

Advancing Biosensors with Machine Learning

Feiyun Cui, Yun Yue, Yi Zhang, Ziming Zhang, and H. Susan Zhou*

Cite This: <https://dx.doi.org/10.1021/acssensors.0c01424>

Read Online

ACCESS |



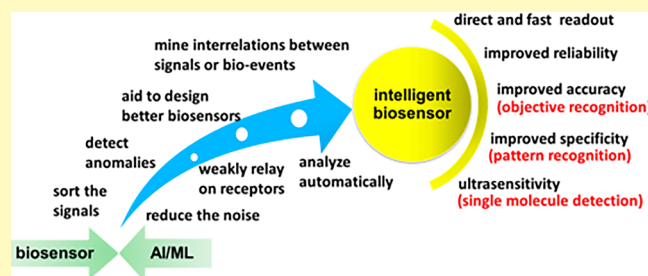
Metrics & More



Article Recommendations

ABSTRACT: Chemometrics play a critical role in biosensors-based detection, analysis, and diagnosis. Nowadays, as a branch of artificial intelligence (AI), machine learning (ML) have achieved impressive advances. However, novel advanced ML methods, especially deep learning, which is famous for image analysis, facial recognition, and speech recognition, has remained relatively elusive to the biosensor community. Herein, how ML can be beneficial to biosensors is systematically discussed. The advantages and drawbacks of most popular ML algorithms are summarized on the basis of sensing data analysis. Specially, deep learning methods such as convolutional neural network (CNN) and recurrent neural network (RNN) are emphasized. Diverse ML-assisted electrochemical biosensors, wearable electronics, SERS and other spectra-based biosensors, fluorescence biosensors and colorimetric biosensors are comprehensively discussed. Furthermore, biosensor networks and multibiosensor data fusion are introduced. This review will nicely bridge ML with biosensors, and greatly expand chemometrics for detection, analysis, and diagnosis.

KEYWORDS: intelligent biosensor, wearable electronics, SERS, artificial intelligence (AI), machine learning (ML), deep learning, chemometrics, sensing data, CNN, multidimensional features



Biosensors are a type of detection or diagnostic device.^{1–3} Compared with conventional or larger analytical instruments, biosensors have the advantages of speed, low cost, nondestructive property, and on-site detection. They have been extensively used in fundamental bioresearch,^{4–7} food safety,^{8,9} environmental monitoring,^{10–12} disease diagnosis,^{13–17} and drug screening.^{18–20} In the past decades, with the extensive progression of nanotechnology, signal amplification strategies, and transducers, biosensors have been substantially advanced. However, all biosensors inevitably have some irregular signal noise. Some biosensors heavily rely on antibodies or aptamers as bioreceptors,^{21,22} which restricts the biosensors to short shelf-life and poor stability. The accuracy and reliability of most current biosensors limit their commercialization. Hence, researchers are looking for breakthroughs in other aspects to improve the performance of biosensors. Herein, the analysis of sensing data based on machine learning (ML) is in focus. ML can provide novel strategies for overcoming the challenges faced by biosensors, and it also can be the way through that common biosensors become intelligent biosensors, which can automatically predict species or concentration of analyte based on a decision system.

Chemometrics belong to the chemical discipline which employs statistical or mathematical methods, (a) to interrogate maximum chemical information by analyzing chemical data, and (b) to select or design optimal experiments and measurement procedures.²³ Chemometric methods have achieved extensive acceptance in analytical chemistry.^{24,25} It

can be one of the strategies or tools to overcome the challenges of biosensors. Applications of chemometrics in quantitative and qualitative processing of complex signals generated from electrochemical,^{26–29} optical,^{30–32} colorimetric,^{33,34} and other biosensors^{35,36} were extensively reported. Numerous chemometric methods have been reported; the most well-known are principal component analysis or regression (PCA or PCR), linear discriminant analysis (LDA), multiple linear regression (MLR), partial least-squares discriminant analysis or regression (PLSDA or PLSR), hierarchical clustering analysis (HCA), and their combination. These chemometric methods have been comprehensively reviewed in previous works.^{37–39} Nowadays, many advanced ML algorithms for processing data are emerging, such as κ -nearest neighbor (κ NN), support vector machine (SVM), Naive Bayes (NB), decision tree (DT), gradient-boosted trees (GBT), random forest (RF), Feedforward artificial neural network (Feedforward ANN), recurrent neural network (RNN), and convolutional neural network (CNN). Some of them are well involved in biosensors, while others have remained untouched by the biosensor community. The merits of advanced ML methods over conventional

Received: July 10, 2020

Accepted: October 29, 2020

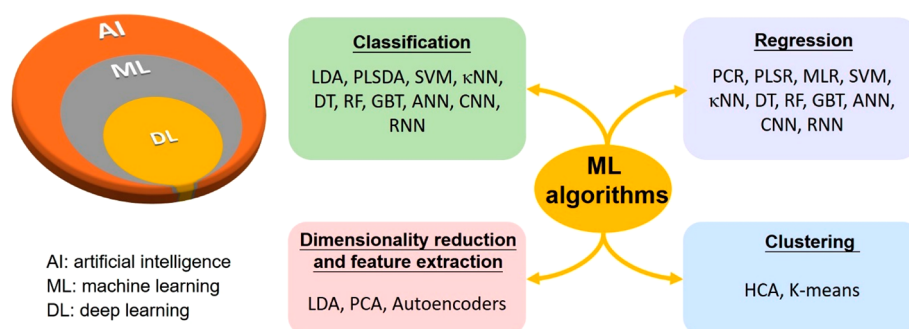


Figure 1. Relationship between AI, ML, and DL, and various ML algorithms involved in this review. CNN: Convolutional neural network. DT: decision tree. Feedforward ANN: Feedforward artificial neural network. GBT: gradient-boosted trees. HCA: hierarchical clustering analysis. κ NN: κ -nearest neighbor. LDA: linear discriminant analysis. MLR: multiple linear regression. NB: Naive Bayes. PCA or PCR: principal component analysis or regression. PLSDA or PLSR: partial least-squares discriminant analysis or regression. RF: random forest. RNN: recurrent neural network. SVM: support vector machine.

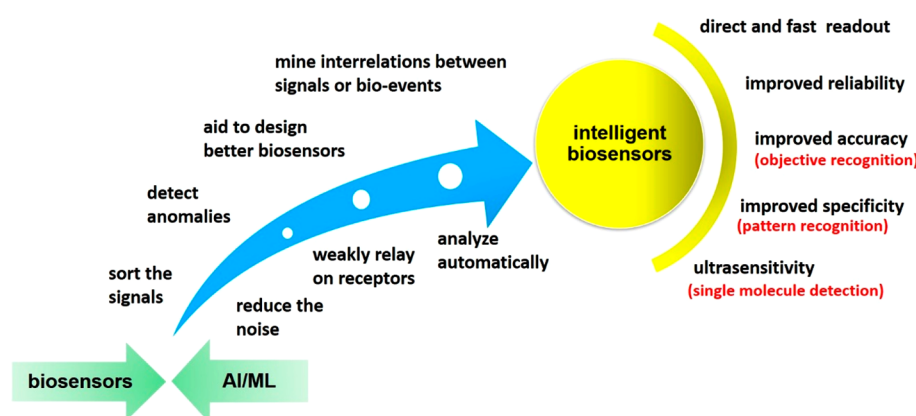


Figure 2. Benefits of ML brought to biosensors.

approaches are their capacity to interrogate appropriate nonlinear dependencies for complex biological samples, offering the inimitable possibility of solving pressing challenges in the area of biosensors. The relationship between AI, ML, and DL, and various ML algorithms is presented in [Figure 1](#). More detailed introduction of ML algorithms is presented in [section 3](#).

The purpose of this review is to present timely discussion and perspectives of advanced ML and their applications in biosensors. Various ML algorithms will be systematically introduced and their applications in diverse biosensors will be emphasized. With the assistance of ML, chemometrics will be expanded and biosensors can become intelligent biosensors. One step further, the intelligent biosensors will be easily integrated into the Internet of Things (IoT).

■ HOW ML CAN BENEFIT BIOSENSORS

First, ML can effectively process big sensing data for complex matrices or samples. The other benefit of ML in biosensors includes the possibility of obtaining reasonable analytical results from noisy and low-resolution sensing data that may be heavily overlapped with each other. Moreover, proper deployment of ML methods can discover hidden relations between sample parameters and sensing signals through data visualization, and mine interrelations between signals and bioevents. Especially, ML can be used to analyze the raw sensing data from a biosensor in several ways: (1) Categorization: the sensing signals can be sorted into various categories by the algorithms based on the target analyte. (2)

Anomaly detection: biosensors are inevitably affected by sample matrix and operating conditions. When biosensors are used on-site, they can significantly interfere with contamination. ML can check the signal and answer the question “does the signal look right?” It can also “correct” sensor performance variations due to biofouling and interferences in real samples. (3) Noise reduction: noise is always included in the sensing signals. The signal from biosensors changes over seconds or minutes, while signal interference such as electrical noise can occur on the subsecond timeline. Therefore, it is possible to train ML models to distinguish the signal from the noise. (4) Object identification and pattern recognition. By discovering latent objects and patterns using ML algorithms, sensing data can be interpreted easily and effectively.^{40,41} [Figure 2](#) shows the benefits of ML to biosensors.

ML can assist biosensor readout directly, automatically, accurately, and rapidly, which is very important for on-site detection or diagnosis. A CNN algorithm-assisted optical imaging method was developed to predict the diagnostic results by Orringer group.⁴² The results can be read out in an automated fashion within 150 s. However, interpretation of the images by pathology workforce needs ~30 min.

In addition, ML has been used to design more desirable biosensors. Metamaterials with negative permeability and permittivity have been employed to amplify the detection signal of surface plasmon resonance (SPR)-based biosensors.⁴³ The preparation of metamaterials with various reflectance characteristics is critical to ensure the resonance to be useful

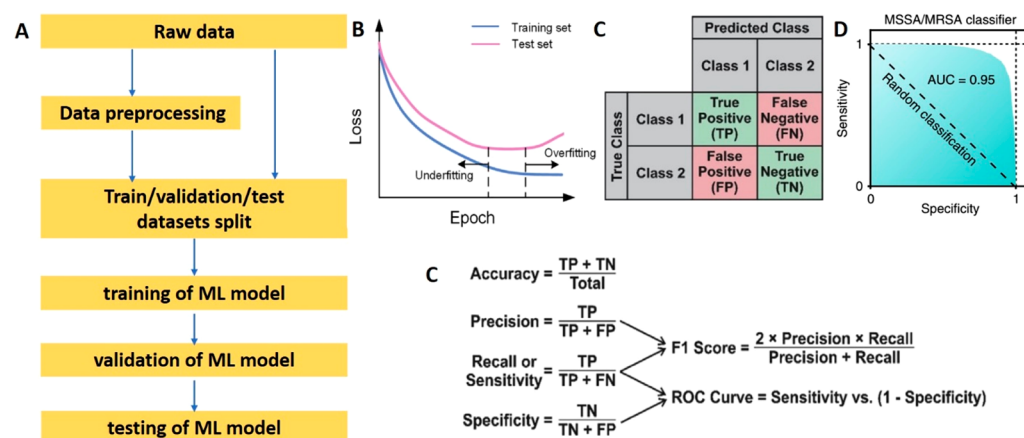


Figure 3. (A) General process of ML-based data analysis. (B) Loss curve in the training stage. Overfitting will occur when the loss of the test set increases. Reproduced with permission from ref 47. Copyright 2019 Elsevier B.V. (C) Representation of a confusion matrix, the first level metrics that can be extracted from it, and the more efficient second level metrics formulas. Reproduced with permission from ref 44. Copyright 2020 Elsevier B.V. (D) A ROC curve. Reproduced with permission from ref 50. Copyright 2019 Springer Nature Limited.

for SPR biosensors. Autoencoder (AE) and multilayer perceptron (MLP) are applied to predict the reflectance characteristics of the metamaterial SPR biosensors. Subsequently, with the dimensional reduction using t-Stochastic Neighbor Embedding (t-SNE) and AE, k-means clustering of the metamaterials was conducted. The clustering of the metamaterials can greatly accelerate researchers to design optimized sensing devices without extensive experimentation.

■ VARIOUS ML ALGORITHMS AND THEIR MERITS FOR BIOSENSORS

This section aims to introduce the general types and implemental procedures of advanced ML algorithms. Abstractly, ML is defined as a system or computer program capable of acquiring knowledge by extracting features from raw data. This newly gained knowledge can be used to make decisions to tackle real-world problems.⁴⁴ Specifically, when ML is employed in the biosensor field, it is considered as a tool or method for data processing and analysis, such as extracting features or predicting the species and concentration of the analytes. ML can be categorized into supervised learning and unsupervised learning.^{45,46} Supervised learning refers to the ML algorithms which can be trained with a group of input data with their target outputs. In the training phase, the algorithms make certain predictions on the input data set and use the given true value to improve the predictive value until the algorithms reach an acceptable accuracy. They are always used to perform classification and regression and has achieved great progress, especially for spectrometric biosensors.^{44,47} For unsupervised learning, the labeled training data set with their given outputs is not available. The goal is to determine the distribution of data set in the input space (called density estimation) or to find a set of similar examples in the input data set (known as clustering). k-Means clustering is one of the mostly known unsupervised learning algorithm.^{48,49}

General Process of ML Data Analysis. Both sequential data sets (e.g., acquired by electrical and spectral biosensors) and image data sets (e.g., acquired by colorimetric and fluorometric biosensors) exist in the biosensor field. Designing an appropriate ML model based on the data set and analysis purpose (qualitative identification, quantitative detection) is the top priority. The workflow presented in Figure 3A needs to

be implemented after the ML architecture is designed for a specific biosensor.

