.9, 20.3, 76.2, and 13.1 $\text{mg}\cdot\text{L}^{-1}$ for predicting nitrate, ammonium, potassium, calcium, sodium, chloride, phosphate, and magnesium, respectively.

In the real hydroponic samples test, the prediction results of the models (Figure 7 and Table 8) tended to emulate those of the training stage. The results of the GP model in potassium prediction (35.2 $\text{mg}\cdot\text{L}^{-1}$ RMSE, and 8.8% CV) were better than those of ANN. However, in most of the remaining cases, the ANN predictions were better than those of GP. For a detailed comparison, the RMSE results of both models were higher than those of the DKL model. This may be due to the differences in background ions and the drifts of ISEs' signals. However, the differences were rather small, and the CVs of most cases were less than 10%. It proves that the MSAM is a simple and effective sampling technique for multi-ion sensing in the hydroponic system that could considerably restrict the drifts, changing background ions, and ionic strength effects. The best results of the DKL model in both real hydroponic solution and training test may be created by compensating the advantages of both ANN and GP together [35]. In this manner, the role ANN stage in DKL was analogous with a simple data mining and preprocessing stage. The high-dimensional dataset of multi-ion sensor array was reduced by the non-linear projection of ANN. The number of nodes in the last hidden layer was chosen following the relationship between the components and variance of the data (as shown in Figure 4b) and the number of output targets for dimensional reduction purposes. The reduced-dimension dataset was then introduced to the last GP stage of DKL to mitigate uncertainties and allow for accurate predictions. The fitted parameters and hyper-parameters of two halves of DKL structure revealed the robustness of DKL for better prediction results with CVs below 8% and the lowest RMSE of 63.8, 8.3, 29.2, 18.5, 11.8, and 8.8 $\text{mg}\cdot\text{L}^{-1}$, for the prediction of nitrate, ammonium, potassium, calcium, sodium, and chloride, respectively.

In phosphate and magnesium predictions, the generalized and expressive abilities of the DKL model produced significant differences in results compared with those of both ANN and GP models. The DKL model achieved RMSE of 29.6 $\text{mg}\cdot\text{L}^{-1}$ and 8.7 $\text{mg}\cdot\text{L}^{-1}$ and CVs below 15% for prediction of phosphate and magnesium, respectively (see Table 8). Additionally, even with the smallest number of samples (27 samples), the prediction results of the DKL model were relatively consistent with the actual concentrations (Figure 7). The results show that the MSAM sampling and the feature enrichment technique provided more useful information to models, which would improve the efficiency of prediction and inferences. Moreover, the results proved that the DKL model successfully fused the data of ISEs to retrieve the concentrations of the two unavailable ISEs elements. The predicted effects of phosphate and magnesium were not as high as those of other available ISEs ions. Nevertheless, in conditions where sensors are lacking, the proposed approach could be used as a foundation to develop phosphate and magnesium diagnosing tool for the closed hydroponic system.

5. Conclusions

This study proposed a combination of the MSAM-FE technique and the DKL model for developing a sensing architecture that could prevent the adverse effects on ISEs, such as signal drifts, interferences,

and the ionic strength for reliably determining eight ions in the hydroponic nutrient solution. The parameters and hyper-parameters of the models were trained by 100 imitated hydroponic background samples and validated by 10 samples collected from several real hydroponic systems.

The results showed that the ANN and GP models did not accomplish high satisfaction as expected because of the effects of interferences, even though they were supported by the MSAM sampling technique for minimizing drifts and ionic strength. The combined MSAM-FE-DKL model enhanced the reliability of the multi-ion sensing with the RMSE of 63.8, 8.3, 29.2, 18.5, 11.8, 8.8, 29.6, and 8.7 mg·L⁻¹ and the CVs below 8% for predictions of nitrate, ammonium, potassium, calcium, sodium, chloride, phosphate, and magnesium, respectively. Specifically, in phosphate and magnesium predictions, the DKL-based structure exhibited many desirable outcomes, and the results were akin to those of the actual ISEs. These proved that the MSAM-FE-DKL sensing architecture can be used as a soft phosphate and magnesium sensing tool for the multi-ion testing tasks.

