scientific reports

OPEN

Machine learning-assisted ammonium detection using zinc oxide/multi-walled carbon nanotube composite based impedance sensors

Akshaya Kumar Aliyana[1], S. K. Naveen Kumar[1], Pradeep Marimuthu[2], Aiswarya Baburaj[1], Michael Adetunji[4], Terrance Frederick[2], Praveen Sekhar[3] & Renny Edwin Fernandez[4] ✉

We report a machine learning approach to accurately correlate the impedance variations in zinc oxide/multi walled carbon nanotube nanocomposite (F-MWCNT/ZnO-NFs) to NH_4^+ ions concentrations. Impedance response of F-MWCNT/ZnO-NFs nanocomposites with varying ZnO:MWCNT compositions were evaluated for its sensitivity and selectivity to NH_4^+ ions in the presence of structurally similar analytes. A decision-making model was built, trained and tested using important features of the impedance response of F-MWCNT/ZnO-NF to varying NH_4^+ concentrations. Different algorithms such as kNN, random forest, neural network, Naïve Bayes and logistic regression are compared and discussed. ML analysis have led to identify the most prominent features of an impedance spectrum that can be used as the ML predictors to estimate the real concentration of NH_4^+ ion levels. The proposed NH_4^+ sensor along with the decision-making model can identify and operate at specific operating frequencies to continuously collect the most relevant information from a system.

Metal oxides-based sensors with excellent response time and longevity, are candidates for the next generation real-time sensing. Analytical signals from a chemical sensor are dominated by the target concentration. Besides the target molecule, interfering ions, process parameters like solution pH, conductivity, temperature, electrode parameters such as geometry, material properties also contribute to the sensor signal[1–4]. In a dynamic setting, these parameters contribute heavily to the signal erraticism[5]. Response of such intricate systems cannot be interpreted using calibration curves or modeled using mathematical equations.

Metal oxide active layers have highly tunable electrical, optical and mechanical properties[6–9]. Metal oxide nanocomposites with superior material properties and enhanced sensing range has been developed using ZnO, TiO_2, SiO_2, NiO_2, Fe_2O_3 and CuO[10,11]. Metal electrodes coated with metal oxide ion selective active layers have been used for electrochemical sensing of Ammonium (NH_4^+)[12–14]. The enhanced sensitivity of the ZnO is attributed to its surface to volume ratio, electrical conductivity, fast response, wide band-gap (3.37 eV)[15], large exciton binding energy (60 meV)[16], and piezoelectricity[17]. Metal oxide functionalization modulates the physicochemical properties of MWCNT thus increasing their ease of dispersion, manipulation, and process ability. Conjugating MWCNT with ZnO have shown to reduce the resistance of the sensing material and metal oxide controls the sensing properties[18–19]. Sensitivity of the electrodes are influenced by the chemical composition and microstructure morphology of the ZnO/MWCNT active layer[16,20,21]. After repeated measurements, metal-oxide integrity is degraded, resulting in a gradual decrease in maximum impedance ($|Z_{max}|$) values over the period of time[22].

Machine learning aided sensing. Authors have applied machine learning (ML) models to interpret the impedance response of a ZnO/MWCNT NH_4^+ sensor, mainly to compensate for the instability and drift in $|Z_{max}|$ values. The proposed ML approach centers the estimation of NH_4^+ on several predictors using advanced multivariate mathematical models. ML approach is adopted to correlate the impedance changes of a ZnO/MWCNT

[1]Department of Electronics, Mangalore University, Mangalore, India. [2]Rajeev Gandhi College of Engineering and Technology, Puducherry, India. [3]School of Engineering and Computer Science, Washington State University, Vancouver, USA. [4]Department of Engineering, Norfolk State University, Norfolk, USA. ✉email: refernandez@nsu.edu

active layer to NH_4^+ concentrations at varying solution conditions. Machine learning aided impedance sensors have been previously reported to classify *Escherichia coli* (*E. coli*) strains JM109, DH5-α, and *Salmonella typhimurium*[23]. However, no study has been reported on machine learning aided impedance sensors for detecting chemicals.

When trying to electrochemically analyze complex mixtures of structurally similar analytes in real matrices, various interferences occur, leading to faulty estimation of target concentration. Instability in electrochemical sensing is caused by structurally similar molecules/ions (Ca^{2+}, Mg^{2+}, K^+, H^+ and Na^+)[24], solution pH, sensor fouling, temperature etc.[25]. Application of machine learning techniques in chemical sensing provide several benefits: (1) can cut down sensor optimization cost and time (2) provide unexpected insights into the experimental data (3) can predict the outcome by deciphering non-linear analytical input signals without mathematical fitting (4) can be readily integrated into a IoT setting. Machine learning powered intelligent electrochemical sensors have better accuracy and reliability[26]. Systematic ML investigation discovers the hidden relationship between the analyte sample parameters and sensing signals through data compression, elimination of baseline drifts, normalization, transformations, and so forth[27].

Various ML models are used to train and extract significant features[28,29]. Support vector machine (SVM) model, used in cancer and pathogen detection, transforms predictor data into higher dimensions in order to maximize the response between the training patterns and the decision boundary[29,30]. The k-Nearest Neighbor (kNN) algorithm has been applied to solve classification problems in biology[31]. Naive Bayes (NB) is a probabilistic classifier based on Bayes theorem with strong (Naive) independence assumptions[32]. Decision tree (DT) based models have been proved to be efficient in various ML applications in particular chlorophyll (Chl) level monitoring[33], food quality assessment[34,35] and antioxidants detection in biodiesel[36]. Each node of DT stands for a feature in an instance to be tested, each branch stands for a value that the node can assume, and each leaf stands for a probability density value distribution. Extensively used ensemble ML model is random forest (RF), operates by developing a number of DTs for classification and regression with addressing overfitting issues[37,38]. Artificial neural networks (ANN) are primarily used to deal with the non-linear problems[39]. These features are particularly beneficial for real-time operations and quick decisions.

Most sensing systems either rely on an equivalent circuit[40] or a deterministic model[41] to make predictions about the anticipated output concentrations. Often, the models have stringent boundary conditions and require extracting sub features to correlate to calibration curves. We hypothesize that a machine learning aided analytical sensing will enhance the accuracy of field deployable sensors operating at dynamic experimental conditions. The success of machine learning aided sensors depend on extracting the right features from an electronic signal or a frequency response. Signals that appear random or decipherable only using complex mathematical equations often have repeatable features that have a collective value after preprocessing[42]. The extracted features are often either peak of waveforms, the rise and fall rates, distance between peaks etc. Impedance spectrum of the sensor provides a fingerprint for a measured sample.

In this paper, we explain how the sensing accuracy of a metal-oxide based impedance sensing system can be improved using a machine learning approach that identifies predominant spectral features to predict unknown NH_4^+ concentrations. We have analyzed the performance of in-house fabricated F-MWCNT/ZnO-NF nanocomposite sensors with varying ZnO:MWCNT ratios. The impedance characteristics of the sensors to varying NH_4^+ concentrations were analyzed to extract impedance and frequency features. Datasets that capture the uniqueness of an impedance response were generated along with experimental parameters, interfering ion concentrations. Features like $|Z_{max}|$, $|Z|$ at fixed frequencies, and average slope of the impedance curve were found to be important predictors. ML models like kNN, random forest, neural network, Naïve Bayes and logistic regression model were trained using datasets with more than 18 predictors. Some of the predictors were transformed to closely resemble a normal distribution as part of the preprocessing. The most dominant predictors were identified using feature importance analysis.