Commonly, the preprocessing of raw sensing data is needed. The general preprocessing methods contain derivatives, denoising, Fourier transform, and so forth. The system-specific preprocessing methods contain data compression, elimination of baseline drifts, normalization, transformations, and so forth. The application of the preprocessing method always has great influences on the total performance of ML models. For Raman spectroscopy, each spectrum needs Savitsky-Golay-smoothed, background-subtracted, and min–max scaled to [0, 1].⁵¹ For the encephalogram (EEG) signal, each signal always should be preprocessed with zero mean, standard deviation of 1, and Z-score normalization before inputting into the ML models for training and testing.⁵² The normalization makes it possible to not only compare the errors of the models but also decrease the effects of outlier samples on the training process.⁵³ The autoencoder is a unique neural network which can be applied to de-noise or reduce dimensionality.^{54,55} However, it must be pointed out that the preprocessing of raw data has no guarantee of yielding better results, since it may also remove some informative features from the raw data accidentally.

The preprocessed or raw data set should be split into three subsets, including training set (about 60%), validation set (about 20%), and test set (about 20%). The training data set is used to extract meaningful information and find optimal hyperparameters of the algorithms. The validation data set is applied when tuning hyperparameters. The test data set is employed to report the performance of algorithms. They are usually acquired from the same large data set so that they are from the same distribution.

Model weights initialization plays an important role in ML training phase. Random initialization is mostly applied, although gradient exploding, vanishing, or slow convergence might happen. Some initialization methods like He⁵⁶ and Xavier initialization⁵⁷ were developed to address the problem. When the accuracy of the testing data set is evidently less than that of the training data set, overfitting might occur. The regularization intensity should be enhanced to decrease the risk of overfitting. The loss curve is a critical indicator for reporting the training status. It can also reflect the impact of different

Table 1. Comparison of Various AI Algorithms

methods ^a	application in biosensors	advantages	drawbacks
SVM	EC biosensor, ^{112–114} Wearable electronics, ¹¹⁵ Fluorescent based biosensor ⁸¹	Works well when data have clear margin of separation in linear and nonlinear spaces; In high dimensional spaces; Promising method for analyzing overlapped signals; ¹¹³ Suitable for classification with a small sample number ¹⁶	Does not perform well when: data set is too large; target class is over two; data set is incomplete or noisy
κ NN	Wearable electronics, ¹¹⁵ Fluorescent based biosensor ⁸¹	Easy to understand and implement; No need for training; Non-parametric algorithm; Works well when features are small	Difficult to use on high dimensional data; Need to determine κ
NB	Fluorescent based biosensor ⁸¹	Easy to understand and implement; Less model complexity	Feature independence is not always satisfied.
DT	Wearable electronics ¹¹⁵	Requires less data preprocessing; Works with categorical and numerical inputs; Easy to visualize and explain	Easy to overfit; Ignores the relationship between features; Sensitive to small changes of input
RF	Electrical biosensor ¹¹⁷	More robust than single DT; Works on high dimensional data without feature selection; Can estimate variable importance; Easy to implement and can be fast trained (trees are independent)	Less interpretability than single DT; Overfitting on noisy inputs
GBT	Fluorescent based biosensor ⁸¹	Prevent overfitting; Avoid feature engineering and selection; Easy to explain and estimate the feature importance	High computation cost when the number of features is high.
Feedforward ANN	EC biosensor ¹¹⁸	NA	Long time required to build the model; Overfitting raw data often need preprocessed ¹¹⁹
CNN	Electrical biosensor ¹¹⁷ SERS-based biosensor ^{51,120,121}	Automatic feature extraction and learning; Shared weights to reduce number of parameters and memory cost; Performs well on high dimensional data; Compatibility with transfer learning ⁵¹	Need big data set, overfitting may occur if the data set is too small; Difficult to understand and interpret the final model; Variable weight and individual impact; Repeatability not guaranteed; Expensive computation; Gradient vanishing/exploding when the network is too deep
RNN	NA	Performs well on long-term dependency data; Shared weights along time dimension	Gradient vanishing/exploding when the input time dimension is too big
κ -means clustering	NA	Easy to implement and interpret; Guaranteed convergence	Results are affected by initial points; Need to determine κ ; Sensitive to outliers; High computation cost when data set is large

^aSVM: support vector machine. κ NN: κ -nearest neighbor. NB: Naive Bayes. DT: decision tree. RF: random forest. GBT: gradient-boosted trees. Feedforward ANN: Feedforward artificial neural network. CNN: convolutional neural network. RNN: recurrent neural network. NA: not available.

hyperparameters.⁴⁴ A classic loss curve shows the scenarios of overfitting and underfitting (Figure 3B).

Hyperparameter tuning is a critical task of the sensing data analysis in the validation phase. Parameters for algorithms include the number of hidden neurons, learning rate, batch size, and so forth. To discover the optimal value for each parameter, approaches including grid search, random search, or Bayesian optimization can be applied. The cross-validation method can be applied to hyperparameter tuning and evaluate the prediction performance of algorithms after parameter tuning. For instance, 5-fold cross-validation is employed to a SVM algorithm.⁵⁸

In practice, the results of a classifier can be displayed in a confusion matrix, a table representation of the actual labels to the assigned labels. This allows the extraction of the probability of true-negative (TN), true-positive (TP), false-negative (FN), and false-positive (FP). In brief, TN/TP refers to the case when the model correctly predicts the negative/positive class, while FN/FP indicates the outcome where the model incorrectly predicts the negative/positive class. In addition, various evaluation parameters can be calculated based on the probability (Figure 3C), including sensitivity (recall), specificity, accuracy, precision, F1 score, and receiver operating characteristic (ROC) curve (Figure 3D). The ROC curve shows the general overview of the model and gives an area under the curve (AUC) which represents the trade-off between sensitivity and specificity. A larger value of AUC indicates a better classifier. The ROC curve is excellent at conveying information about binary classifiers but fails to do so in multiclass models. Confusion matrices are more suitable in this case, since they help visualize how good a model is for each class. In the field of biosensing, three other metrics including the correlation coefficient (R^2), the relative error of prediction (REP), and the root-mean-square difference (RMSD) can be used to evaluate the performance of the model. They can be calculated as⁵⁹

$$\text{RMSD} = \left[\frac{1}{n} \sum_{i=1}^n (\hat{c}_i - c_i)^2 \right]^{0.5} \quad (1)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{c}_i - \bar{c}_i)^2}{\sum_{i=1}^n (c_i - \bar{c}_i)^2} \quad (2)$$

$$\text{REP}(\%) = \left(\frac{100}{\bar{c}_i} \right) \left[\frac{1}{n} \sum_{i=1}^n (\hat{c}_i - c_i)^2 \right]^{0.5} \quad (3)$$

where n denotes the total number of samples, c_i represents the real value of the analyte i , $\hat{c}_i$ is the predicted value (derived from the sensor output) of the analyte i , and $\bar{c}_i$ denotes the mean of the real value.

Besides the general process of ML, transfer learning, as an important strategy for developing ML models must be mentioned. Transfer learning refers to the situation where a model is developed using a data set and repurposed on another related data set.⁶⁰ Two approaches are mostly employed in the transfer learning: the searching and using of a developed model in algorithms library and developing a pretrained model using readily available samples.⁶¹ Insufficient data always leads to poor accuracy and reliability of deep learning methods. However, clinical samples are limited in some cases. Developing transfer learning-based deep learning models

would be of great potential. A deep learning model was developed based on a SERS data set of exosomes from lung-related cells, and then the model was transferred to predict the lung cancer stage using the SERS data set collected in patient plasma samples.⁶² The data set similarity is quantitatively evaluated by the Mahalanobis distance between cancer cell exosomes and plasma exosomes clusters. For 43 cancer patients who are in stages I and II, 90.7% patients can be accurately predicted using the transferred model. Notably, the similarity of cancer cell exosomes and plasma exosomes has a positive correlation to the stage of cancer. The results demonstrated that the transferred model can predict lung cancer using SERS of plasma exosomes. The AUC for stage I patients was 0.910, and the AUC for the whole cohort was 0.912. Following sections are some ML models related to biosensors that are widely used or with great research potentials. Their merits are listed in Table 1.

Support Vector Machine (SVM). This is designed to search for a hyperplane, which can maximize the margin between the training patterns and the decision boundary. It has been extensively used in cancer diagnosis⁶³ and waterborne pathogens detection.⁶⁴ Kernel-based SVM applies kernel functions to transform the data into a higher dimensional feature space if a data set cannot be separable linearly. The performance of the SVM is affected by two hyperparameters which are kernel parameters and kernel types.⁶⁵ The choice of kernel type is decided by the input data.

κ -Nearest Neighbor (κ NN). This is a type of algorithm mostly applied for classification.⁶⁶ A variety of approaches related to κ NN have been proposed.⁶⁷ The choice of neighbor number (κ) is based on the data set. Usually, smaller values of κ make boundaries more distinct between classes but increase the influence of the noise on the classification. Recently, the algorithm has been optimized and widely used for breast cancer diagnosis^{68,69} and anomaly detection.⁷⁰

Naive Bayes (NB). This is a probabilistic classifier which is on the basis of Bayes theorem with strong (naive) independence assumptions.⁷¹ It is the simplest Bayesian networks. Maximum likelihood is usually applied to estimate parameters in the practical applications of NB models. Higher accuracy can be achieved by coupling NB with Kernel density estimation.

Decision Tree (DT), Random Forest (RF), and Gradient-Boosted Trees (GBT). In the algorithm structure of the DT, each node 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 class distribution or value distribution. The prediction using DT is fast. However, developing a DT model requires a considerable amount of time. It is hard to operate on high-dimensional data. A single DT often encounters the problem of overfitting as the tree grows deeper. In this case, it will get an ideal training result but will potentially induce an unacceptable test error.⁷²

Ensemble methods have shown excellent performance in various ML applications. The widely used ensemble methods include RF, boosting, and bagging.⁷³ RF (also called Bagged DTs) operates by developing a number of DTs for classification and regression.^{74,75} It can address the overfitting issue and deal with noisy data well. It should be noted that, with the number increases of trees, the RF would be slow for real-time classification or prediction. Hence, weak and uncorrelated DTs are aggregated as RF.⁷⁶

Boosting is a powerful learning strategy for improving the prediction accuracy by combining the output of many weak learners with weighting and applying the learners repeatedly in a series.⁷⁰ GBT is another ensemble ML algorithm that includes AdaBoost, XGBoost, CatBoost, and LightGBM.^{77–79} It has the advantage of high accuracy, small model size, and a fast training and prediction process.⁸⁰ In GBT, trees are trained sequentially to compensate for the residual of the preceding tree. The maximum depth of each tree, the minimum number of leaves in one tree, and the minimum number of data points per leaf can be employed to avoid overfitting and maximize accuracy.⁸¹

Feedforward Artificial Neural Networks (ANN). Usually, three layers (input, hidden, and output layers) are included in an ANN structure. It consists of connected neurons (nodes) which is designed to imitate the human brain. The nodes process the input signals and transmit them to the next connected nodes. The output of the nodes is subject to the weighted sum given by the nodes of former layer. Deep learning, a branch of ML characterized using deep ANN, has been greatly developed. Open-source deep learning frameworks developed by institutions ranging from universities (Theano) to Google (TensorFlow), Microsoft (CNTK), and Pytorch now provide scientists and engineers with access to deep learning tools. The size of the hidden layer is one of key parameters which affect the performance of the ANN significantly.⁸²

Convolutional Neural Network (CNN). CNN belongs to a type of deep learning which is good at image analysis such as X-ray images,⁸³ magnetic resonance images,⁸⁴ and computed tomography (CT) images.⁸⁵ The CNN model usually including three layers: (1) Convolutional layer: Filters (kernels) that slip across preprocessed signals are included in this layer. Stride controls how the filter shifts around the input image. The feature map can be achieved after the convolution step. (2) Pooling layer: It also can be called the down-sampling layer. The output of convolutional layer needs to reduce its dimension by pooling operation which can prevent overfitting and decrease the computational intensity. (3) A fully connected layer: Activation functions (such as Sigmoid, Tanh, Relu, LeakyRelu,⁸⁶ and Softmax⁸⁷) are usually used to introduce nonlinearity into the output. Although the CNN is originally developed for 2D image recognition, the 1D CNN⁸⁸ achieved great advances, which have been shown to perform well on data having spatial relations in a single dimension, such as stock price,⁸⁹ electroencephalograms (EEGs),⁹⁰ audio signals,⁹¹ and spectra.⁹² Lussier et al.⁹² developed a 1D CNN model which include two layers of each convolutional, pooling, and densely connected neurons structure to analyze molecular spectra for multiplexing SERS sensing. Softmax function was selected to transfer the final output and converted the them into probabilistic values. The highest probability was assigned to a positive value of 1, while all others were assigned to a negative value of 0. 1000 SERS spectra were acquired and randomly split into training set (60%), validation set (20%), and test set (20%). The SERS spectra were preprocessed and labeled. Bias and weights were initialized with random values. Ten times were repeated for the training and testing operations to check the reliability and reproducibility of the model.