The proposed approach enabled us to quantify eight essential ions simultaneously in hydroponic solution with satisfying results and a feasible structure, suggesting that the proposed multi-ion sensing architecture could be applied for improving the quality and efficiency of closed hydroponic systems. The study also paved the way for effectively measuring and controlling individual ion tasks in a hydroponic system. Although DKL provided several promising results, it still has some limitations, such as computational complexity and storage of GP core [56]. This limitation may make DKL difficult to deploy in the big data system. However, in this study, this problem was alleviated by reducing the dimension of data (from seventeen to eight) and utilizing the finite samples (one hundred). Thus, the computational burden hinders no more with the application of DKL in this scenario. Moreover, the DKL model was combined with a relatively simple sampling technique. This approach has feasible application in the closed hydroponic system. Nevertheless, deploying MSAM-FE-DKL in commercial hydroponics needs further research to validate the efficiency of DKL via developing fertigation through the hydroponic solution. In future, DKL could be combined with semi-supervised or unsupervised learning techniques for effectively adopting ISEs into commercial hydroponic systems.

Author Contributions: Conceptualization, V.N.T. and W.G.; methodology, V.N.T. and W.G.; software, V.N.T. and W.G.; validation, V.N.T. and M.W.; formal analysis, V.N.T., W.G., and M.W.; investigation, V.N.T. and H.Z.; resources, V.N.T.; data curation, V.N.T., A.M.K., and M.W.; writing-original draft preparation, V.N.T.; writing-review and editing, V.N.T., A.M.K., and M.W.; visualization, V.N.T., A.M.K., and H.Z.; supervision, W.G., and M.W.; project administration, W.G.; funding acquisition, W.G., and M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Liquor Making Biological Technology and Application of Key laboratory of Sichuan Province (Grant No. NJ2019-02).

Acknowledgments: The authors would like to express their sincere gratitude to Guo Yanbin, College of Information and Electrical Engineering VR and Simulation Laboratory Center, China Agricultural University, Beijing, China for donations of implementing the sensor chamber, And the teachers and students of International Joint Research Center of Aerospace Biotechnology and Medical Engineering, Beihang University, Beijing, China, for analyzing the hydroponic samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Palermo, M.; Paradiso, R.; De Pascale, S.; Fogliano, V. Hydroponic Cultivation Improves the Nutritional Quality of Soybean and Its Products. *J. Agric. Food Chem.* **2012**, *60*, 250–255. [\[CrossRef\]](#)
2. Despommier, D. *The Vertical Farm: Feeding the World in the 21st Century*; Macmillan: New York, NY, USA, 2010.
3. Jones, J.B., Jr. *Complete Guide for Growing Plants Hydroponically*; CRC Press: New York, NY, USA, 2014.
4. Hosseinzadeh, S.; Verheust, Y.; Bonarrigo, G.; Van Hulle, S. Closed hydroponic systems: Operational parameters, root exudates occurrence and related water treatment. *Rev. Environ. Sci. Bio-Technol.* **2017**, *16*, 59–79. [\[CrossRef\]](#)
5. Bamsey, M.; Graham, T.; Thompson, C.; Berinstain, A.; Scott, A.; Dixon, M. Ion-specific nutrient management in closed systems: The necessity for ion-selective sensors in terrestrial and space-based agriculture and water management systems. *Sensors* **2012**, *12*, 13349–13392. [\[CrossRef\]](#)