Experimental and methods

Chemicals.
Multi-Wall Carbon nanotube (MWCNT), Zinc nitrate hexahydrate ($H_{12}N_2O_{12}Zn$), Sodium hydroxide (NaOH), Hexamine((CH_2)$_6N_4$), Copper (Cu), Fibre epoxy resins, Di-methyl sulfoxide (DMSO), Nitric acid (HNO_3), Ammonium hydroxide (NH_4^+OH) and Ethanol are the precursor materials and solvents used to formulate the F-MWCNT/ZnO-NFs nanocomposites. The precursor materials and analytes were purchased from Sigma Aldrich. Interdigitated electrode structures were made from Copper (Cu) and Fiber epoxy resin (Fig. 1c).

Device characterization.
The surface orientation of the F-MWCNT/ZnO structure is studied using field emission scanning electron microscopy (FeSEM) with energy-dispersive X-ray spectroscopy (EDS). Crystal phase and orientation of F-MWCNT, ZnO and F-MWCNT/ZnO nanocomposites were examined using an X-ray diffractometer (XRD). Further, electrochemical performance of NH_4^+ detection device was measured using an Agilent 4294A precision impedance analyzer system. Measurements were repeated and analyzed using multiple sets of sensors.

Fabrication of F-MWCNT/ZnO-NFs nanocomposites based NH_4^+ smart sensor.
The fabrication process of screen-printed IDEs modified with F-MWCNT/ZnO active layer is illustrated in Fig. 1. The Copper (Cu) based IDEs structures were screen printed on the fiber epoxy substrate as described in our previous publication[16]. The CIRCAD software-oriented IDE's design consist of total 18 fingers and provide a total sensing area of 18×15 mm^2 for the NH_4^+ analytes.

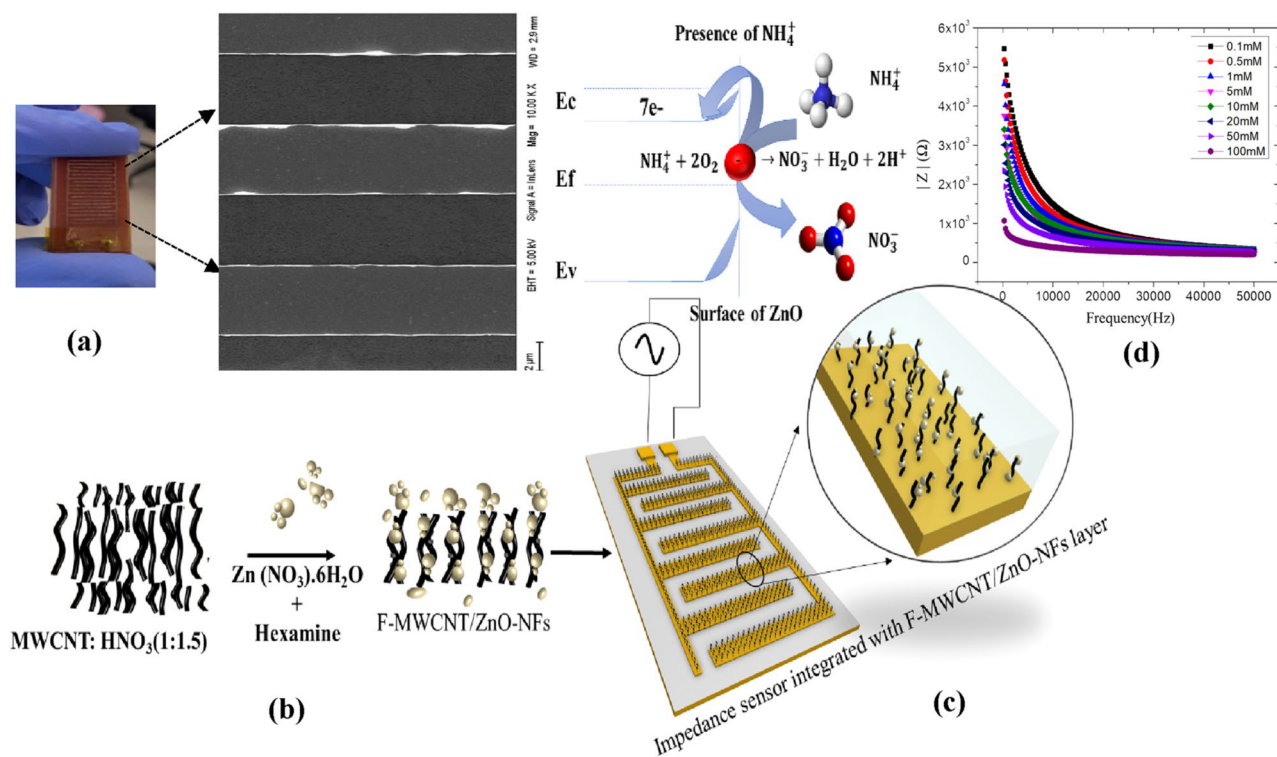**Figure 1.** (**a**) Screen printed interdigitated electrodes (IDEs), (**b**) F-MWCNT/ZnO-NFs nanocomposites are synthesized by functionalizing MWCNT using ZnO nanoparticles. (**c**) NH_4^+ selective impedance sensors are fabricated by embedding F-MWCNT/ZnO-NF active layers on interdigitated arrays and (**d**) electrochemical response of the NH_4^+ selective impedance sensors.

F-MWCNT was functionalized using ZnO nanoparticles to form F-MWCNT/ZnO-NFs nanoflowers (NFs) structures. MWCNT is added with Nitric acid (HNO_3) in the ratio of 1:1.5 solution is made homogeneous. The obtained solution is further refluxed for 8 h at 110 °C, consequently F-MWCNT powder treated with 0.1, 0.3 and 0.5 M Zn (NO_3)·$6H_2O$ to get the ZnO in the proportion of 1%, 3% and 5%. Three variants of Nanocomposites were synthesized by varying the MWCNT: ZnO ratio. The binding agent (0.01 M Hexamine) is added to the respective homogeneous nanocomposite solutions to create three types of sensors (S1, S2 and S3) with varying ZNO:MWCNT ratio. The mixture was refluxed for 3 h at 120 °C and precipitate is dried at 80 °C for 10 h to get the F-MWCNT/ZnO-NFs composites. The synthesized products were investigated in terms of morphological and structural properties. Nanocomposites further dispersed in the Di-methyl sulfoxide (DMSO) solvent were drop casted over the screen-printed IDEs surface.

Results and discussion

Characterizations of IDEs modified F-MWCNT/ZnO nanocomposite active layer.

The layout and prototype of the screen-printed IDEs and FeSEM images of the electrode patterns are shown in Fig. 1c. The width and space between the two electrodes designed for 500 μm which comes to ≈ 460 μm after printing. Evident from the field emission microscopy (FeSEM) studies on F-MWCNT (Fig. 2a1), F-MWCNT/ ZnO-NFs composites with proportion of 1:1 (Fig. 2a2), 1:3 (Fig. 2a3) and 1:5 (Fig. 2a4), the morphology of F-MWCNT/ ZnO-NFs composites with varying ZnO percentages were distinctly different. Formation of nanoflowers (NF) inclusions over the F-MWCNT were observed. The majority of the MWCNT and F-MWCNT particles are uniformly distributed in high compactness with almost parallel alignment to the surface. Nanoflowers (NFs) are assembled in bunches which contain sub-NFs. Further characterization of the NFs attributed a hexagonal structure to NFs confirming good crystallinity. We noticed that ZnO percentage in the composite defines the orientation of the NFs in such a way that its branching and petal size is proportional to the ratio of ZnO (1 to 5%).