Adoption of the Inception module⁹³ in CNN, also named the DeepSpectra model, was developed to quantitatively analyze the one-dimensional spectra.⁹⁴ In the Inception module, simultaneous convolutional operations with different

filter sizes are allowed, which makes the depth and width of DeepSpectra retain large values and the complexity of computational unchanged. Both high- and low-level features of raw spectra can be extracted in the deeper network. The dimension reduction using PCA can be omitted. With the expansion the network can improve its adaptability to different feature maps. Compared with a conventional CNN model includes three convolutional layers and two pooling layers, the DeepSpectra achieved good performance with both a small data set (data from 80 samples, 700 features/data) and a large data set (data from 3793 samples, 2151 features/data). However, the DeepSpectra model showed a comparable performance with the PLS model for the small data set (data from 80 samples, 700 features/data). It revealed that the CNN-based spectral analysis needs a larger sample size to get excellent repeatability and accuracy.

Recurrent Neural Networks (RNN). Among the different deep learning methods, RNN have drawn researchers' attention in sequential data related studies.⁹⁵ RNN is well-suited for time-series or sequential data because the network structure is specifically designed to represent historical information in each recurrent round.^{96,97} Due to the attribute of propagating past information along time through recurrent connections, RNN is widely applied for sequence mapping problems such as sequence generation,⁹⁸ speech recognition,^{99,100} handwriting recognition,¹⁰¹ and reinforcement learning.^{102,103} Biomedical researchers have applied RNN in detecting the interactions between genes and proteins.^{104–106} Recent studies have achieved promising brain tumor segmentation performance by training RNN.¹⁰⁷ Long Short-Term Memory (LSTM) networks are a special type of RNN with a capability for long-term dependencies.¹⁰⁸ Bidirectional RNN with LSTM was developed to detect DNA modifications.¹⁰⁹ RNN based algorithms also have improved nanopore sequencing read accuracy.^{110,111}

■ DIFFERENT BIOSENSORS WITH ML

Electrochemical (EC) Biosensors. These are one type of widely used biosensor. Combining EC biosensors and traditional chemometrics was reviewed by Ni and Kokot in 2008.¹²² However, the use of novel ML methods in current EC biosensors is still in the inception stage. Although the relatively elaborated theoretical backgrounds of electrochemistry grant description of large variety of signals, EC biosensors are not very reproducible or stable in real sample detection. Real samples may have many interferants, in a wide window of ionic strength, temperature, pH, and so forth. Another reason is that the electrode or modified electrode used in EC biosensors fouls with time. Hence, one-dimensional data analysis is not enough to acquire sensitive signals highly correlated with the analyte type and quantity. This highlights an emerging opportunity for coupling ML with EC biosensors to study how ML can be used to improve the accuracy and reliability of sensor in real sample measurements.

SVM regression model was used by Massah and co-workers to improve the performance of cyclic voltammetry-based portable EC biosensor.¹¹² For predicting the concentration of nitrate, different kernel types including linear, polynomial, and Gaussian with various parameters were applied. The correlation coefficient (R^2) and mean squared error (MSE) were used to estimate their performance. The results showed that the polynomial kernel with the kernel parameter at $\gamma = 0.20$ was best, with the MSE of 0.0016 and R^2 of 0.93. By

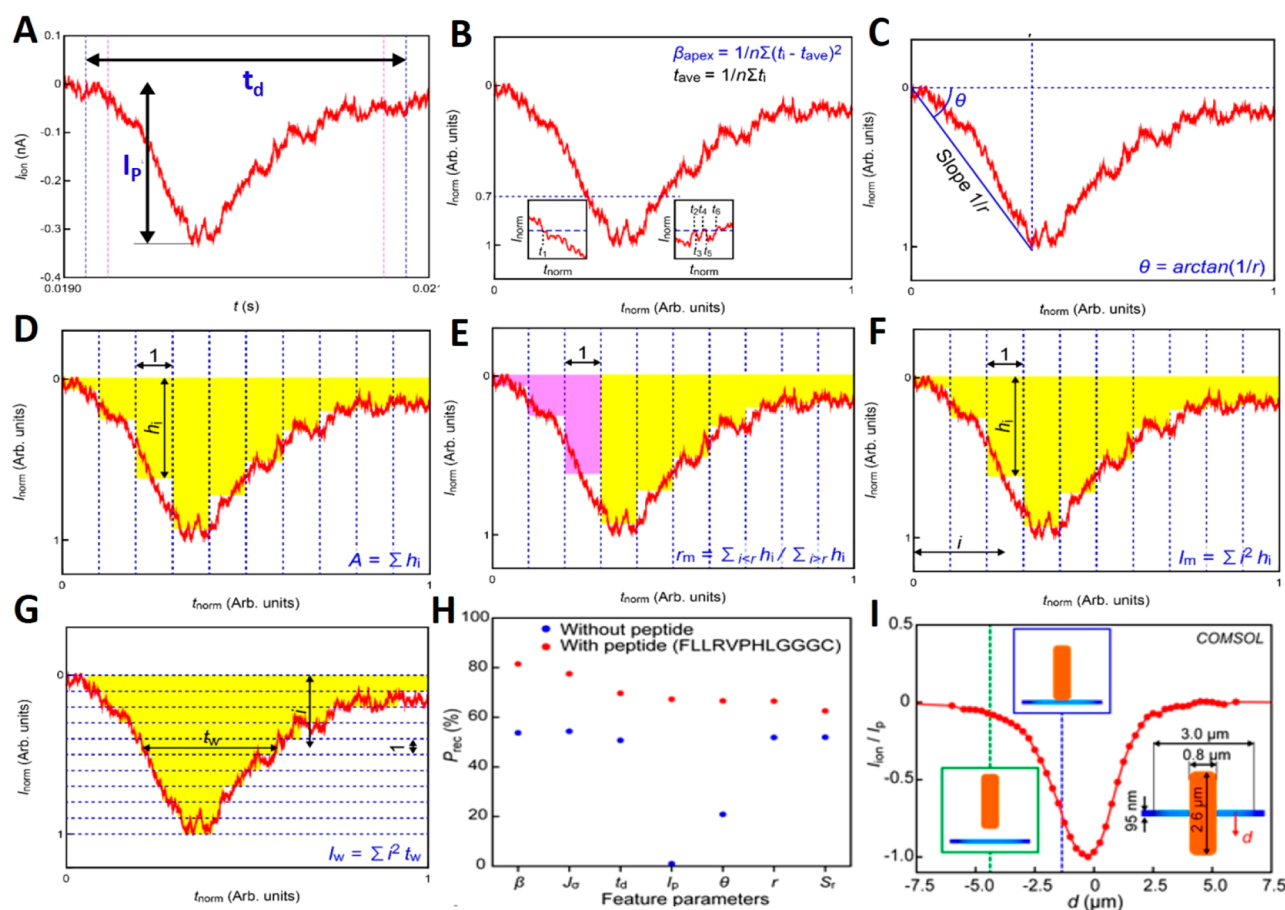


Figure 4. Multidimensional features extracted from the current–time waveform. Reproduced with permission from refs 136 and 137. Copyright 2017 Springer Nature Limited and 2018 The American Chemical Society. Pulse height I_p and width t_d (A), the bluntness of resistive pulse apex β_{apex} (B), the onset angle θ (C), the area A (D), the ratio r_m of area before current peak to that of behind the current peak (E), the inertia I_m (F) and I_w (G). The precision analysis of the various features for identifying the Δ fliC and wild-type *E. coli* based on multiphysics simulation of the waveform during the bacteria passing through the nanopore biosensor (I).

applying the SVM regression model, the service life of the EC biosensor was improved, which can work 10 days after immobilization of the enzyme. 400 samples (nitrate) and even more can be detected without the replacement of the enzyme. Gonzales-Navarro and co-workers¹¹³ compared four types of ML regression models to enhance the stability and accuracy of an **amperometric** glucose-oxidase biosensor (GOB) in the uncertain circumstance. The radial basis function-based SVM (SVM-R), linear kernel-based SVM (SVMR-L), the Levenberg–Marquardt backpropagation method-based ANN and PLS were used. The SVM-R model was demonstrated as an excellent ML model for improving performance of the amperometric biosensor.

Electrochemical impedance spectroscopy (EIS) are popular among EC biosensors.^{123,124} The equivalent circuit models are always applied to extract key parameters of EIS data with χ^2 testing. The extracted parameters, such as electron transfer impedance (R_{et}) and capacitance, are employed to indicate the binding events on the working electrodes. However, selecting or designing an equivalent circuit model for fitting the EIS data from complex electrode geometries or complicated solutions is challenging.¹²⁵ Especially, for the small molecule–protein interaction-based EIS biosensors, the equivalent circuit model analysis was not well documented. In this case, Rong et al. developed a SVM model to analyze the EIS data without equivalent circuit fitting.¹¹⁴ SVM models with

four different kernels (polynomial, sigmoidal, linear, and radial basis function) were compared to find the optimized ML model. 80% of 54 EIS data were randomly selected as training data set and the other 20% data were attributed to the testing data set. The SVM with radial base function kernel was demonstrated to have the optimal performance for classifying the training data set with the accuracy of 98%. The nonlinear kernel coefficient (γ) and penalty parameter (C) were tuned to improve the performance of radial base function-based SVM with the optimum value of 0.01 and 10, respectively. Ali et al.¹¹⁸ applied back-propagation ANN, maximum likelihood, and LDA to classify *Escherichia coli* (*E. coli*) strains JM109, DHS- α , and *Salmonella typhimurium* by their impedance features. The impedance signals were measured by interdigital silver electrodes and silver nanowires uniformly decorated on the electrodes on a polyamide-based polyethylene terephthalate substrate. 40 samples of each bacteria were measured; each measurement contained 251 data points that were represented by a vector. All three of these ML algorithms achieved a 100% accuracy to classify these bacteria.

As of now, a deep-learning aided EC biosensor has not yet been reported. One possible reason may be the limited number of available data sets. Generally, EC biosensors rely on a bioreceptor (antibody or aptamer) for capturing analyte and rely on nanomaterials for signal amplification. The need to measure a large amount of data for detection is small.

Developing arrayed or multiplexing EC biosensors to test a large number of real samples (e.g., clinical specimens) will bring opportunities for the application of deep learning in EC biosensors.

Combining the **single-molecule (SM) electrical biosensors** and ML can improve the accuracy and precision of SM identification. The combination can also quantitatively evaluate the molecular recognition ability and optimize the design parameters of the electrical biosensor device.¹²⁶ Methods for SM electrical detection can be divided into two categories, roughly: nanogap and nanopore.^{127–129} A maximum current (I_p) and a current duration (t_d) from the tunneling current–time waveform or the ion current–time waveform are used as signals to identify the analyte. They are widely applied to DNA, RNA, carbohydrate, and peptide sequencing^{130–132} and virus detection.^{133,134} However, analytes with little difference of molecular volumes and frontier orbital energy have similar I_p and t_d signals. The overlapping of current signals cannot meet the detection and identification of multiple analytes. This challenge is overcome by analyzing the current–time waveform using ML methods, such as SVM, RF, and CNN.¹¹⁷ SVM-based analysis of tunneling current–time waveform was used to identify single amino acids by nanogap biosensor. Different from just applying I_p as a signal feature, two features obtained by fast Fourier transform (FFT) of each of the obtained waveforms were used to train the SVM algorithm and develop a two-dimensional map plot that can clearly discriminate the chiral enantiomers D-asparagine and L-asparagine, glycine (Gly), and N-methylglycine (mGly), and the isobaric isomers leucine and isoleucine. Their prediction accuracies improved from 53% to 87%, 55% to 95%, and 51% to 80% with the assistance of the SVM, respectively.¹³⁵ Since many bacteria have a similar morphology and size for the solid-state pore biosensors, the overlapping of ionic current spikes induces a poor single-bacterial cell detection. Kawai and co-workers¹³⁶ demonstrated that the Rotation Forest model can discriminate the electrical signatures from similar microbial shapes. Besides the width t_d and height I_p of the current waveform (Figure 4A), other easily overlooked features can be extracted for identifying the species of bacteria (*E. coli* and *Bacillus subtilis*). They are the bluntness of resistive pulse apex β_{apex} (Figure 4B), the onset angle θ (Figure 4C), the area A (Figure 4D), the ratio r_m of area before current peak to that of behind the current peak (Figure 4E), the inertia I_m (Figure 4F) and I_w (Figure 4G) calculated associated with abscissa and ordinate. 60 features were extracted by coupling the time vector and the current vector. The total extracted features of 161 waveforms of *B. subtilis* and *E. coli* (in total 322 spikes) were employed as training data to predict the other 18 resistive waveforms as the test. The results showed that the discrimination accuracy of single bacteria was more than 90%. The same group developed a peptide-functionalized solid-state pore biosensor to discriminate the flagellin-deletion mutant (ΔfliC) and wild-type of the *E. coli* using the similar data analysis method.¹³⁷ Several features, such as the onset angle θ , pulse peak position r , pulse bluntness β , inertia $J\sigma$, and pulse area ratio S_r , are extracted and contribute to the precision (P_{rec}) of bacteria discrimination personally (Figure 4H). Multiphysics simulation was used to interpret the features chemically and physically (Figure 4I).

Wearable Electronics. Wearable electronics have a wide range of applications in basic biomedical research and clinical medicine, including human–machine interfaces,^{138,139} artificial

skins,^{140,141} disease diagnosis,^{142–144} and health monitoring.¹⁴⁵ Recently, wearable electronics, such as electronic tattoos (E-tattoos)¹⁴⁶ and epidermal electronics systems (EES),¹⁴⁷ and flexible electrochemical bioelectronics¹⁴⁸ have been widely reported to monitor various physiological signals in real time. Three different ML algorithms, including κ NN, DT, and SVM, were applied to mine interrelations between the fatigue levels and physiological signals, and then predict fatigue states of workers based on physiological signals acquired by a multimodal EES.¹¹⁵ Two modules are included in the EES. One module contains a strain sensor and three flexible electrodes. It can be attached to the chest and used to monitor the respiration and rate electrocardiogram (ECG). Another module contains two flexible electrodes which can be pasted on the palms and applied to detect the galvanic skin response (GSR). Hence, related physiological signals can be monitored by the developed EES, and features can be extracted from the electrophysiological signals. Then, the features are fed into the proposed ML models to find an optimized algorithm and predict the fatigue levels. The results demonstrated that the DT models showed the highest predictive accuracy with the value of 89%. The signal of single-molecule (SM) electrical biosensors and wearable electronics is almost time-series sequential data. RNN will be a good choice for processing these data, since the network structure is specifically designed to represent historical information in each recurrent round.