6. Rius-Ruiz, F.X.; Andrade, F.J.; Riu, J.; Rius, F.X. Computer-operated analytical platform for the determination of nutrients in hydroponic systems. *Food Chem.* **2014**, *147*, 92–97. [CrossRef] [PubMed]
7. Kim, H.-J.; Kim, D.-W.; Kim, W.K.; Cho, W.-J.; Kang, C.I. PVC membrane-based portable ion analyzer for hydroponic and water monitoring. *Comput. Electron. Agric.* **2017**, *140*, 374–385. [CrossRef]
8. Kim, H.-J.; Hummel, J.W.; Sudduth, K.A.; Motavalli, P.P. Simultaneous Analysis of Soil Macronutrients Using Ion-Selective Electrodes. *Soil Sci. Soc. Am. J.* **2007**, *71*, 1867. [CrossRef]
9. Rundle, C.C. A Beginners Guide To Ion-Selective Electrode Measurements. 2000. Available online: <http://www.nico2000.net/Book/Guide1.html> (accessed on 14 September 2013).
10. Lindner, E.; Pendley, B.D. A tutorial on the application of ion-selective electrode potentiometry: An analytical method with unique qualities, unexplored opportunities and potential pitfalls; Tutorial. *Anal. Chim. Acta* **2013**, *762*, 1–13. [CrossRef]
11. Bratov, A.; Abramova, N.; Ipatov, A. Recent trends in potentiometric sensor arrays—A review. *Anal. Chim. Acta* **2010**, *678*, 149–159. [CrossRef]
12. Codinachs, L.M.; Baldi, A.; Merlos, A.; Abramova, N.; Ipatov, A.; Jimenez-Jorquera, C.; Bratov, A. Integrated multisensor for FIA-based electronic tongue applications. *IEEE Sens. J.* **2008**, *8*, 608–615. [CrossRef]
13. Gutiérrez, M.; Moo, V.M.; Alegret, S.; Leija, L.; Hernández, P.R.; Muñoz, R.; del Valle, M. Electronic tongue for the determination of alkaline ions using a screen-printed potentiometric sensor array. *Microchim. Acta* **2008**, *163*, 81–88. [CrossRef]
14. Jung, D.H.; Kim, H.J.; Choi, G.L.; Ahn, T.I.; Son, J.E.; Sudduth, K.A. Automated Lettuce Nutrient Solution Management Using An Array of Ion-Selective Electrodes. *Trans. Asabe* **2015**, *58*, 1309–1319.
15. Cho, W.J.; Kim, H.J.; Jung, D.H.; Kim, D.W.; Ahn, T.I.; Son, J.E. On-site ion monitoring system for precision hydroponic nutrient management. *Comput. Electron. Agric.* **2018**, *146*, 51–58. [CrossRef]
16. Wang, L.; Cheng, Y.; Lamb, D.; Lesniewski, P.J.; Chen, Z.L.; Megharaj, M.; Naidu, R. Novel recalibration methodologies for ion-selective electrode arrays in the multi-ion interference scenario. *J. Chemom.* **2017**, *31*. [CrossRef]
17. Mueller, A.V.; Hemond, H.F. Statistical generation of training sets for measuring NO_3^- , NH_4^+ and major ions in natural waters using an ion selective electrode array. *Environ. Sci. Process. Impacts* **2016**, *18*, 590–599. [CrossRef] [PubMed]
18. Wang, L.; Yang, D.; Fang, C.; Chen, Z.L.; Lesniewski, P.J.; Mallavarapu, M.; Naidu, R. Application of neural networks with novel independent component analysis methodologies to a Prussian blue modified glassy carbon electrode array. *Talanta* **2015**, *131*, 395–403. [CrossRef] [PubMed]
19. Gutiérrez, M.; Alegret, S.; Cáceres, R.; Casadesús, J.; Marfà, O.; del Valle, M. Application of a potentiometric electronic tongue to fertigation strategy in greenhouse cultivation. *Comput. Electron. Agric.* **2007**, *57*, 12–22. [CrossRef]
20. Mueller, A.V.; Hemond, H.F. Extended artificial neural networks: Incorporation of a priori chemical knowledge enables use of ion selective electrodes for in-situ measurement of ions at environmentally relevant levels. *Talanta* **2013**, *117*, 112–118. [CrossRef]
21. Cuartero, M.; Ruiz, A.; Oliva, D.J.; Ortuño, J.A. Multianalyte detection using potentiometric ionophore-based ion-selective electrodes. *Sens. Actuators B Chem.* **2017**, *243*, 144–151. [CrossRef]
22. Duarte, L.T.; Jutten, C. Design of Smart Ion-Selective Electrode Arrays Based on Source Separation through Nonlinear Independent Component Analysis. *Oil Gas Sci. Technol.-Rev. D Ifp Energ. Nouv.* **2014**, *69*, 293–306. [CrossRef]
23. Duarte, L.T.; Suyama, R.; Attux, R.; Romano, J.M.T.; Jutten, C. A novel blind source separation method based on monotonic functions and its application to ion-selective electrode arrays. In Proceedings of the 2017 ISOCs/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), Montreal, QC, Canada, 28–31 May 2017.
24. Wang, L.; Yang, D.; Chen, Z.L.; Lesniewski, P.J.; Naidu, R. Application of neural networks with novel independent component analysis methodologies for the simultaneous determination of cadmium, copper, and lead using an ISE array. *J. Chemom.* **2014**, *28*, 491–498. [CrossRef]
25. Magalhaes, J.; Machado, A. Array of potentiometric sensors for the analysis of creatinine in urine samples. *Analyst* **2002**, *127*, 1069–1075. [CrossRef] [PubMed]
26. Rudnitskaya, A.; Costa, A.M.S.; Delgadillo, I. Calibration update strategies for an array of potentiometric chemical sensors. *Sens. Actuators B Chem.* **2017**, *238*, 1181–1189. [CrossRef]