Morphology and chemical composition of the nanocomposites were investigated using Energy Dispersive Spectroscopy (EDS) (Fig. S2). To analyze the crystalline nature of the particles, X-ray diffraction (XRD) study for ZnO, F-MWCNT, F-MWCNT/ZnO-NFs (S1, S2 and S3) samples were done and pattern is plotted in Fig. 2b. Obtained pattern shows maximum intensity for the peak positioned at $2\theta = 36.56°$ for ZnO along the (101) plane and relates the growth along c-axis parallel to the substrate.

Evaluation of electrochemical sensing performance of F-MWCNT/ZnO-NFs composites active layer.

Impedance response of the F-MWCNT/ZnO-NFs composite samples were recorded using Agilent 4294A precision impedance analyzer. The performance of the sensors was evaluated using standard NH_4^+ solutions (1–100 mM). Enhanced sensitivity and stability of the F-MWCNT/ZnO-NFs is attributed to the higher
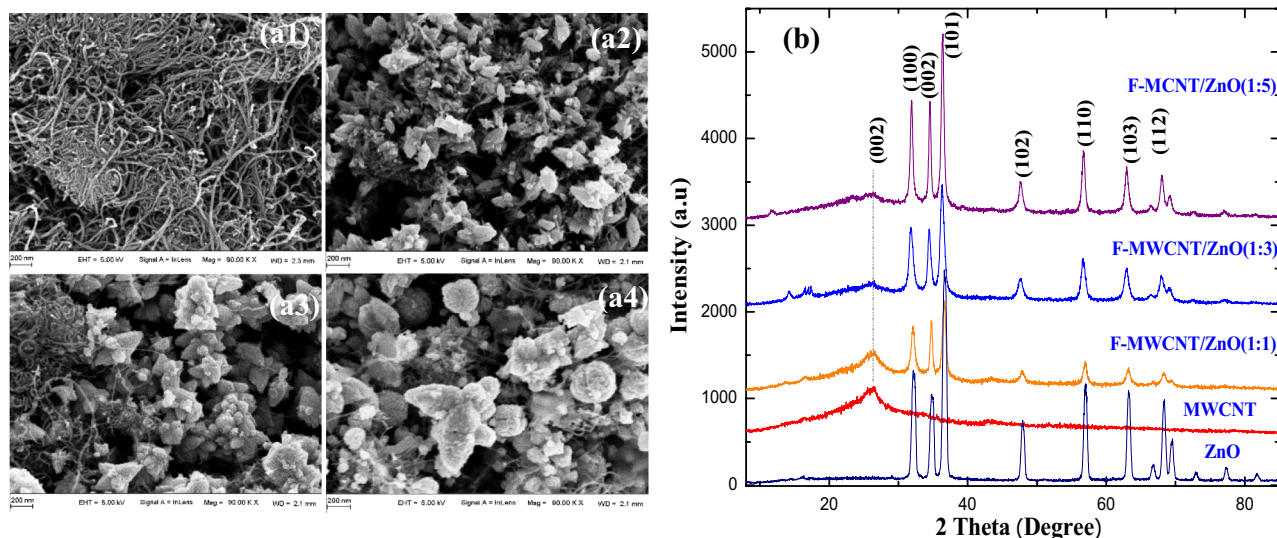
Figure 2. (**a**) FeSEM images of F-MWCNT (**a1**), F-MWCNT/ZnO-NFs composites with proportion of 1:1 (**a2**), 1:3 (**a3**), 1:5 (**a4**) and (**b**) X-ray diffraction (XRD) spectrum for ZnO, F-MWCNT, F-MWCNT/ZnO-NFs.
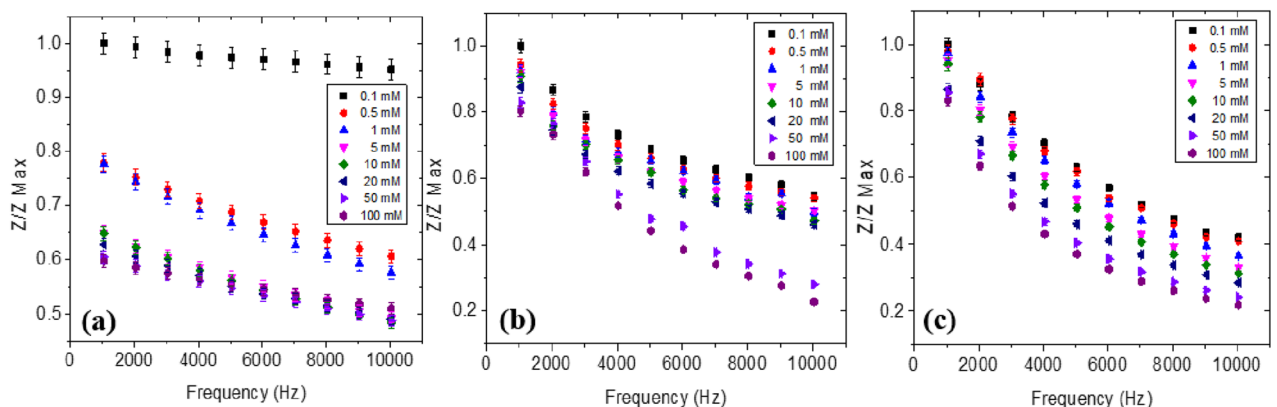

Figure 3. Impedance (Z/Z_{max}) versus frequency response of the sensors (**a**) S1, (**b**) S2 and (**c**) S3 at 0.1–100 mM NH_4^+ ion concentrations.

surface area and direct electron mobility[11]. NH_4^+ ions in the immediate vicinity of ZnO-NF active layers, are converted into $NO_3^{-[6,43]}$. NH_4^+ analytes adsorbed on ZnO-NFs are nitrified causing a decrease in the overall impedance of the sensor which induces a change in the local electrical fields and a proportional change in the electrical properties of the F-MWCNT/ZnO-NFs nanocomposites. The surface of ZnO-NFs will react with oxidizing and reducing analytes due to the presence of adsorbed oxygen species on the outermost layer, which subsequently gets ionized into O^{2-}, O^- or O_2^-.

Three types of the F-MWCNT/ZnO-NF nanocomposite sensors were constructed by varying the ZnO:MWCNT weight percentages. Sensors S1, S2 and S3 had ZnO:MWCNT ratios of 1:1, 1:3 and 1:5 respectively. Impedance spectra of the sensors showed an inversely proportional response with frequency in the 1–10 kHz range (Fig. S3), clearly at higher frequencies the total impedance of the sensor decreased[44,45]. Higher $|Z_{max}|$ were observed for S3 (1:5) compared to S1 (1:1) and S2 (1:3). Results confirm the predominant effect of ZnO weight percentages on the impedance characteristics of F-MWCNT/ZnO-NFs composites. $|Z_{max}|$ values also tend to slightly vary from device to device. Such anomalies lead to inaccurate analysis of target concentrations.