The absence of the necessary volume of data to train ML algorithms represents one of the major challenges of coupling ML with wearable electronics for health monitoring, disease diagnosis, and replicating human sensory functions. This challenge should be considered in the initial sensor design phase to build reliable sensor arrays, thereby collecting a large volume of data set. Recently, Sundaram et al. integrated a wearable tactile (touch-based) glove sensor array with CNN to identify individual objects and estimate their weights of unknown objects.¹⁴⁹ More specifically, an array of piezoresistive sensors (548 sensors) were fabricated and assembled on a knitted glove, and a ResNet-18-based architecture was established to identify objects and/or estimate their weights using large databases of detailed pressure information collected from the tactile sensors. The developed tactile sensor array and the integration with deep-learning models highlight the emerging applications of intelligent sensors for understanding the role of touch and replicating human sensory functions in active prosthetics and robotics. In addition to these emerging applications, integrating wearable electronics with ML for continuous monitoring of temperature, blood oxygenation, and respiratory biomarkers (cough frequency and intensity and respiratory rate) of COVID-19 patients highlights another important application of intelligent sensors for addressing the current public health challenge. For example, researchers from Prof. John A. Rogers's lab at Northwestern University developed a wearable sensor with a high-bandwidth accelerometer and a temperature sensor to catch early signs and symptoms associated with COVID-19 patients, including temperature, coughing intensity and patterns, and heart rate.¹⁵⁰ These biophysical measurements provide important insights into the physiological status of COVID-19 patients in a continuous and real-time manner. Importantly, coupling this physiological data with state-of-the-art ML techniques will create a valuable platform to detect COVID-19 infections, to predict disease severity and fatal outcome of COVID-19, and to provide guides for reopening the economy.

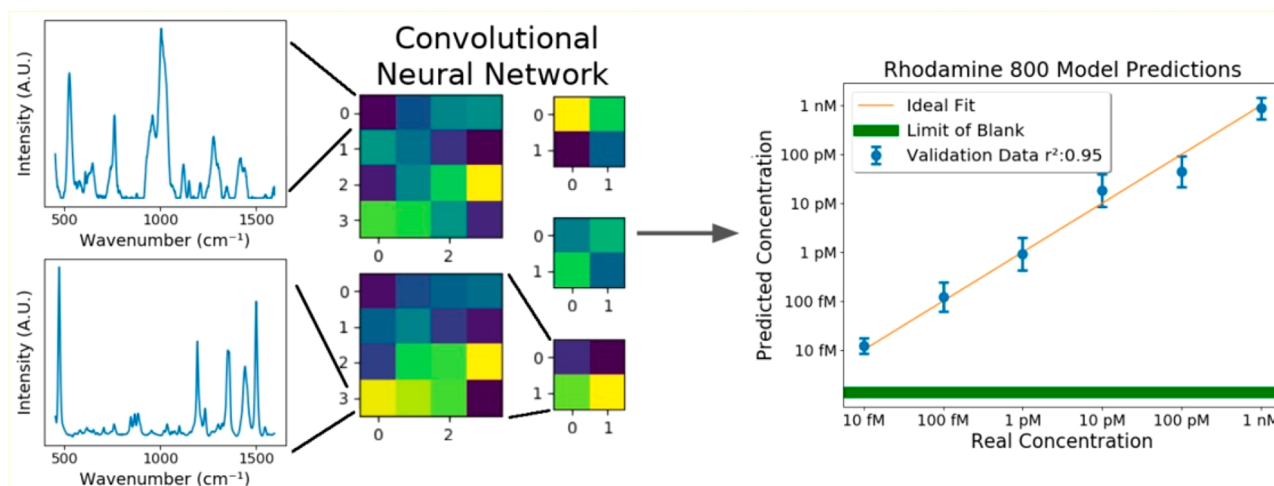


Figure 5. Preprocessed SERS spectrum of Rhodamine 800, SERS spectrum was converted into pixels which were bundled into pixel maps with the size of 8×8 , the CNN model converted spectrum into concentration value. Reproduced from ref 51. Copyright 2019 The American Chemical Society.

SERS and Other Spectra-Based Biosensors. Surface enhanced Raman spectroscopy (SERS) can acquire intrinsic fingerprint information on an analyte in a complex matrix. SERS sensing is one of most promising analytical tools for the rapid, label-free, on-site, and nondestructive detection.^{151,152} However, many analytes and the substance in the matrix have similar or overlapping spectra. It is tedious or impossible to manually distinguish them. Hopefully, the application of ML can significantly improve the effectivity of SERS. The uniformity of the enhancement factor of the SERS substrate is essential for ML methods, as large variance in the data set increases the variance in predictions, which limits the methods to semiquantitative or quantitative analysis.⁹² Among various ML methods, CNN always shows a better prediction accuracy with medium or large data sets.¹⁵³ Hence, CNN are now the most popular for spectral analysis.^{154,155} A practical guide for CNN applied in the spectral analysis was reported by Ying's group.⁴⁷

A CNN-assisted SERS biosensor was developed to detect the oligonucleotide (OND) damage on a gold gratings substrate.¹²⁰ A portable spectrometer was applied to collect the SERS spectra of OND by different operators without the optimization of test conditions (such as optimal location on substrate, laser intensity, acquisition time, and manual baseline correction). In their CNN structure, a new way of feature extraction, named binary stochastic filtering (BSF), was introduced. BSF would evaluate the significance of every inputted feature to identify important areas in the original spectrum. The proposed SERS-CNN method can identify very small DNA damage that hardly can be detected by other technologies. Their results showed that the accuracy of OND damage classification was up to 98% with the confidence level of more than 95%. The same group then developed an improved SERS-CNN method to identify the normal and cancer cells by detecting cell cultivation media.¹²¹ Gold multibranched nanoparticles (AuMs) functionalized with different chemical groups were used to capture the biomarkers and enhance Raman signals. The gold grating surface was also used to form a plasmonic coupling effected to obtain higher resolution SERS spectra. However, the challenge is that the signal of interference increased with the target analyte, which subsequently decreased the detection reliability. A combina-

tion of the SERS and CNN approach was applied to overcome this challenge. With a similar but modified CNN algorithm, the results showed that the prediction accuracy was 100% for data validation.

Combining SERS biosensors and ML methods for single molecular and single-cell analysis is highly desirable. Thrift and Ragan⁵¹ presented an CNN-based SERS for quantification of subnanomolar Rhodamine 800 concentrations (Figure 5). Obtain the spectrum in the SM concentration range, and transfer each SERS spectrum into a pixel, and bundle them into pixel maps with the size of 8×8 for training the proposed CNN model. It is very interesting that the non-analyte and signal noise can be easily discerned by the CNN model, which has distinctly enhanced the detection accuracy. The CNN model also can convert a spectral signal into a concentration value based on the deviations of the Langmuir isotherm. Raman spectroscopy can also identify microorganisms at the single-cell or single-particle level. Combining a large Raman data set of microbes with ML can produce more accurate identification results. Laser tweezer Raman spectroscopy along with a CNN model was reported for identifying species or subtypes of microorganisms at a single-cell level.¹⁵⁶ A new occlusion-based method for feature extraction was proposed to process the Raman spectroscopy. Different species or subtypes of microorganisms were distinguished based on the weights of features. The overall accuracy of identification was $95.64 \pm 5.46\%$.

A fluorescent array-based differential sensing platform, also described as a chemical nose, was developed to identify eight different proteins at 100 nM concentration without a bioreceptor. Compared with 83% efficiency using the LDA, κ NN, GBT, SVM, and LR with optimization parameters achieved a 100% pattern recognition accuracy.⁸¹

Fluorometric and Colorimetric Biosensors. This section focuses on the fluorometric and colorimetric biosensors using images as detection signals. The automated identification of colors and their intensity from these biosensing images is of significant interest. Digital polymerase chain reaction (dPCR) is included as a type of fluorometric biosensor. Lateral flow assay (LFA), paper-based vertical flow assay (VFA), and other colorimetric strips are included as colorimetric biosensors.

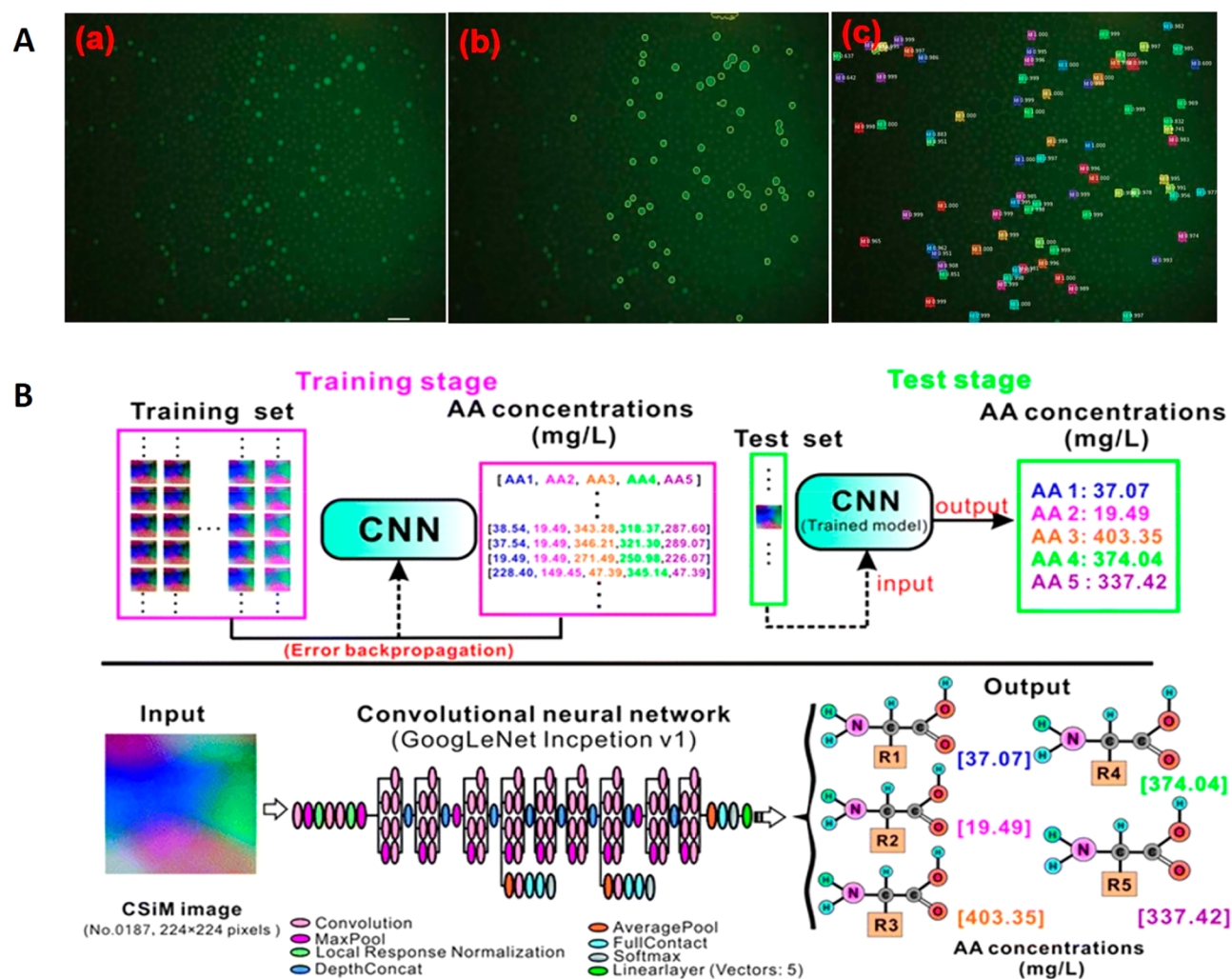


Figure 6. (A) Analysis of uneven light images by Mask R-CNN model and threshold segmentation. Reproduced with permission from ref 158. Copyright 2019 The Royal Society of Chemistry. (a) Image with uneven light from the real experiment. (b) Results of threshold segmentation. (c) Results of the Mask R-CNN model. (B) Developed CNN model for mixed AA analysis. Reproduced with permission from ref 172. Copyright 2020 The Royal Society of Chemistry.

Fluorescence imaging-based dPCR is a promising technology for gene diagnostics.¹⁵⁷ Accurate and quick recognition of the positive reaction chamber in the fluorescence image is critical to applying the dPCR to the actual application. Traditional methods such as threshold segmentation, numerical clustering, and grid location have been used in the images analysis.¹⁵⁸ The most used image processing method is threshold segmentation.¹⁵⁹ However, it is needed to tune the parameters of the threshold segmentation in each analysis. It is also limited to analysis of images with uneven brightness induced by poor camera imaging or nonuniform illumination. The light intensity is always unevenly distributed in the real testing surroundings. This situation can cause a low accuracy of the positive reaction chamber recognition. Recently, Mask R-CNN¹⁶⁰ was deployed to analyze the images precisely and automatically.¹⁵⁸ A real image showed that there was a dark area on the left side (Figure 6A-a) where the microreaction chambers including the positive one have weaker brightness compared with other areas. Figure 6A-b and Figure 6A-c showed the results based on threshold segmentation and the proposed Mask R-CNN model, respectively. The threshold segmentation method read out 56 bright spots from 82 positive reaction chambers with 2 misclassified spots. The true positive

rate was 68.29%. What is exciting is that the Mask R-CNN method read out 80 bright spots without a zero false-positive rate. Its true positive rate was 97.56%. Besides the dPCR, the developed CNN model has promising application in digital enzyme linked immunosorbent assay techniques (digital ELISA),¹⁶¹ fluorescent microarrays,^{162,163} and location surface plasmon resonance imaging¹⁶⁴ technologies for molecular diagnosis and biological detection.