27. Chen, F.; Wei, D.; Tang, Y. Virtual Ion Selective Electrode for Online Measurement of Nutrient Solution Components. *IEEE Sens. J.* **2011**, *11*, 462–468. [\[CrossRef\]](#)
28. Jung, D.-H.; Kim, H.-J.; Kim, H.S.; Choi, J.; Kim, J.D.; Park, S.H. Fusion of Spectroscopy and Cobalt Electrochemistry Data for Estimating Phosphate Concentration in Hydroponic Solution. *Sensors* **2019**, *19*, 2596. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Tuan, V.N.; Dinh, T.D.; Khattak, A.M.; Zheng, L.; Chu, X.; Gao, W.; Wang, M. Multivariate Standard Addition Cobalt Electrochemistry Data Fusion for Determining Phosphate Concentration in Hydroponic Solution. *IEEE Access* **2020**, *8*, 28289–28300. [\[CrossRef\]](#)
30. Cho, W.J.; Kim, H.J.; Jung, D.H.; Han, H.J.; Cho, Y.Y. Hybrid Signal-Processing Method Based on Neural Network for Prediction of NO₃, K, Ca, and Mg Ions in Hydroponic Solutions Using an Array of Ion-Selective Electrodes. *Sensors* **2019**, *19*, 5508. [\[CrossRef\]](#)
31. Sales, F.; Callao, M.P.; Rius, F.X. Multivariate standardization for correcting the ionic strength variation on potentiometric sensor arrays. *Analyst* **2000**, *125*, 883–888. [\[CrossRef\]](#)
32. Jones, J.B., Jr. *Hydroponics: A Practical Guide for the Soilless Grower*; CRC Press Inc.: Boca Raton, FL, USA, 2005.
33. Yu, C.; Seslija, M.; Brownbridge, G.; Mosbach, S.; Kraft, M.; Parsi, M.; Davis, M.; Page, V.; Bhawe, A. Deep Kernel Learning Approach to Engine Emissions Modelling. *Data-Cent. Eng.* **2020**. [\[CrossRef\]](#)
34. Alsaedi, B.S.O.; McGraw, C.M.; Schaerf, T.M.; Dillingham, P.W. Multivariate limit of detection for non-linear sensor arrays. *Chemom. Intell. Lab. Syst.* **2020**, *201*, 104016. [\[CrossRef\]](#)
35. Wilson, A.G.; Hu, Z.T.; Salakhutdinov, R.; Xing, E.P. Deep kernel learning. In *Artificial Intelligence and Statistics*; Gretton, A., Robert, C.C., Eds.; Microtome Publishing: Brookline, MA, USA, 2016; Volume 51, pp. 370–378.
36. Jiu, M.; Sahbi, H. Nonlinear deep kernel learning for image annotation. *IEEE Trans. Image Process.* **2017**, *26*, 1820–1832. [\[CrossRef\]](#)
37. Zheng, S.H.; Liu, K.X.; Xu, Y.L.; Chen, H.; Zhang, X.L.; Liu, Y. Robust Soft Sensor with Deep Kernel Learning for Quality Prediction in Rubber Mixing Processes. *Sensors* **2020**, *20*, 695. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Conagin, A.; Barbin, D.; Zocchi, S.S.; Demétrio, C.G.B. Fractional factorial designs for fertilizer experiments with 25 treatments in poor soils. *Rev. Bras. Biom.* **2014**, *32*, 180–189.
39. Trejo-Téllez, L.I.; Gómez-Merino, F.C. Nutrient solutions for hydroponic systems. In *Hydroponics-A Standard Methodology for Plant Biological Researches*; InTech: Rijeka, Croatia, 2012.