Ammonium sensing. The performance of the sensors S1, S2 and S3 was evaluated using standard NH_4^+ solutions (1–100 mM) (Fig. 3). The presence of NH_4^+ molecule in the dielectric (liquid medium) causes a change in the sensor impedance. A proportional decrease in the impedance magnitude was observed with increasing frequency (1 to 10 kHz) in all the three sample conditions. Linearity of the response was found to decrease with frequency. It was observed that response of impedance magnitude is influenced by the pH environment of the NH_4^+ testing samples. The impedance sensitivity of the electrode was drastically affected by the pH environments, the response being acute in the 1–10 kHz range.

Results indicate that higher pH conditions cause an enhanced impedance value for S1, S2 and S3. It can be noted that impedance values of all sensors decreased with increasing NH_4^+ levels in acid (Figs. 4a, 5a, 6a), neutral
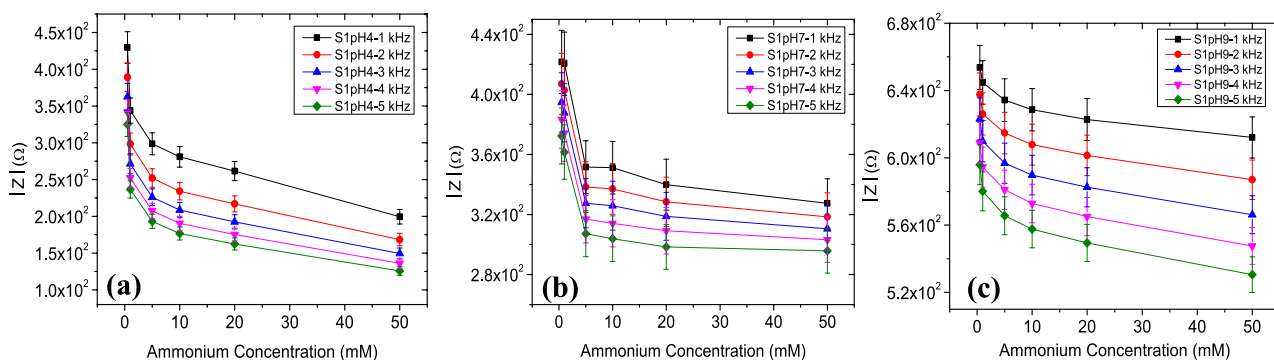
Figure 4. Impedance magnitude (Ω) versus NH_4^+ ion concentration (0.5–50 mM) response of the sensor S1 in the (**a**) acid, (**b**) neutral and (**c**) alkaline sample conditions.
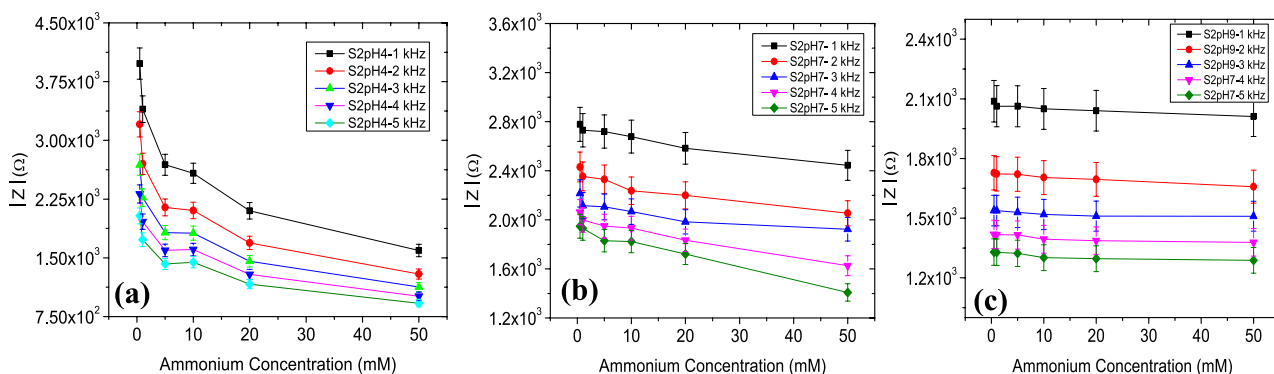

Figure 5. Impedance magnitude (Ω) versus NH_4^+ ion concentration (0.5–50 mM) response of the sensor S2 in the (**a**) acid, (**b**) neutral and (**c**) alkaline sample conditions.

Figure 6. Impedance magnitude (Ω) versus NH_4^+ ion concentration (0.5–50 mM) response of the sensor S3 in the (**a**) acid, (**b**) neutral and (**c**) alkaline sample conditions.
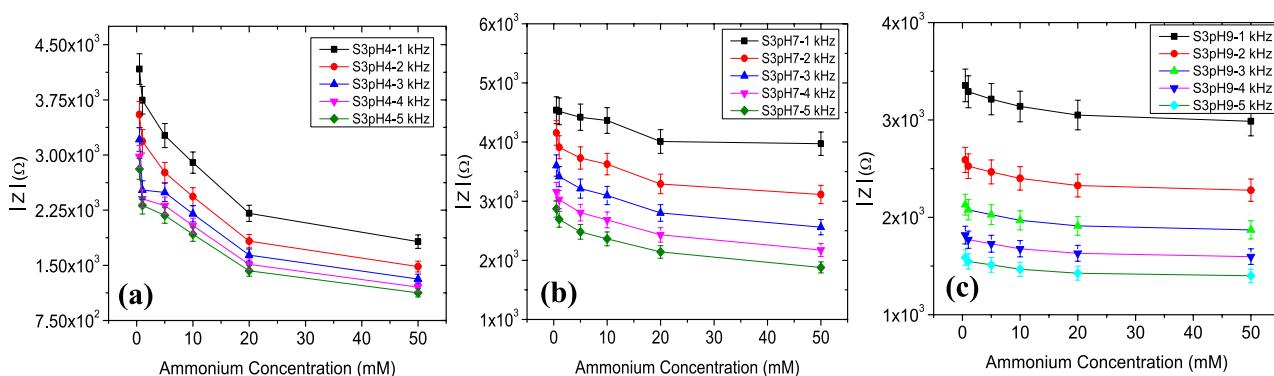
(Figs. 4b, 5b, 6b) and alkaline (Figs. 4c, 5c, 6c) conditions. The device sensitivity to NH_4^+ was also found to vary with pH. Critical values such as the $|Z_{max}|$ and the frequency dependency of $|Z|$ were found to vary from device to device. $|Z_{max}|$ values for sensor S3 operated at pH 4 was ~ 10.5% higher when operated in pH 9 and its ~ 23.8% higher for sensor S2. $|Z_{max}|$ values were comparatively stable for S1 sensors with a variation of ~ 2%. The rate of change of $|Z_{max}|$ is highest for pH 4 whereas increase in Z_{max} was very gradual at pH 7 and 9.