Combining colorimetric biosensor and smartphone reader is more attractive for point-of-care testing (POCT). Smartphone applications (Apps) and cloud-based ML models pave the way for quantitative colorimetric detection with higher accuracy and repeatability.^{165,166} Recently, a smartphone based colorimetric biosensor was developed to monitor water contamination by bacteria.¹⁶⁷ CNN was used to classify the presence or absence of bacteria based on color intensity of the biosensor images. The reported method achieved an improved accuracy of 99.99% for predicting the presence of *E. coli*. Three different ML methods (ANN, SVM, and LDA) were explored using a smartphone-based colorimetric biosensor to detect concentrations of saliva alcohol. Four-color spaces (HSV, RGB, Lab, and YUV) of images were evaluated by the ML methods. A major color change appeared in the blue channel,

while the optimal differentiability and sensitivity occurred in the green channel of the RGB color space. The ANN model showed optimal performance with LAB color space as features. In all, the overall classification rate for standard concentrations was 100% and for enhanced concentrations was 80%.¹⁶⁸ A vertical flow assay (VFA)-based colorimetric sensor powered by deep learning was reported for C-Reactive Protein (CRP) detection.¹⁶⁹ The deep learning method was used to optimize the configuration of immunoreaction spots and predict the concentration of CRP. It showed a variation coefficient of 11.2% and a correlation coefficient (R^2) of 0.95 for blind VFAs test in the CRP range of 0–10 mg/L. Luo et al.¹⁷⁰ adapted a CNN model into a colorimetric sensor to detect total organic carbon (TOC) in the environmental water. 697 valid various images (64 px × 64 px, RGB mode) corresponding to 697 environmental water samples was obtained and randomly segmented into a training data set of 80% and a test data set of 20%. The R^2 was 0.90314 between the predicted values and the true values of the training data set. The CNN-aided colorimetric sensor can read the result fast and avoid the usage of large instruments. The POCT tool for detection of waterborne parasite *Giardia lamblia* cysts were reported.¹⁷¹ The ML algorithm bagging (bootstrap aggregating) was used to analyze the fluorescence image. The results demonstrated that the *Giardia lamblia* cysts can be captured on the filter membrane by size and counted automatically with an overall sensitivity of 84% using the proposed ML model. The specificity was tested as 76%, 94%, 94%, and 90% for 500, 100, 50, and 10 cysts per 10 mL samples, respectively.

Besides the fluorometric and colorimetric image-based detection methods, a color spectral image-based method for mixed amino acid (AAs) analysis was reported.¹⁷² Six common CNNs, including Residual Network (ResNet), Vanilla CNN, LeNet, GoogLeNet Inception v1, VGGNet, and SqueezeNet were explored. Among these six CNNs algorithms, the developed Inception v1 model showed higher accuracy and a better astringency. The R^2 for five AAs was 0.999 with the RMSE of 10.22% (Figure 6B).

Other Biosensors and Devices. The quartz crystal microbalance (QCM)-based biosensor is one type of attractive sensing device which is gravimetrically sensitive and can detect analyte at sub-nanogram resolution.¹⁷³ The SVM classification/regression algorithm was applied to discriminate/quantify trypsin and plasmin based on frequency shift data generated by QCM.¹⁷⁴ Because the effect of trypsin and plasmin on the κ -casein was similar, the ML was needed to distinguish similar data. The results proved that the frequency shifts can be classified with more than 95% accuracy within 15–20 min, which is less than that using statistical methods. An artificial bee colony (ABC) algorithm trained ANN model was employed in QCM sensors to classify five different alcohols.¹⁷⁵ As a type of hybrid algorithms, the ABC algorithm always outputs encouraging results compared with the backpropagation (BP) algorithm. E-16 MSE level was obtained in both training and test data sets. A magnetic nanoparticle (MNP)-labeled immunochromatography test strip (ICTSs) was developed for human chorionic gonadotropin (HCG) and multiplex cardiac markers.¹¹⁶ The custom waveform reconstruction method and SVM models were used to analyze the poor signals, which greatly improved the sensitivity and accuracy. The flowchart for the data processing procedure is presented in Figure 7. Several kernel functions were compared, and the linear kernel showed a higher accuracy of 100% for all

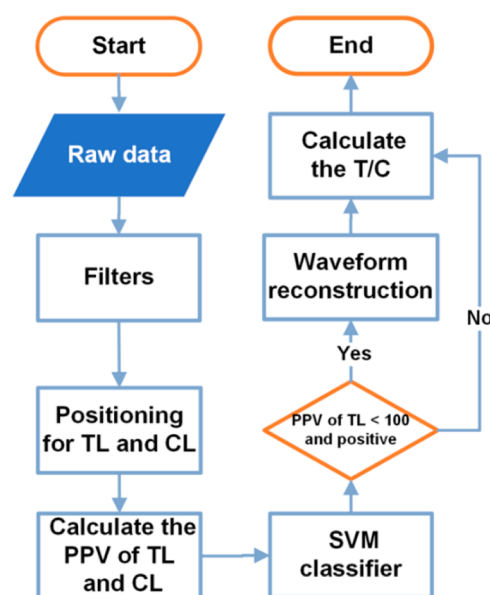


Figure 7. Flowchart for data processing procedure. Reproduced with permission from ref 116. Copyright 2019 Springer.

tested concentrations of samples (0.025, 0.5, 1, 2.5, 5, and 10 mIU mL⁻¹). However, the accuracy of the visual readings can achieve 100% only in concentrations over 2.5 mIU mL⁻¹. Hence, applying the SVM classifier can improve the sensitivity of portable ICTSs test.

Biosensors in Microfluidic Bioassay. Blood cell counting provides important indicators for fast diagnosis of disease.¹⁷⁶ Several ML-based microfluidic cytometers were reported.^{177–179} Extreme learning machine based super-resolution (ELMSR) and CNN based super-resolution (CNNSR) were compared for a lensless blood cell counting device integrating microfluidic channel and a complementary metal oxide semiconductor (CMOS) image sensor.¹⁷⁷ The cell resolution was improved four times, and CNNSR showed 9.5% improved quality over the ELMSR on resolution enhancement. Oliver and co-workers¹⁸⁰ developed a blood brain barrier (BBB) organ-on-a-chip to investigate the brain metastatic spread of breast cancer. Cellular dynamic phenotypes and features were detected by confocal tomography. The images were analyzed by eight ML algorithms, which contain neural network, NB, RF, AdaBoost latched to RF and DT, κ NN, logistic regression, and stochastic gradient descent (SGD) to predict the brain metastatic potential probability. The area under the weighted average of precision and recall (F1), accuracy (CA), and curve (AUC) were used to score these ML algorithms. The neural network (AUC = 0.951), AdaBoost (RF) (AUC = 0.950), and RF (AUC = 0.946) were ranked the top three based on their AUC. Both the positive predictive value and negative predictive value were 0.87, which can be considered excellent predictive models of clinical behavior (Figure 8).

The performance of various biosensors based on ML algorithms is listed in Table 2.

■ BIOSENSOR NETWORKS AND MULTI-BIOSENSOR DATA FUSION

Multi-biosensor synchronous measurement is important for practical applications. Fusion of sensing data from multiple biosensors directly impacts application performance. This

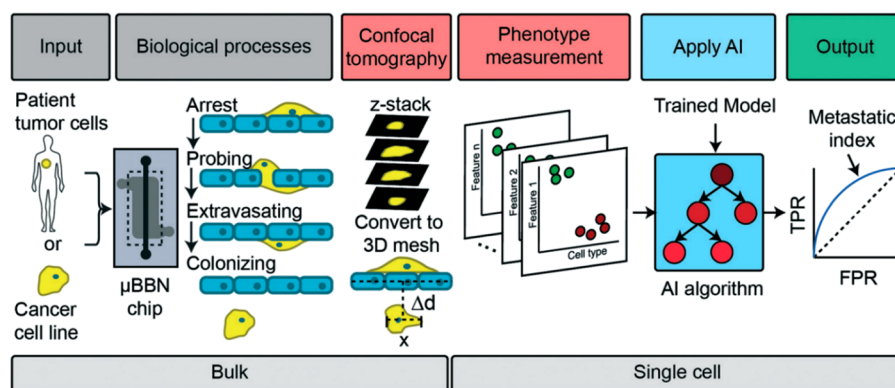


Figure 8. ML-assisted ex vivo blood brain barrier organ-on-a-chip model to investigate the brain metastatic spread of cancer. Adapted with permission from ref 180. Copyright 2019 The Royal Society of Chemistry.

concept has been applied in biomedical area, like monitoring of cardiac based on fusion of arterial blood pressure, photoplethysmography (PPG), and ECG.¹⁸⁶ Generally, the fusion of sensing data can be sorted into three types: decision-level fusion, feature-level fusion, and data-level fusion.^{187,188} The homogeneous sensing data (from biosensors with the same sensing mechanism) can be directly fused for detecting the same analytes. For the heterogeneous sensing data (from biosensors with the different sensing mechanism), decision-level fusion or feature-level fusion should always be considered.¹⁸⁹

The synchronous measurement of SPR and SERS would be very interesting, since they share similar principles of signal enhancement: local SPR of metal nanoparticles (especially Au, Ag, and Cu). Combining EIS and SERS is also attractive to get heterogeneous sensing data. The EIS is a simple, effective, and label-free method to quantitatively detect biological events. The SERS can qualitatively collect unique molecular spectroscopy of biological species.

The multisensor data fusion depends on various ML methods. Feed forward ANN, NB, and DT were used to recognize activities for Parkinson patients. Reputation-based voting and majority voting were applied (Figure 9) for sensing data fusion.¹⁹⁰ Combining covariance matrix adaptation evolutionary strategy (CMAES) with decision-level and feature-level fusion was applied to process sensing data from two electronic noses. PCA was applied to extract features, and probabilistic neural networks (PNN) were used as the classifier. The results showed that the sensor fusion has an error rate of 11%; the two individual sensors have an error rate of 13% for prediction.¹⁹¹

CONCLUSION AND OUTLOOK

From linear calibrations and nonlinear fittings to advanced ML methods for classification, regression, and clustering of complex biological samples, chemometrics provides robust mathematical tools for interpretation of sensing data. Both qualitative discrimination of complex overlapping signals and quantitative prediction of trace analytes can be improved using ML methods. Especially, the deep learning methods, including CNN and RNN, are more and more popular in sensing data analysis. Traditional data regression analysis uses a mathematical equation to calculate the dependent variables of the sample. The input features are usually less than two. On the contrary, the advanced ML models can process a database containing hundreds of input features. A sufficient data set is

essential to deep learning methods. Design and application of multiplex or high-throughput biosensors, such as microarray and multichannel fluidic chips, can help researchers to break through the data bottleneck for connecting the ML and biosensors. Databanks provided by federal agencies such as the National Institutes of Health and the Centers for Disease Control and Prevention represent another important data resource to train ML algorithms.

Compared with traditional laboratory assays, POCT usually has less reliability and accuracy. Applying ML methods into POCT represents an opportunity to study how ML can be used to improve sensor reliability and accuracy in real sample measurements. Smartphone APPs that are integrated with ML algorithms can be a greatly interesting tool for direct readout of POCT biosensors. It would extensively push the POCT biosensor to home-testing or self-testing.

Analysis of single molecular/single particle/single cell detection data is challenging, mainly limited to the poor signal-to-noise ratio, signal overlap, and dispersive signals. For single molecular sequencing biosensors, a large data set had to be analyzed. Traditional hypothesis-driven data exploration and selection may not be reasonable because unexpected signals could be missed. Developing ML methods to reduce the noise and extract multidimensional signal features can improve the resolution of pattern recognition and sensitivity of objective recognition.

Combining wearable biosensors and ML for healthy monitoring is another opportunity. Wearable biosensors have gained remarkable interest owing to their huge potential in noninvasive monitoring of human physiology by multifarious biological fluids (e.g., sweat, tears, and saliva). A desired vision of wearable biosensors is to heterogeneously integrate a series of sensor networks on flexible patches that can continuously monitor the biomarkers. ML can be used to parse the time series of multiplexed sensing data to identify the state of health. In these applications, ML has to be explainable (rather than a black-box). The machine decision must be understandable by medical professionals and decision makers. In the meantime, human knowledge and reasoning rules need to be adopted into a deep learning system in a transparent manner to enforce and regulate its learning and decision process. In addition, the incorporation of human knowledge and reasoning rules into ML can significantly reduce the sample size needed to train the ML algorithm. Therefore, coupling explainable ML with wearable electronics for health monitoring and associated medical interventions is urgently needed.