40. Puri, M.; Pathak, Y.; Sutariya, V.K.; Tippetaraju, S.; Moreno, W. *Artificial Neural Network for Drug Design, Delivery and Disposition*; Academic Press: San Diego, CA, USA, 2015.
41. Basheer, I.A.; Hajmeer, M. Artificial neural networks: Fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [\[CrossRef\]](#)
42. Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; MIT Press; MIT: Cambridge, MA, USA, 2006.
43. Neal, R.M. *Bayesian Learning for Neural Networks*; Springer Science & Business Media: New York, NY, USA, 2012; Volume 118.
44. Wilson, A.; Adams, R. Gaussian process kernels for pattern discovery and extrapolation. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 13 February 2013; pp. 1067–1075.
45. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
46. Horvai, G.; Tóth, K.; Pungor, E. A simple continuous method for calibration and measurement with ion-selective electrodes. *Anal. Chim. Acta* **1976**, *82*, 45–54. [\[CrossRef\]](#)
47. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
48. Wang, L.; Cheng, Y.; Lamb, D.; Chen, Z.; Lesniewski, P.J.; Megharaj, M.; Naidu, R. Simultaneously determining multi-metal ions using an ion selective electrode array system. *Environ. Technol. Innov.* **2016**, *6*, 165–176. [\[CrossRef\]](#)
49. Morf, W.E. *The Principles of Ion-Selective Electrodes and of Membrane Transport*; Elsevier: New York, NY, USA, 2012.
50. Baret, M.; Massart, D.; Fabry, P.; Conesa, F.; Eichner, C.; Menardo, C. Application of neural network calibrations to an halide ISE array. *Talanta* **2000**, *51*, 863–877. [\[CrossRef\]](#)
51. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
52. Dimeski, G.; Badrick, T.; St John, A. Ion Selective Electrodes (ISEs) and interferences-A review. *Clin. Chim. Acta* **2010**, *411*, 309–317. [\[CrossRef\]](#)

53. Mohammed, R.O.; Cawley, G.C. Over-fitting in model selection with Gaussian process regression. In Proceedings of the International Conference on Machine Learning and Data Mining in Pattern Recognition, New York, NY, USA, 15–20 July 2017; pp. 192–205.
54. Cortina, M.; Duran, A.; Alegret, S.; del Valle, M. A sequential injection electronic tongue employing the transient response from potentiometric sensors for anion multidetermination. *Anal. Bioanal. Chem.* **2006**, *385*, 1186–1194. [[CrossRef](#)]
55. Zhang, Y.; Ling, C. A strategy to apply machine learning to small datasets in materials science. *NPJ Comput. Mater.* **2018**, *4*, 1–8. [[CrossRef](#)]
56. Álvarez, M.A.; Lawrence, N.D. Computationally efficient convolved multiple output Gaussian processes. *J. Mach. Learn. Res.* **2011**, *12*, 1459–1500.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).