Figure 7a shows the response of $|Z_{max}|$ versus NH_4^+ concentration (1–20 mM) for three S3 devices in the acid, neutral and alkaline sample conditions. $|Z_{max}|$ values can be seen to decrease consistently with increasing NH_4^+ concentrations. $|Z_{max}|$ values for S3 I, II and III sensors showed ~ 5% variation which can be due to variation in the active layer thickness among the sensor prototypes. Response of the sensors were interpreted using linear impedance versus NH_4^+ calibration plots (Figs. S7, S8 and S9). The correlation coefficient (R^2) of these plots is in range of 0.75–0.90. Sensitivity of the sensors were between 76.09 and 1.06 mM/Ω with an average correlation coefficient (R^2) of 0.81. A decrease in sensitivity was observed when the solution pH was varied from acid to alkaline. This can be attributed to the non-uniformity in the distribution of ZnO-NFs on electrode active area
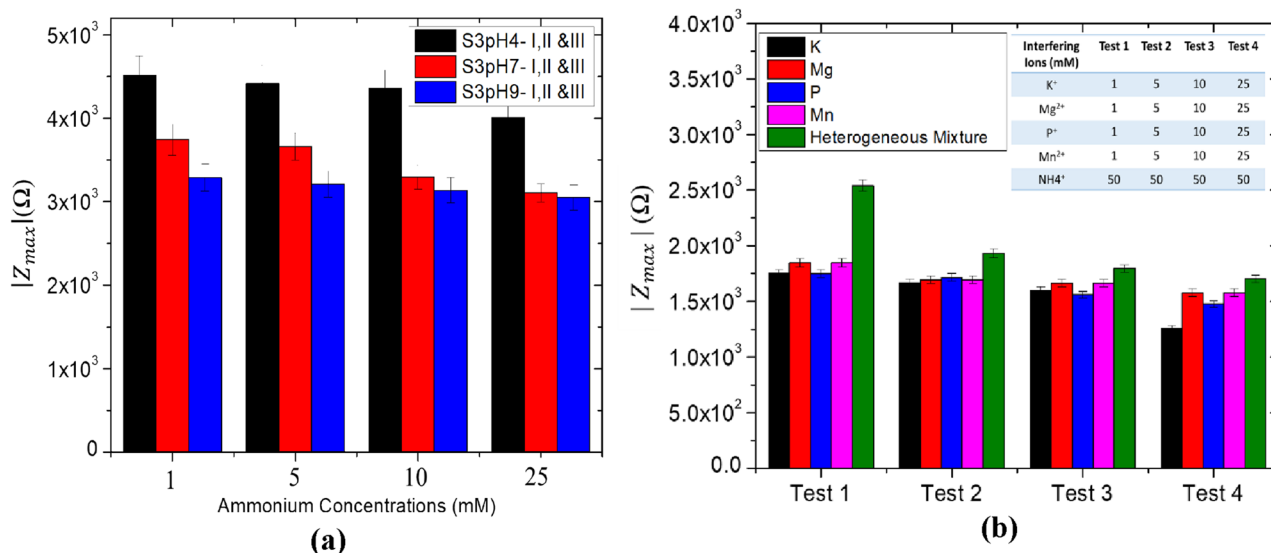
Figure 7. (**a**) $|Z_{max}|$ versus NH_4^+ concentration (1–20 mM) for three similar set of S3 devices in acid, neutral and alkaline pH conditions. (**b**) Selectivity performance of the sensor in the presence of Interfering ions (K^+, Mg^{2+}, P^+, Mn^{2+} and Heterogeneous mixture) concentrations levels.

(Fig. 2a). To interpret the stability of the sensor, impedance characteristics were recorded every 500 s for a total time duration of 4000 s and further measured every day for 7 continuous days. The results confirm the stable performance of the fabricated device.

To evaluate the selectivity of the NH_4^+ sensor, we introduced 1–25 mM of interfering ions (K^+, Mg^{2+}, P^+, and Mn^{2+}) in 50 mM NH_4^+ sample. Figure 7b reveals the selectivity performance of the device in the presence of interfering ions. Four experiments (Test 1–4) were performed to analyze the effect of interfering ions. Test 1–4 was conducted in the presence of 1, 10 and 25 mM each of the interfering ions (K^+, Mg^{2+}, P^+, and Mn^{2+}). Experiments were also conducted using a heterogenous mixture of all the interfering ions. In the presence of 1–25 mM interfering ions with 50 mM NH_4^+, $|Z_{max}|$ values were found to change sigificantly in the presence of interfering ions. A drift of ~ 5.14 to ~ 23.15% in $|Z_{max}|$ was observed in NH_4^+ sensor due to chemisorption of interfering ions in the MWCNT/ZnO active layer. Heterogenous mixture of ions resulted in a significant variation of ~ 4.5% (Fig. 7b) in the $|Z_{max}|$ values.

Machine learning based target prediction. Prediction of target variable concentration (NH_4^+) was done using 28 predictors in which 8 are experimental features and 20 are impedance and frequency features of the Frequency-impedance (F-Z) spectra (Fig. 8). S1, S2 and S3 are categorical predictors that indicate the type of sensor. The extracted features of the spectra are stored in a structure with a sensor and session ID number encoded in CSV file for each impedance spectra. Missing values were replaced with the median value of the predictor. 80% of dataset is assigned as training data set and the other 20% were assigned for testing. The accuracy of the ML regression models was analyzed using a confusion matrix in order to compare the predicted and true values. ML analyses were performed using the Orange toolbox by writing Python scripts accessing the Orange API. Python Script widget are used to run python scripts in Orange to perform additional functionalities like feature importance[46]. The updated variables from the Python script are used as outputs of the Python Script widget.

Feature extraction and model training. Predictors of our model were extracted from the impedance spectra, sample features, and experimental settings (Fig. 9). Both impedance magnitude and frequency features are used as predictors to capture the dynamics of the original spectra. Individual features were expanded into multiple predictors in order to better represent a non-linear input–output relationship thereby eliminating outliers and reducing computational time[47]. The frequencies at which a fixed change in impedance occurs are taken as the derivative frequency values (f_n) of a spectrum (Fig. 9b). Predictors $f_{0.1}$ to $f_{0.9}$ are the frequencies at which the impedance values are 0.1–0.9 times of Z_{max} from the impedance spectra, we extracted derivative frequencies f_n at $Z_n = n * Z_{max} \{ 0.1 \leq n \leq 0.9 \}$. We hypothesize that representing the impedance variations as f_n will reduce the dependency of the spectrum on electrode area. f_n starts to increase at the onset of depletion and a maximum value of f_n is observed at $0.9 * Z_{max}$. Although Z_{max} value is a function of the area of the electrode, the frequency response is influenced by the dielectric property of the material and the environment. Total surface area of an electrode may decrease after cycles of repeated measurements as the nanocomposite structures in the active layer tend to deteriorate.