Table 2. Performance of AI Algorithms Employed in Various Biosensors^a

data type	AI algorithms	samples	accuracy	LOD	ref
voltammograms	ANN (Bayesian regularization) and PLS	tryptophan (TRP), tyrosine (TYR), and cysteine (CYS) amino acid mixtures	ANN: R^2 of 0.960, 0.952, 0.967; PLS: R^2 of 0.867, 0.718, 0.926	NA	181
ECL images and amperograms	RF, feedforward ANN	$\text{Ru}(\text{bpy})_2^{3+}$	RF: R^2 of 0.99 feedforward ANN: R^2 of 0.96	0.02 μM	182
fluorescent spectra	LDA, DT, RF, NB, SVM, κ -NN, GBT	48 examples from 8 classes of protein, 24 unknown samples	100% pattern recognition accuracy for κ -NN, GBT, SVM, and LR	LDA: 25 nM in PBS, 50 nM in human plasma;	81
fluorometric images	bagging (bootstrap aggregating)	1370 cysts and 1485 other microparticles	cyst counting sensitivity of ~84%	~12 cysts per 10 mL	171
chemiluminescence (CL)	feedforward ANN	three organophosphorus pesticides	R^2 of 0.9995, 0.9989, and 0.9987	1×10^{-8} g/mL	59
color-spectra	CNN (i.e., LeNet, Vanilla CNN, ResNet, SqueezeNet, VGG and Inception v1)	5 kinds of amino acids, 3,600 color-spectral images (80% for training, 20% for test set) and their 5-dimensional concentration features	Inception v1 model: concentration coefficient (R^2) of actual vs predicted is of 0.999 and RMSE is of 10.22%	100 mg/L	172
colorimetric images	CNN	presence or absence of bacteria in water	~99.99%	NA	167
SERS	SVM	methamphetamine (MAMP)	90% in real addicts' urine	0.1 ppm	58
	PCA-SVM	histamine	96.8% in artificially spoiled tuna solution	10 ppm	183
	feedforward ANN	quantitative detection of pyocyanin	R^2 of 0.991	100 ppt (100 pg/mL ⁻¹)	184
	1D CNN		R^2 of 0.998		
	1D CNN	60,000 spectra from 30 bacterial and yeast isolates	$82.2 \pm 0.3\%$	~10 cells	50
	SVM		74.9%		
	LR		75.7%		
	CNN	Rhodamine 800	R^2 of 0.958	10 fM	51
	DNN	tumor suppressor genes, p16, p21, and p53 fragments	90.28% in water ~76% in human plasma	10 μM	185

^aAlanine (Ala), arginine (Arg), amino acids (AA), deep neural network (DNN), decision tree (DT), electrochemiluminescence (ECL), gradient-boosted trees (GBT), histidine (His), nearest neighbors (κ -NN), linear discriminant analysis (LDA), Logistic Regression (LR), phenylalanine (l-phe), Naive Bayes (NB), random forest (RF), root mean squared error (RMSE), support vector machines (SVM), threonine (Thr).

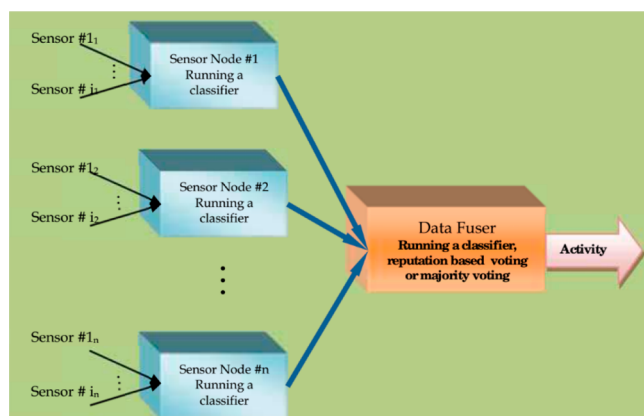


Figure 9. Diagram of multi-biosensor data fusion. Reproduced with permission from 190. Copyright 2012 In Tech.

AUTHOR INFORMATION

Corresponding Author

H. Susan Zhou – Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States; orcid.org/0000-0002-6659-6965; Email: szhou@wpi.edu

Authors

Feiyun Cui – Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States; orcid.org/0000-0001-5055-1423

Yun Yue – Department of Electrical & Computer Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States

Yi Zhang – Department of Biomedical Engineering, University of Connecticut, Storrs, Connecticut 06269, United States

Ziming Zhang – Department of Electrical & Computer Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acssensors.0c01424>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was funded by the National Science Foundation (CBET-1805514).

VOCABULARY

Biosensors: biosensors are a type of detection or diagnostic devices consisting of a sensing surface with bioreceptors, a transducer, and associated electronics which contains the signal processor, a signal amplifier, and a user-friendly readout interface.; **Chemometrics:** belong to the chemical discipline which employs statistical or mathematical methods, (a) to interrogate maximum chemical information by analyzing chemical data, and (b) to select or design optimal experiments and measurement procedures.; **Machine learning:** is referring that give computers the ability to learn without being explicitly programmed (Arthur Samuel in 1959), or a computer program is said to learn from an experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E (Tom Mitchell

in 1997). Machine learning algorithms develop a mathematical model based on sample data, known as “training data”, in order to make predictions or decisions without being explicitly programmed to do so.; **Deep learning:** is advanced machine learning that applies a hierarchical recombination of features to extract pertinent information and then learn the patterns represented in the data.; **Supervised learning:** refers to the ML algorithms which can be trained with a group of input data with their target outputs. In the training phase, the algorithms make certain predictions on the input data set and use the given true value to improve the predictive value until the algorithms reach an acceptable accuracy.; **Unsupervised learning:** there is no labeled training data set with their given outputs for the unsupervised learning. The goal is to determine the distribution of data set in the input space (called density estimation), or to find a set of similar examples in the input data set (known as clustering).

REFERENCES

- (1) Schuster, B. S-layer protein-based biosensors. *Biosensors* **2018**, *8* (2), 40.
- (2) Cui, F.; Zhou, Z.; Zhou, H. S. Measurement and analysis of cancer biomarkers based on electrochemical biosensors. *J. Electrochem. Soc.* **2020**, *167* (3), No. 037525.
- (3) Mohanty, S. P.; Kougianos, E. Biosensors: a tutorial review. *IEEE Potentials* **2006**, *25* (2), 35–40.
- (4) Cui, F.; Xu, Y.; Wang, R.; Liu, H.; Chen, L.; Zhang, Q.; Mu, X. Label-free impedimetric glycan biosensor for quantitative evaluation interactions between pathogenic bacteria and mannose. *Biosens. Bioelectron.* **2018**, *103*, 94–98.
- (5) Malmqvist, M. Biospecific interaction analysis using biosensor technology. *Nature* **1993**, *361* (6408), 186–187.
- (6) Jin, Y.; Xie, Y.; Wu, K.; Huang, Y.; Wang, F.; Zhao, R. Probing the dynamic interaction between damaged DNA and a cellular responsive protein using a piezoelectric mass biosensor. *ACS Appl. Mater. Interfaces* **2017**, *9* (10), 8490–8497.
- (7) Mehta, S.; Zhang, Y.; Roth, R. H.; Zhang, J.-f.; Mo, A.; Tenner, B.; Haganir, R. L.; Zhang, J. Single-fluorophore biosensors for sensitive and multiplexed detection of signalling activities. *Nat. Cell Biol.* **2018**, *20* (10), 1215–1225.
- (8) Lv, M.; Liu, Y.; Geng, J.; Kou, X.; Xin, Z.; Yang, D. Engineering nanomaterials-based biosensors for food safety detection. *Biosens. Bioelectron.* **2018**, *106*, 122–128.
- (9) Zhang, X.; Wu, D.; Zhou, X.; Yu, Y.; Liu, J.; Hu, N.; Wang, H.; Li, G.; Wu, Y. Recent progress on the construction of nanozymes-based biosensors and their applications to food safety assay. *TrAC, Trends Anal. Chem.* **2019**, *121*, 115668.
- (10) Weber, P.; Vogler, J.; Gauglitz, G. In *Development of an optical biosensor for the detection of antibiotics in the environment*, *Optical Sensors 2017*; International Society for Optics and Photonics: 2017; p 102312L.
- (11) Rogers, K. Recent advances in biosensor techniques for environmental monitoring. *Anal. Chim. Acta* **2006**, *568* (1–2), 222–231.
- (12) Long, F.; Zhu, A.; Shi, H. Recent advances in optical biosensors for environmental monitoring and early warning. *Sensors* **2013**, *13* (10), 13928–13948.
- (13) Qian, L.; Li, Q.; Kwaku, B.; Qiu, W.; Li, K.; Zhang, J.; Yu, Q.; Xu, D.; Liu, W.; Brand, R. E. Biosensors for early diagnosis of pancreatic cancer: a review. *Translational Research* **2019**, *213*, 67.
- (14) Sadighbayan, D.; Sadighbayan, K.; Tohid-Kia, M. R.; Khosroushahi, A. Y.; Hasanzadeh, M. Development of electrochemical biosensors for tumor marker determination towards cancer diagnosis: Recent progress. *TrAC, Trends Anal. Chem.* **2019**, *118*, 73.
- (15) Hamm, L.; Gee, A.; De Silva Indrasekara, A. S. Recent Advancement in the Surface-Enhanced Raman Spectroscopy-Based

Biosensors for Infectious Disease Diagnosis. *Appl. Sci.* **2019**, *9* (7), 1448.

(16) Cui, F.; Zhou, Z.; Feng, H.; Zhou, H. S. Disposable Polyurethane Nanospiked Gold Electrode-Based Label-Free Electrochemical Immunosensor for *Clostridium difficile*. *ACS Applied Nano Materials* **2020**, *3* (1), 357–363.

(17) Cui, F.; Zhou, Z.; Zhou, H. S. Molecularly imprinted polymers and surface imprinted polymers based electrochemical biosensor for infectious diseases. *Sensors* **2020**, *20* (4), 996.

(18) Kloß, D.; Kurz, R.; Jahnke, H.-G.; Fischer, M.; Rothermel, A.; Anderegg, U.; Simon, J. C.; Robitzki, A. A. Microcavity array (MCA)-based biosensor chip for functional drug screening of 3D tissue models. *Biosens. Bioelectron.* **2008**, *23* (10), 1473–1480.

(19) Pan, Y.; Hu, N.; Wei, X.; Gong, L.; Zhang, B.; Wan, H.; Wang, P. 3D cell-based biosensor for cell viability and drug assessment by 3D electric cell/matrigel-substrate impedance sensing. *Biosens. Bioelectron.* **2019**, *130*, 344–351.

(20) Cooper, M. A. Optical biosensors in drug discovery. *Nat. Rev. Drug Discovery* **2002**, *1* (7), 515–528.

(21) Zourob, M. *Recognition receptors in biosensors*; Springer: 2010.

(22) Justino, C. I.; Freitas, A. C.; Pereira, R.; Duarte, A. C.; Santos, T. A. R. Recent developments in recognition elements for chemical sensors and biosensors. *TrAC, Trends Anal. Chem.* **2015**, *68*, 2–17.

(23) Otto, M. *Chemometrics: statistics and computer application in analytical chemistry*; John Wiley & Sons: 2016.

(24) Reinholds, I.; Bartkevics, V.; Silvis, I. C.; van Ruth, S. M.; Esslinger, S. Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review. *J. Food Compos. Anal.* **2015**, *44*, 56–72.

(25) Panchuk, V.; Yaroshenko, I.; Legin, A.; Semenov, V.; Kirsanov, D. Application of chemometric methods to XRF-data—A tutorial review. *Anal. Chim. Acta* **2018**, *1040*, 19–32.

(26) Pravdová, V.; Pravda, M.; Guilbault, G. G. ROLE OF CHEMOMETRICS FOR ELECTROCHEMICAL SENSORS. *Anal. Lett.* **2002**, *35* (15), 2389–2419.

(27) Lindholm-Sethson, B.; Nyström, J.; Malmsten, M.; Ringstad, L.; Nelson, A.; Geladi, P. Electrochemical impedance spectroscopy in label-free biosensor applications: multivariate data analysis for an objective interpretation. *Anal. Bioanal. Chem.* **2010**, *398* (6), 2341–2349.

(28) Torrecilla, J. S.; Mena, M. L.; Yáñez-Sedeño, P.; García, J. A neural network approach based on gold-nanoparticle enzyme biosensor. *J. Chemom.* **2008**, *22* (1), 46–53.

(29) Baronas, R.; Ivanauskas, F.; Maslovskis, R.; Vaitkus, P. An Analysis of Mixtures Using Amperometric Biosensors and Artificial Neural Networks. *J. Math. Chem.* **2004**, *36* (3), 281–297.

(30) Wang, Y.; Ni, Y. Combination of UV–vis spectroscopy and chemometrics to understand protein–nanomaterial conjugate: A case study on human serum albumin and gold nanoparticles. *Talanta* **2014**, *119*, 320–330.

(31) Abdullah, J.; Ahmad, M.; Heng, L. Y.; Karuppiah, N.; Sidek, H. Evaluation of an optical phenolic biosensor signal employing artificial neural networks. *Sens. Actuators, B* **2008**, *134* (2), 959–965.

(32) Villa, J. E.; Afonso, M. A.; Dos Santos, D. P.; Mercadal, P. A.; Coronado, E. A.; Poppi, R. J. Colloidal gold clusters formation and chemometrics for direct SERS determination of bioanalytes in complex media. *Spectrochim. Acta, Part A* **2020**, *224*, 117380.

(33) Kangas, M. J.; Burks, R. M.; Atwater, J.; Lukowicz, R. M.; Garver, B.; Holmes, A. E. Comparative chemometric analysis for classification of acids and bases via a colorimetric sensor array. *J. Chemom.* **2018**, *32* (2), e2961.

(34) Ghasemi, F.; Hormozi-Nezhad, M. R.; Mahmoudi, M. A colorimetric sensor array for detection and discrimination of biothiols based on aggregation of gold nanoparticles. *Anal. Chim. Acta* **2015**, *882*, 58–67.

(35) Vasilescu, A.; Purcarea, C.; Popa, E.; Zamfir, M.; Mihai, I.; Litescu, S.; David, S.; Gaspar, S.; Gheorghiu, M.; Jean-Louis, M. Versatile SPR aptasensor for detection of lysozyme dimer in

oligomeric and aggregated mixtures. *Biosens. Bioelectron.* **2016**, *83*, 353–360.

(36) Yuan, Y.-K.; Xiao, X.-L.; Wang, Y.-S.; Xue, J.-H.; Li, G.-R.; Kang, R.-H.; Zhang, J.-Q.; Shi, L.-F. Quartz crystal microbalance with β -cyclodextrin/TiO₂ composite films coupled with chemometrics for the simultaneous determination of urinary 1-and 2-naphthol. *Sens. Actuators, B* **2010**, *145* (1), 348–354.

(37) Jurs, P. C.; Bakken, G.; McClelland, H. Computational methods for the analysis of chemical sensor array data from volatile analytes. *Chem. Rev.* **2000**, *100* (7), 2649–2678.

(38) Stewart, S.; Ivy, M. A.; Anslyn, E. V. The use of principal component analysis and discriminant analysis in differential sensing routines. *Chem. Soc. Rev.* **2014**, *43* (1), 70–84.

(39) Goodacre, R.; Graham, D.; Faulds, K. Recent developments in quantitative SERS: Moving towards absolute quantification. *TrAC, Trends Anal. Chem.* **2018**, *102*, 359–368.

(40) Asefpour Vakilian, K.; Massah, J. A portable nitrate biosensing device using electrochemistry and spectroscopy. *IEEE Sens. J.* **2018**, *18* (8), 3080–3089.

(41) Morellos, A.; Pantazi, X.-E.; Moshou, D.; Alexandridis, T.; Whetton, R.; Tziotziou, G.; Wiebensohn, J.; Bill, R.; Mouazen, A. M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosys. Eng.* **2016**, *152*, 104–116.

(42) Hollon, T. C.; Pandian, B.; Adapa, A. R.; Urias, E.; Save, A. V.; Khalsa, S. S. S.; Eichberg, D. G.; D'Amico, R. S.; Farooq, Z. U.; Lewis, S. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat. Med.* **2020**, *26*, 52.

(43) Moon, G.; Choi, J.-r.; Lee, C.; Oh, Y.; Kim, K.; Kim, D. Machine learning-based design of meta-plasmonic biosensors with negative index metamaterials. *Biosens. Bioelectron.* **2020**, *164*, 112335.

(44) Lussier, F.; Thibault, V.; Charron, B.; Wallace, G. Q.; Masson, J.-F. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrAC, Trends Anal. Chem.* **2020**, *124*, 115796.

(45) Dasgupta, A.; Nath, A. Classification of Machine Learning Algorithms. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)* **2016**, *3* (3), 6–11.

(46) Ayodele, T. O. Types of machine learning algorithms. *New advances in machine learning* **2010**, 19–48.

(47) Yang, J.; Xu, J.; Zhang, X.; Wu, C.; Lin, T.; Ying, Y. Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Anal. Chim. Acta* **2019**, *1081*, 6–17.

(48) Shanthamallu, U. S.; Spanias, A.; Tepedelenioglu, C.; Stanley, M. In *A brief survey of machine learning methods and their sensor and IoT applications*; 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA), IEEE: 2017; pp 1–8.

(49) Bishop, C. M. *Pattern recognition and machine learning*; Springer: 2006.

(50) Ho, C.-S.; Jean, N.; Hogan, C. A.; Blackmon, L.; Jeffrey, S. S.; Holodniy, M.; Banaei, N.; Saleh, A. A.; Ermon, S.; Dionne, J. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* **2019**, *10* (1), 1–8.

(51) Thrift, W. J.; Ragan, R. Quantification of Analyte Concentration in the Single Molecule Regime Using Convolutional Neural Networks. *Anal. Chem.* **2019**, *91* (21), 13337–13342.

(52) Acharya, U. R.; Oh, S. L.; Hagiwara, Y.; Tan, J. H.; Adeli, H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput. Biol. Med.* **2018**, *100*, 270–278.

(53) Mitchell, R.; Michalski, J.; Carbonell, T. *An artificial intelligence approach*; Springer: 2013.

(54) Thirukovalluru, R.; Dixit, S.; Sevakula, R. K.; Verma, N. K.; Salour, A. In *Generating feature sets for fault diagnosis using denoising stacked auto-encoder*; 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), IEEE: 2016; pp 1–7.

- (55) Namuduri, S.; Narayanan, B. N.; Davuluru, V. S. P.; Burton, L.; Bhansali, S. Deep Learning Methods for Sensor Based Predictive Maintenance and Future Perspectives for Electrochemical Sensors. *J. Electrochem. Soc.* **2020**, 167 (3), No. 037552.
- (56) He, K.; Zhang, X.; Ren, S.; Sun, J. In *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*; Proceedings of the IEEE international conference on computer vision, 2015; pp 1026–1034.
- (57) Glorot, X.; Bengio, Y. In *Understanding the difficulty of training deep feedforward neural networks*; Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010; pp 249–256.
- (58) 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.* **2015**, 87 (5), 2937–2944.
- (59) Li, B.; He, Y.; Xu, C. Simultaneous determination of three organophosphorus pesticides residues in vegetables using continuous-flow chemiluminescence with artificial neural network calibration. *Talanta* **2007**, 72 (1), 223–230.
- (60) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep learning*; MIT press: 2016.
- (61) Tania, M. H.; Lwin, K. T.; Shabut, A. M.; Abu-Hassan, K. J.; Kaiser, M. S.; Hossain, M. In *Assay Type Detection Using Advanced Machine Learning Algorithms*; 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), IEEE: 2019; pp 1–8.
- (62) Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y.-g.; Choi, B. H.; Kang, K.-W.; Jeong, H.; Park, Y. Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes. *ACS Nano* **2020**, 14 (5), 5435–5444.
- (63) Majumder, S.; Ghosh, N.; Gupta, P. Support vector machine for optical diagnosis of cancer. *J. Biomed. Opt.* **2005**, 10 (2), No. 024034.
- (64) 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.* **2017**, 11 (1), 9.
- (65) Awad, M.; Khanna, R., Support Vector Regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Awad, M.; Khanna, R., Eds.; Apress: Berkeley, CA, 2015; pp 67–80.
- (66) Peterson, L. E. K-nearest neighbor. *Scholarpedia* **2009**, 4 (2), 1883.
- (67) Bermejo, S.; Cabestany, J. Adaptive soft k-nearest-neighbour classifiers. *Pattern Recognition* **2000**, 33 (12), 1999–2005.
- (68) Cherif, W. Optimization of K-NN algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis. *Procedia Computer Science* **2018**, 127, 293–299.
- (69) Islam, M. M.; Iqbal, H.; Haque, M. R.; Hasan, M. K. In *Prediction of breast cancer using support vector machine and K-Nearest neighbors*; 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), IEEE: 2017; pp 226–229.
- (70) Su, M.-Y. Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers. *Expert Systems with Applications* **2011**, 38 (4), 3492–3498.
- (71) 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 Systems with Applications* **2014**, 41 (4), 1937–1946.
- (72) Singh, A.; Thakur, N.; Sharma, A. In *A review of supervised machine learning algorithms*; 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE: 2016; pp 1310–1315.
- (73) Si, S.; Zhang, H.; Keerthi, S. S.; Mahajan, D.; Dhillon, I. S.; Hsieh, C.-J. In *Gradient Boosted Decision Trees for High Dimensional Sparse Output*; International Conference on Machine Learning, 2017; pp 3182–3190.
- (74) Shaikhina, T.; Lowe, D.; Daga, S.; Briggs, D.; Higgins, R.; Khovanova, N. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. *Bio-medical Signal Processing and Control* **2019**, 52, 456–462.
- (75) Rodriguez, J. J.; Kuncheva, L. I.; Alonso, C. J. Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence* **2006**, 28 (10), 1619–1630.
- (76) Blanchet, L.; Vitale, R.; van Vorstenbosch, R.; Stavropoulos, G.; Pender, J.; Jonkers, D.; Schooten, F.-J. v.; Smolinska, A. Constructing bi-approach for random forest: Tutorial. *Anal. Chim. Acta* **2020**, 1131, 146.
- (77) Krauss, C.; Do, X. A.; Huck, N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* **2017**, 259 (2), 689–702.
- (78) Dev, V. A.; Eden, M. R. Gradient Boosted Decision Trees for Lithology Classification. *Comput.-Aided Chem. Eng.* **2019**, 47, 113–118.
- (79) Guelman, L. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications* **2012**, 39 (3), 3659–3667.
- (80) Whiting, D. G.; Hansen, J. V.; McDonald, J. B.; Albrecht, C.; Albrecht, W. S. Machine learning methods for detecting patterns of management fraud. *Computational Intelligence* **2012**, 28 (4), 505–527.
- (81) Pandit, S.; Banerjee, T.; Srivastava, I.; Nie, S.; Pan, D. Machine Learning-Assisted Array-Based Biomolecular Sensing Using Surface-Functionalized Carbon Dots. *ACS sensors* **2019**, 4 (10), 2730–2737.
- (82) 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 Protect.* **2013**, 46 (13), 1580–1588.
- (83) Kallenberg, M.; Petersen, K.; Nielsen, M.; Ng, A. Y.; Diao, P.; Igel, C.; Vachon, C. M.; Holland, K.; Winkel, R. R.; Karssemeijer, N. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans. Med. Imaging* **2016**, 35 (5), 1322–1331.
- (84) Pereira, S.; Pinto, A.; Alves, V.; Silva, C. A. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **2016**, 35 (5), 1240–1251.
- (85) Setio, A. A. A.; Ciompi, F.; Litjens, G.; Gerke, P.; Jacobs, C.; Van Riel, S. J.; Wille, M. M. W.; Naqibullah, M.; Sánchez, C. I.; van Ginneken, B. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* **2016**, 35 (5), 1160–1169.
- (86) Maas, A. L.; Hannun, A. Y.; Ng, A. Y. In *Rectifier nonlinearities improve neural network acoustic models*, 2013; p 3.
- (87) Bridle, J. S. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition, In *Neurocomputing*, Soulié, F. F.; Héroult, J., Eds.; Berlin, Heidelberg, 1990; pp 227–236.
- (88) Malek, S.; Melgani, F.; Bazi, Y. One-dimensional convolutional neural networks for spectroscopic signal regression. *J. Chemom.* **2018**, 32 (5), e2977.
- (89) Tsantekidis, A.; Passalis, N.; Tefas, A.; Kannianen, J.; Gabbouj, M.; Iosifidis, A. In *Forecasting stock prices from the limit order book using convolutional neural networks*; 2017 IEEE 19th Conference on Business Informatics (CBI), IEEE: 2017; pp 7–12.
- (90) Tabar, Y. R.; Halici, U. A novel deep learning approach for classification of EEG motor imagery signals. *Journal of neural engineering* **2017**, 14 (1), No. 016003.
- (91) Lee, J.; Park, J.; Kim, K. L.; Nam, J., Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms. *arXiv*, 1703.01789, **2017**.
- (92) Lussier, F.; Missirlis, D.; Spatz, J. P.; Masson, J.-F. Machine-learning-driven surface-enhanced Raman scattering optophysiology reveals multiplexed metabolite gradients near cells. *ACS Nano* **2019**, 13 (2), 1403–1411.
- (93) Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. In *Going deeper with convolutions*; Proceedings of the IEEE conference on computer vision and pattern recognition, 2015; pp 1–9.