We transformed the impedance features into admittance values in order to improve the model. Raw impedance data were preprocessed using a running mean prior to modeling. But excess data smoothing showed a tendency to eliminate important nonlinear patterns. Hence, we adopted a Box-Cox transformation method that
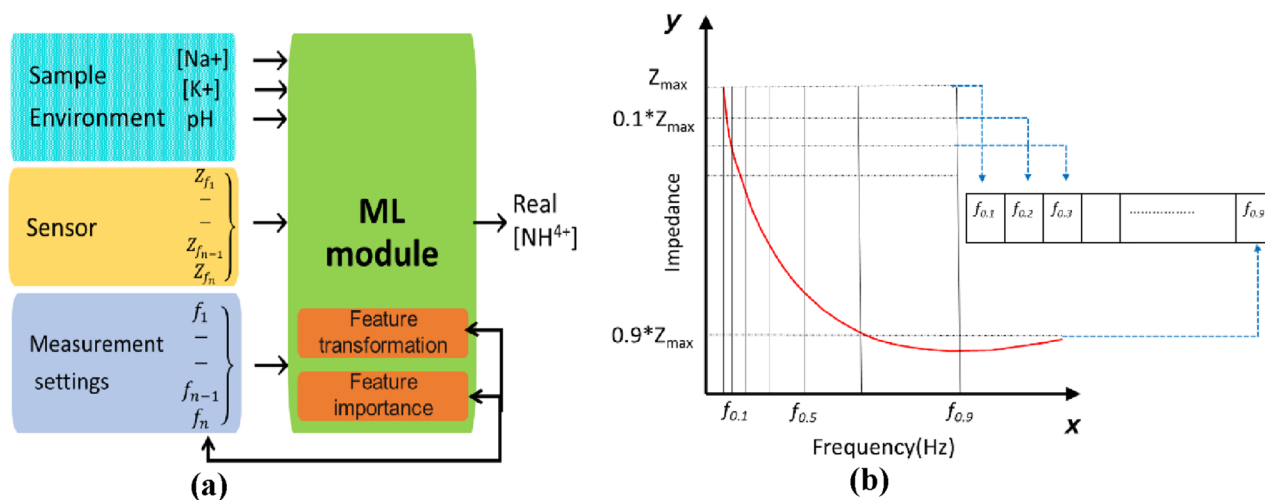
	Ds1	Ds2	Ds3
Z_{1k-10k}	■	■	
$f_{0.1-0.9}$		■	
Y_{1k-10k}			■
pH 4	■	■	■
pH 7	■	■	■
pH 9	■	■	■
K^+	■	■	■
P^+	■	■	■
Mn^+	■	■	■
Mg^+	■	■	■
S1, S2, S3	■	■	■

(a)
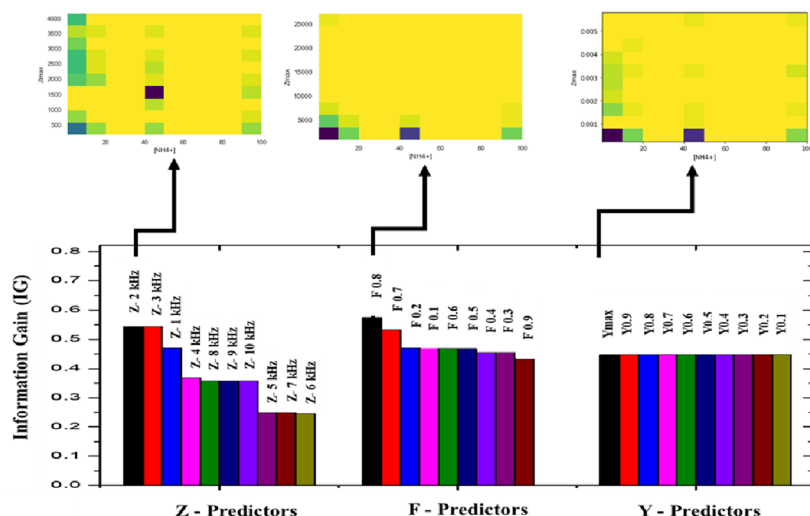
(b)

Figure 8. (**a**) Impedance spectral features and experimental parameters used as predictors in various datasets. (**b**) Information gain scores for Z, F and Y predictors. Insets show the variation in Z_{max} values of predictors with maximum IG scores (Z_{2k}, $f_{0.8}$, Y_{max}).

Figure 9. (**a**) Predictors of the ML model are extracted from Impedance Spectral features, sample parameters, and experimental settings. (**b**) Predictors $f_{0.1}$ to $f_{0.9}$ are the frequencies at which the impedance values are 0.1–0.9 times of Z_{max}.

uses separate estimation procedures to transform the impedance (Z) data so that it closely resembles a normal distribution. A Box-Cox transformation was used to estimate the transformation.

$$x^* = \begin{cases} \frac{x^\lambda - 1}{\lambda \tilde{x}^{\lambda-1}}, & \lambda \neq 0 \\ \tilde{x} \log x, & \lambda = 0 \end{cases}$$

The Box-Cox procedure, originally intended as a transformation of a model's outcome, uses maximum likelihood estimation to estimate a transformation parameter λ in the equation which is estimated from the data. Values of λ map to transformations. When $\lambda = 1$ (no transformation), $\lambda = 0$ (log), $\lambda = 0.5$ (square root), and $\lambda = -1$ (inverse). For impedance (Z) values, λ was estimated to be -1 which is effectively the inverse transformation. Hence, we adopted the admittance representation in our datasets using the feature constructor tool in Orange.

Prediction efficiency of the decision-making model. From the confusion matrices, scoring frameworks such as CA (Accuracy classification score), Precision (precision computation), recall (computation of true positives and false negatives), F1 score (weighted average of precision and recall) and AUC (stored predic-

DS1

Feature	Info. gain
Z-2 kHz	0.5446540380784097
Z-3 kHz	0.5446540380784097
Z-1 kHz	0.46971456018798596
Z-4 kHz	0.3679208135854708
Z-8 kHz	0.35827071155755297
Z-9 kHz	0.35827071155755297
Z-10 kHz	0.35827071155755297
pH	0.33585071183290305
Z-5 kHz	0.24933706493829133
Z-7kHz	0.24851772710451758
Z-6 kHz	0.24586825898826836
K+ (mM)	0.15744648076058
P+ (mM)	0.15744648076058
Mg+ (mM)	0.15744648076058
Mn+ (mM)	0.15744648076058
S3	0.1355032637390603
S1	0.04702540135221289
S2	0.04702540135221289

DS2

Feature	Info. gain
F-0.8	0.5749759804072121
F-0.7	0.5328707172493177
F-0.2	0.47076577566282696
F-0.1	0.47076577566282696
Z-Max	0.46971456018798596
Z-0.9	0.46971456018798596
Z-0.8	0.46971456018798596
Z-0.7	0.46971456018798596
Z-0.6	0.46971456018798596
Z-0.5	0.46971456018798596
Z-0.4	0.46971456018798596
Z-0.3	0.46971456018798596
Z-0.2	0.46971456018798596
Z-0.1	0.46971456018798596
F-0.6	0.46797985454187563
F-0.5	0.46797985454187563
F-0.4	0.4548734071962275
F-0.3	0.4548734071962275
F-0.9	0.43125239996512166
pH	0.33585071183290305
K+ (mM)	0.15744648076058
P+ (mM)	0.15744648076058
Mg+ (mM)	0.15744648076058
Mn+ (mM)	0.15744648076058
S3	0.1355032637390603
S1	0.04702540135221289
S2	0.04702540135221289

DS3

Feature	Info. gain
F-0.8	0.5749759804072121
F-0.7	0.5328707172493177
F-0.2	0.47076577566282696
F-0.1	0.47076577566282696
F-0.6	0.46797985454187563
F-0.5	0.46797985454187563
F-0.4	0.4548734071962275
F-0.3	0.4548734071962275
1/Zmax	0.4474166794734358
1/Z-0.9	0.4474166794734358
1/Z-0.8	0.4474166794734358
1/Z-0.7	0.4474166794734358
1/Z-0.6	0.4474166794734358
1/Z-0.5	0.4474166794734358
1/Z-0.4	0.4474166794734358
1/Z-0.3	0.4474166794734358
1/Z-0.2	0.4474166794734358
1/Z-0.1	0.4474166794734358
F-0.9	0.43125239996512166
pH	0.33585071183290305
K+ (mM)	0.15744648076058
P+ (mM)	0.15744648076058
Mg+ (mM)	0.15744648076058
Mn+ (mM)	0.15744648076058
S3	0.1355032637390603
S1	0.04702540135221289
S2	0.04702540135221289

Table 1. Information gain (IG) scores for DS1, DS2 and DS3.

tions and actual data in model testing) were obtained to evaluate the prediction efficiency of each model. Scoring framework takes the decimal values of probability between 0 and 1. Evident from the Confusion matrices (Tables S5–S10), are the prediction accuracy of all the models among which RF is the most accurate (Table S5). Among all the models that we tested, RF model was the most efficiency in predicting the concentration range, which has a precision score of 0.83–0.89 depending upon the predictor variables. Using impedance only dataset, RF displayed a precision of ~0.83. Precision values of the RF model increased from 0.83 to 0.85 when transformed $1/Z$ values were used. The precision of the RF model was increased further to ~0.87 after including derivative frequency (f_n) predictors. Accuracy was highest for RF model, when admittance ($1/Z$) were used as predictor instead of impedance values. Prediction accuracy of the model was lower for samples in low pH solutions. This may be attributed to high variance in Z values in low pH solutions. The accuracy was found to be highest in basic pH.

All other models like ANN and kNN had lower prediction accuracy (< 0.70). SVM models with four different kernels (polynomial, sigmoidal, linear, and radial basis function) were analyzed. However, SVM showed the least accuracy (0.24–0.47) while testing for unknown samples.

Feature importance estimation for maximum relevance and minimum redundancy criterion.

Extraneous predictors erode the performance of a model and also increases computational cost. It is important to identify a subset of predictors with high prevalence. Among the predictors, some of them have a significant influence on the response of the model. Feature importance was tested by permuting the predictors and measuring the decrease in accuracy. In order to identify the predictor combination that is relevant to the response, we made 3 datasets containing predictors derived from the features of the impedance spectra. DS (1) Z_{1k-10k}, (2) $Z_{1k-10k}, f_{0.1-0.9}$ (3) $Y_{1k-10k}, f_{0.1-0.9}$ (Fig. 8a).

A complete list of all features and their Information Gain (IG) scores for RF model is shown in Table 1. IG scores indicate the significance of a predictor in the model. The IG scores of various predictors are ranked based on highest and lowest precedence features. We evaluated the information gain for each variable in order to select only those variables that maximizes the information gain. IG score provides a baseline for comparison; some of the features are removed based on feature importance scores.

Among three datasets analyzed, each consisting of ~150 data points and 18 features, Z_{2k}, the Z value corresponding to the impedance at 2 kHz, had the highest IG score 0.544 which is higher among all Z predictors (Z_{1k-10k}). IG scores were higher for f_n predictors. Among them $f_{0.8}$ had the highest score (0.574). S predictors, S1, S2, S3, was found to have low info gains (~0.047). Following points summarizes our findings from the feature analysis.

(1) It was clear from the IG scores that categorical S-predictor parameter S3 has a higher significance than S2 and S1
(2) Impedance predictors had higher IG scores than Admittance.
(3) Frequency predictors (f_n) had a higher significance than impedance or Admittance predictors. Among the f_n predictors, $f_{0.8}$ (0.574) had the highest significance and $f_{0.9}$ (0.43) had lowest significance.

(4) The effect of interfering ions and pH were consistent across all types of devices (S1, S2 and S3). For all datasets, IG scores for pH was 0.33 and interfering ion concentrations scores were 0.157.

Besides analyzing the impedance spectra for prominent features, our study can be used to design impedance sensing circuits specifically for a sensor. The model is capable of identifying and operating at specific sensing frequencies to continuously collect the most relevant information from a system.

Conclusions

We report a smart NH_4^+ sensing system based on F-MWCNT/ZnO-NFs sensor and a ML model for the purpose of accurately correlating sensor impedance to NH_4^+ concentrations. F-MWCNT/ZnO-NFs composites with varying ZnO:MWCNT ratio (S1, S2, and S3) were analyzed for their NH_4^+ sensitivity. The machine learning based decision support tool which is the new frame of reference has been designed to interpret the features of impedance spectra and predict the real NH_4^+ concentrations by mitigating the effects of non-linearity and impedance drifts. ML analysis has led to identification of dominant predictors of our sensing system. Our work has confirmed the possibility of integrating a custom ML module in an impedance sensing system in order to mitigate the detrimental effect of interference ions and pH on sensitivity and selectivity of F-MWCNT/ZnO-NF active layers. Different models such as kNN, random forest, neural network, Naïve Bayes and logistic regression support have been contrasted and discussed. Depending on the analyte concentration and the metal-oxide, the proposed model can be extended to most impedance-based sensing systems, although dominant features have to be reanalyzed using feature scores and feature transformation.

References

1. Kumar, A. A., Kumar, S. N. & Fernandez, R. E. Real time sensing of soil potassium levels using zinc oxide-multiwall carbon nano-tube based sensors. *IEEE Trans. NanoBiosci.* **20**, 50–56 (2020).
2. Fazio, E. *et al.* Metal-oxide based nanomaterials: Synthesis, characterization and their applications in electrical and electrochemical sensors. *Sensors* **21**, 2494 (2021).
3. Chavali, M. S. & Nikolova, M. P. Metal oxide nanoparticles and their applications in nanotechnology. *SN Appl. Sci.* **1**, 1–30 (2019).
4. Hatchett, D. W. & Josowicz, M. Composites of intrinsically conducting polymers as sensing nanomaterials. *Chem. Rev.* **108**, 746–769 (2008).
5. Hu, Y., Tan, O. K., Pan, J. S., Huang, H. & Cao, W. The effects of annealing temperature on the sensing properties of low temperature nano-sized $SrTiO_3$ oxygen gas sensor. *Sens. Actuators B Chem.* **108**, 244–249 (2005).
6. Franco, F. F., Manjakkal, L., Shakthivel, D. & Dahiya, R. ZnO based screen printed aqueous ammonia sensor for water quality monitoring. *Proc. IEEE Sens.* **2019-Octob**, 1–4 (2019).
7. Vinoth, E. & Gopalakrishnan, N. Fabrication of interdigitated electrode (IDE) based ZnO sensors for room temperature ammonia detection. *J. Alloys Compd.* **824**, 153900 (2020).
8. Punetha, D. & Pandey, S. K. Enhancement and optimization in sensing characteristics of ammonia gas sensor based on light assisted nanostructured WO_3 thin film. *IEEE Sens. J.* **20**, 14617–14623 (2020).
9. Tohidi, S., Parhizkar, M., Bidadi, H. & Mohamad-Rezaei, R. Electrodeposition of polyaniline/three-dimensional reduced graphene oxide hybrid films for detection of ammonia gas at room temperature. *IEEE Sens. J.* **20**, 9660–9667 (2020).
10. Saleh, T. A. & Fadillah, G. Recent trends in the design of chemical sensors based on graphene–metal oxide nanocomposites for the analysis of toxic species and biomolecules. *TrAC Trends Anal. Chem.* **120**, 115660 (2019).
11. Rahman, M. M., Balkhoyor, H. B., Asiri, A. M. & Marwani, H. M. A gold electrode modified with silver oxide nanoparticle decorated carbon nanotubes for electrochemical sensing of dissolved ammonia. *Microchim. Acta* **183**, 1677–1685 (2016).
12. Sekhar, P. K. & Kysar, J. S. An electrochemical ammonia sensor on paper substrate. *J. Electrochem. Soc.* **164**, B113–B117 (2017).
13. Ahmad, R. *et al.* Ammonium ion detection in solution using vertically grown ZnO nanorod based field-effect transistor. *RSC Adv.* **6**, 54836–54840 (2016).
14. Kumar, A. A. & Kumar, S. K. N. Comprehensive review on pH and nutrients detection sensitive materials and methods for agriculture applications. *Sens. Lett.* **17**, 663–670 (2019).
15. Naveen Kumar, S. K., Akshaya Kumar, A., Aniley, A. A., Bhansali, S. & Fernandez, R. E. Hydrothermal growth of zinc oxide (ZnO) nanorods (NRs), structural, and chemical composition studies for pH measurement sensor applications. *ECS Trans.* **88**, 437–447 (2019).
16. Kumar, A. A., Aniley, A. A., Fernandez, R. E. & Bhansali, S. Hydrothermal growth of zinc oxide (ZnO) nanorods (NRs) on screen printed IDEs for pH measurement application. *J. Electrochem. Soc.* **166**, 3264–3270 (2019).
17. Xiang, H. J., Yang, J., Hou, J. G. & Zhu, Q. Piezoelectricity in ZnO nanowires: A first-principles study. 2006–2008. *Appl. Phys. Lett.* **89**, 223111. https://doi.org/10.1063/1.2397013 (2006).
18. Wu, J. J. & Liu, S. C. Low-temperature growth of well-aligned ZnO nanorods by chemical vapor deposition. *Adv. Mater.* **14**, 215–218 (2002).
19. Timmer, B., Olthuis, W. & Van Den Berg, A. Ammonia sensors and their applications—a review. *Sens. Actuators B Chem.* **107**, 666–677 (2005).
20. Kumar, A. A., Kalappa, S. & Kumar, N. Integration of interdigitated electrodes (IDEs) with ZnO nanoflakes (Nfs) active layers integration of interdigitated electrodes (IDEs) with ZnO nanoflakes (Nfs) active layers. In *AIP Conference Proceedings* 070008 (2020).
21. Gangu, K. K., Maddila, S. & Jonnalagadda, S. B. A review on novel composites of MWCNTs mediated semiconducting materials as photocatalysts in water treatment. *Sci. Total Environ.* **646**, 1398–1412 (2019).
22. Fernández-Sánchez, C., McNeil, C. J. & Rawson, K. Electrochemical impedance spectroscopy studies of polymer degradation: Application to biosensor development. *TrAC Trends Anal. Chem.* **24**, 37–48 (2005).
23. Ali, S. *et al.* Disposable all-printed electronic biosensor for instantaneous detection and classification of pathogens. *Sci. Rep.* **8**, 1–11 (2018).
24. Silvester, D. S. Recent advances in the use of ionic liquids for electrochemical sensing. *Analyst* **136**, 4871–4882 (2011).
25. Wang, J. Modified electrodes for electrochemical sensors. *Electroanalysis* **3**, 255–259 (1991).
26. Nath, A., Agarwal, S. & Ghosh, A. Classification of machine learning algorithms. *Int. J. Innov. Res. Adv. Eng.* **3**, 6–11 (2016).

27. Oladipupo, T. Types of machine learning algorithms. *New Adv. Mach. Learn.* https://doi.org/10.5772/9385 (2010).
28. Dong, R., Weng, S., Yang, L. & Liu, J. Detection and direct readout of drugs in human urine using dynamic surface-enhanced Raman spectroscopy and support vector machines. *Anal. Chem.* **87**, 2937–2944 (2015).
29. Majumder, S. K., Ghosh, N. & Gupta, P. K. Support vector machine for optical diagnosis of cancer. *J. Biomed. Opt.* **10**, 024034 (2005).
30. Wang, C., Madiyar, F., Yu, C. & Li, J. Detection of extremely low concentration waterborne pathogen using a multiplexing self-referencing SERS microfluidic biosensor. *J. Biol. Eng.* **11**, 1–11 (2017).
31. Peterson, L. K-nearest neighbor. *Scholarpedia* **4**, 1883 (2009).
32. Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A. & Strachan, R. Hybrid decision tree and Naïve Bayes classifiers for multi-class classification tasks. *Expert Syst. Appl.* **41**, 1937–1946 (2014).
33. Amirruddin, A. D. *et al.* Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (*Elaeis guineensis*) based on frond numbers: Analysis of decision tree and random forest. *Comput. Electron. Agric.* **169**, 105221 (2020).
34. Maciejak, T. R., Kukawska-Tarnawska, B., Tyszkiewicz, J. & Tyszkiewicz, S. Multi-sensor odour detection and measurement of polluted food. *Pol. J. Food Nutr. Sci.* **1253**, 45–48 (2003).
35. Debska, B. & Guzowska-Świder, B. Decision trees in selection of featured determined food quality. *Anal. Chim. Acta* **705**, 261–271 (2011).
36. de Schaumlöffel, L. S., Dambros, J. W. V., Bolognese Fernandes, P. R., Gutterres, M. & Piatnicki, C. M. S. Direct and simultaneous determination of four phenolic antioxidants in biodiesel using differential pulse voltammetry assisted by artificial neural networks and variable selection by decision trees. *Fuel* **236**, 803–810 (2019).
37. Rodríguez, J. J., Kuncheva, L. I. & Alonso, C. J. Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 1619–1630 (2006).
38. Blanchet, L. *et al.* Constructing bi-plots for random forest: Tutorial. *Anal. Chim. Acta* **1131**, 146–155 (2020).
39. Asefpour Vakilian, K. & Massah, J. An artificial neural network approach to identify fungal diseases of cucumber (*Cucumis sativus* L.) plants using digital image processing. *Arch. Phytopathol. Plant Prot.* **46**, 1580–1588 (2013).
40. Manjakkal, L., Sakthivel, B., Gopalakrishnan, N. & Dahiya, R. Printed flexible electrochemical pH sensors based on CuO nanorods. *Sens. Actuators B Chem.* **263**, 50–58 (2018).
41. Simic, M., Manjakkal, L., Zaraska, K., Stojanovic, G. M. & Dahiya, R. TiO$_2$-based thick film pH sensor. *IEEE Sens. J.* **17**, 248–255 (2017).
42. Kong, Q. *et al.* Machine learning in seismology: Turning data into insights. *Seismol. Res. Lett.* **90**, 3–14 (2019).
43. Wang, J. X. *et al.* Zinc oxide nanocomb biosensor for glucose detection. *Appl. Phys. Lett.* **88**, 233106 (2006).
44. Lin, T. K. Fabrication of interdigitated electrodes (IDE's) by conventional photolithography technique for pH measurement using micro-gap structure. 1570016883, 8–9 (2014).
45. Schreiber, M. A. *et al.* Development and validation of a cholate binding capacity method for DMP 504, a bile acid sequestrant. *J. Pharm. Biomed. Anal.* **25**, 343–351 (2001).
46. Demšar, J. *et al.* Orange: Data mining toolbox in python. *J. Mach. Learn. Res.* **14**, 2349–2353 (2013).
47. Nonlinear Time Series Analysis—Holger Kantz, Thomas Schreiber—Google Books.

Acknowledgements

Author contributions

R.F. designed the experiments, developed the machine learning module and co-wrote the manuscript. A.K.A., A.B., M.A. and N.K. fabricated the sensor and conducted the sensing experiments. P.M. and T.F. conducted the machine-learning analysis. P.S. did part of the machine learning based data analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-03674-1.

Correspondence and requests for materials should be addressed to R.E.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.