- (94) Zhang, X.; Lin, T.; Xu, J.; Luo, X.; Ying, Y. DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. *Anal. Chim. Acta* **2019**, 1058, 48–57.
- (95) Lipton, Z. C.; Berkowitz, J.; Elkan, C., A critical review of recurrent neural networks for sequence learning. *arXiv*, 1506.00019, 2015.
- (96) Ordóñez, F. J.; Roggen, D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, 16 (1), 115.
- (97) Hanson, J.; Yang, Y.; Paliwal, K.; Zhou, Y. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics* **2016**, 33 (5), 685–692.
- (98) Kiddon, C.; Zettlemoyer, L.; Choi, Y. In *Globally coherent text generation with neural checklist models*; Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016; pp 329–339.
- (99) Takeuchi, D.; Yatabe, K.; Koizumi, Y.; Oikawa, Y.; Harada, N. In *Real-time speech enhancement using equilibrated RNN*; ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE: 2020; pp 851–855.
- (100) Miao, Y.; Gawayyed, M.; Metze, F. In *EESSEN: End-to-end speech recognition using deep RNN models and WFST-based decoding*; 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), IEEE: 2015; pp 167–174.
- (101) Doetsch, P.; Kozielski, M.; Ney, H. In *Fast and robust training of recurrent neural networks for offline handwriting recognition*; 2014 14th International Conference on Frontiers in Handwriting Recognition, IEEE: 2014; pp 279–284.
- (102) Williams, J. D.; Zweig, G. End-to-end lstm-based dialog control optimized with supervised and reinforcement learning. *arXiv*, 1606.01269, 2016.
- (103) Zoph, B.; Le, Q. V. Neural architecture search with reinforcement learning. *arXiv*, 1611.01578, 2016.
- (104) Corbett, P.; Boyle, J. Improving the learning of chemical-protein interactions from literature using transfer learning and specialized word embeddings. *Database* **2018**, 2018, 1 DOI: 10.1093/database/bay066.
- (105) Lim, S.; Kang, J. Chemical–gene relation extraction using recursive neural network. *Database* **2018**, 2018, 1 DOI: 10.1093/database/bay060.
- (106) Peng, Y.; Rios, A.; Kavuluru, R.; Lu, Z., Chemical-protein relation extraction with ensembles of SVM, CNN, and RNN models. *arXiv*, 1802.01255, 2018.
- (107) Zhao, X.; Wu, Y.; Song, G.; Li, Z.; Fan, Y.; Zhang, Y. In *Brain tumor segmentation using a fully convolutional neural network with conditional random fields*, International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer: 2016; pp 75–87.
- (108) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, 9 (8), 1735–1780.
- (109) Liu, Q.; Fang, L.; Yu, G.; Wang, D.; Xiao, C.-L.; Wang, K. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun.* **2019**, 10 (1), 1–11.
- (110) Boža, V.; Brejová, B.; Vinař, T. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. *PLoS One* **2017**, 12 (6), No. e0178751.
- (111) Rang, F. J.; Kloosterman, W. P.; de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **2018**, 19 (1), 90.
- (112) Massah, J.; Asefpour Vakilian, K. An intelligent portable biosensor for fast and accurate nitrate determination using cyclic voltammetry. *Biosys. Eng.* **2019**, 177, 49–58.
- (113) Gonzalez-Navarro, F. F.; Stilianova-Stoytcheva, M.; Renteria-Gutierrez, L.; Belanche-Muñoz, L. A.; Flores-Rios, B. L.; Ibarra-Esquer, J. E. Glucose oxidase biosensor modeling and predictors optimization by machine learning methods. *Sensors* **2016**, 16 (11), 1483.
- (114) Rong, Y.; Padron, A.; Hagerty, K.; Nelson, N.; Chi, S.; Keyhani, N. O.; Katz, J.; Datta, S.; Gomes, C.; McLamore, E. S. Post hoc support vector machine learning for impedimetric biosensors based on weak protein–ligand interactions. *Analyst* **2018**, 143 (9), 2066–2075.
- (115) Zeng, Z.; Huang, Z.; Leng, K.; Han, W.; Niu, H.; Yu, Y.; Ling, Q.; Liu, J.; Wu, Z.; Zang, J. Non-intrusive Monitoring of Mental Fatigue Status Using Epidermal Electronic Systems and Machine-learning Algorithms. *ACS Sensors* **2020**, 5, 1305.
- (116) Yan, W.; Wang, K.; Xu, H.; Huo, X.; Jin, Q.; Cui, D. Machine learning approach to enhance the performance of MNP-labeled lateral flow immunoassay. *Nano-Micro Lett.* **2019**, 11 (1), 7.
- (117) Albrecht, T.; Slabaugh, G.; Alonso, E.; Al-Arif, S. M. R. Deep learning for single-molecule science. *Nanotechnology* **2017**, 28 (42), 423001.
- (118) Ali, S.; Hassan, A.; Hassan, G.; Eun, C.-H.; Bae, J.; Lee, C. H.; Kim, I.-J. Disposable all-printed electronic biosensor for instantaneous detection and classification of pathogens. *Sci. Rep.* **2018**, 8 (1), 1–11.
- (119) González-Calabuig, A.; Faura, G.; del Valle, M. In *High dimensionality voltammetric biosensor data processed with artificial neural networks*; ESANN, 2017.
- (120) Guselnikova, O.; Trelin, A.; Skvortsova, A.; Ulbrich, P.; Postnikov, P.; Pershina, A.; Sykora, D.; Svorcik, V.; Lyutakov, O. Label-free surface-enhanced Raman spectroscopy with artificial neural network technique for recognition photoinduced DNA damage. *Biosens. Bioelectron.* **2019**, 145, 111718.
- (121) Erzina, M.; Trelin, A.; Guselnikova, O.; Dvorankova, B.; Strnadova, K.; Perminova, A.; Ulbrich, P.; Mares, D.; Jerabek, V.; Elashnikov, R.; Svorcik, V.; Lyutakov, O. Precise cancer detection via the combination of functionalized SERS surfaces and convolutional neural network with independent inputs. *Sens. Actuators, B* **2020**, 308, 127660.
- (122) Ni, Y.; Kokot, S. Does chemometrics enhance the performance of electroanalysis? *Anal. Chim. Acta* **2008**, 626 (2), 130–146.
- (123) Daniels, J. S.; Pourmand, N. Label-free impedance biosensors: Opportunities and challenges. *Electroanalysis* **2007**, 19 (12), 1239–1257.
- (124) Bahadır, E. B.; Sezgentürk, M. K. A review on impedimetric biosensors. *Artif. Cells, Nanomed., Biotechnol.* **2016**, 44 (1), 248–262.
- (125) Orazem, M. E.; Agarwal, P.; Garcia-Rubio, L. H. Critical issues associated with interpretation of impedance spectra. *J. Electroanal. Chem.* **1994**, 378 (1), 51–62.
- (126) Taniguchi, M. Combination of Single-Molecule Electrical Measurements and Machine Learning for the Identification of Single Biomolecules. *ACS Omega* **2020**, 5 (2), 959–964.
- (127) Restrepo-Pérez, L.; Joo, C.; Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **2018**, 13 (9), 786–796.
- (128) Smeets, R. M.; Keyser, U. F.; Dekker, N. H.; Dekker, C. Noise in solid-state nanopores. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105 (2), 417–421.
- (129) Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* **2007**, 2 (4), 209.
- (130) Di Ventra, M.; Taniguchi, M. Decoding DNA, RNA and peptides with quantum tunnelling. *Nat. Nanotechnol.* **2016**, 11 (2), 117.
- (131) Heerema, S. J.; Dekker, C. Graphene nanodevices for DNA sequencing. *Nat. Nanotechnol.* **2016**, 11 (2), 127–136.
- (132) Im, J.; Biswas, S.; Liu, H.; Zhao, Y.; Sen, S.; Biswas, S.; Ashcroft, B.; Borges, C.; Wang, X.; Lindsay, S. Electronic single-molecule identification of carbohydrate isomers by recognition tunnelling. *Nat. Commun.* **2016**, 7 (1), 1–7.
- (133) Arima, A.; Harlisa, I. H.; Yoshida, T.; Tsutsui, M.; Tanaka, M.; Yokota, K.; Tonomura, W.; Yasuda, J.; Taniguchi, M.; Washio, T. Identifying single viruses using biorecognition solid-state nanopores. *J. Am. Chem. Soc.* **2018**, 140 (48), 16834–16841.
- (134) Arima, A.; Tsutsui, M.; Harlisa, I. H.; Yoshida, T.; Tanaka, M.; Yokota, K.; Tonomura, W.; Taniguchi, M.; Okochi, M.; Washio, T.

Selective detections of single-viruses using solid-state nanopores. *Sci. Rep.* **2018**, *8* (1), 1–7.

(135) Zhao, Y.; Ashcroft, B.; Zhang, P.; Liu, H.; Sen, S.; Song, W.; Im, J.; Gyrfas, B.; Manna, S.; Biswas, S. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotechnol.* **2014**, *9* (6), 466.

(136) Tsutsui, M.; Yoshida, T.; Yokota, K.; Yasaki, H.; Yasui, T.; Arima, A.; Tonomura, W.; Nagashima, K.; Yanagida, T.; Kaji, N. Discriminating single-bacterial shape using low-aspect-ratio pores. *Sci. Rep.* **2017**, *7* (1), 1–9.

(137) Tsutsui, M.; Tanaka, M.; Marui, T.; Yokota, K.; Yoshida, T.; Arima, A.; Tonomura, W.; Taniguchi, M.; Washio, T.; Okochi, M. Identification of individual bacterial cells through the intermolecular interactions with peptide-functionalized solid-state pores. *Anal. Chem.* **2018**, *90* (3), 1511–1515.

(138) Sim, K.; Rao, Z.; Zou, Z.; Ershad, F.; Lei, J.; Thukral, A.; Chen, J.; Huang, Q.-A.; Xiao, J.; Yu, C. J. S. a. Metal oxide semiconductor nanomembrane-based soft unnoticeable multifunctional electronics for wearable human-machine interfaces. *Sci. Adv.* **2019**, *5* (8), eaav9653.

(139) Liu, Y. H.; Norton, J. J. S.; Qazi, R.; Zou, Z. N.; Ammann, K. R.; Liu, H.; Yan, L. Q.; Tran, P. L.; Jang, K. I.; Lee, J. W.; Zhang, D.; Kilian, K. A.; Jung, S. H.; Bretl, T.; Xiao, J. L.; Slepian, M. J.; Huang, Y. G.; Jeong, J. W.; Rogers, J. A. Epidermal mechano-acoustic sensing electronics for cardiovascular diagnostics and human-machine interfaces. *Sci. Adv.* **2016**, *2* (11), 1 DOI: 10.1126/sciadv.1601185.

(140) Wang, S. H.; Xu, J.; Wang, W. C.; Wang, G. J. N.; Rastak, R.; Molina-Lopez, F.; Chung, J. W.; Niu, S. M.; Feig, V. R.; Lopez, J.; Lei, T.; Kwon, S. K.; Kim, Y.; Foudeh, A. M.; Ehrlich, A.; Gasperini, A.; Yun, Y.; Murmann, B.; Tok, J. B. H.; Bao, Z. A. Skin electronics from scalable fabrication of an intrinsically stretchable transistor array. *Nature* **2018**, *555* (7694), 83.

(141) Kim, J.; Lee, M.; Shim, H. J.; Ghaffari, R.; Cho, H. R.; Son, D.; Jung, Y. H.; Soh, M.; Choi, C.; Jung, S.; Chu, K.; Jeon, D.; Lee, S. T.; Kim, J. H.; Choi, S. H.; Hyeon, T.; Kim, D. H. Stretchable silicon nanoribbon electronics for skin prosthesis. *Nat. Commun.* **2014**, *5*, 1 DOI: 10.1038/ncomms6747.

(142) Zhang, Y.; Guo, H. X.; Kim, S. B.; Wu, Y. X.; Ostojich, D.; Park, S. H.; Wang, X. J.; Wang, Z. Y.; Li, R.; Bhandodkar, A. J.; Sekine, Y.; Choi, J.; Xu, S.; Quaggin, S.; Ghaffari, R.; Rogers, J. A. Passive sweat collection and colorimetric analysis of biomarkers relevant to kidney disorders using a soft microfluidic system. *Lab Chip* **2019**, *19* (9), 1545–1555.

(143) Yang, Y. R.; Song, Y.; Bo, X. J.; Min, J. H.; Pak, O. S.; Zhu, L. L.; Wang, M. Q.; Tu, J. B.; Kogan, A.; Zhang, H. X.; Hsiai, T. K.; Li, Z. P.; Gao, W. A laser-engraved wearable sensor for sensitive detection of uric acid and tyrosine in sweat. *Nat. Biotechnol.* **2020**, *38* (2), 217.

(144) Koh, A.; Kang, D.; Xue, Y.; Lee, S.; Pielak, R. M.; Kim, J.; Hwang, T.; Min, S.; Banks, A.; Bastien, P. J. S. t. m. A soft, wearable microfluidic device for the capture, storage, and colorimetric sensing of sweat. *Sci. Transl. Med.* **2016**, *8* (366), 366ra165.

(145) Pegan, J. D.; Zhang, J.; Chu, M.; Nguyen, T.; Park, S. J.; Paul, A.; Kim, J.; Bachman, M.; Khine, M. Skin-mountable stretch sensor for wearable health monitoring. *Nanoscale* **2016**, *8* (39), 17295–17303.

(146) Windmiller, J. R.; Bhandodkar, A. J.; Valdés-Ramírez, G.; Parkhomovsky, S.; Martinez, A. G.; Wang, J. Electrochemical sensing based on printable temporary transfer tattoos. *Chem. Commun.* **2012**, *48* (54), 6794–6796.

(147) Kim, D.-H.; Lu, N.; Ma, R.; Kim, Y.-S.; Kim, R.-H.; Wang, S.; Wu, J.; Won, S. M.; Tao, H.; Islam, A. Epidermal electronics. *Science* **2011**, *333* (6044), 838–843.

(148) Yu, Y.; Nyein, H. Y. Y.; Gao, W.; Javey, A. Flexible Electrochemical Bioelectronics: The Rise of In Situ Bioanalysis. *Adv. Mater.* **2020**, *32* (15), 1902083.

(149) Sundaram, S.; Kellnhofer, P.; Li, Y.; Zhu, J.-Y.; Torralba, A.; Matusik, W. Learning the signatures of the human grasp using a scalable tactile glove. *Nature* **2019**, *569* (7758), 698–702.

(150) Jeong, H.; Rogers, J. A.; Xu, S. Continuous on-body sensing for the COVID-19 pandemic: Gaps and opportunities. *Science Advances* **2020**, *6* (36), eabd4794.

(151) Larkin, P. J. Chapter 1 - Introduction: Infrared and Raman Spectroscopy. In *Infrared and Raman Spectroscopy*, 2nd ed., Larkin, P. J., Ed.; Elsevier: 2018; pp 1–5.

(152) Kasera, S.; Herrmann, L. O.; Del Barrio, J.; Baumberg, J. J.; Scherman, O. A. Quantitative multiplexing with nano-self-assemblies in SERS. *Sci. Rep.* **2015**, *4*, 6785.

(153) Dong, J.; Hong, M.; Xu, Y.; Zheng, X. A practical convolutional neural network model for discriminating Raman spectra of human and animal blood. *J. Chemom.* **2019**, *33* (11), e3184.

(154) Liu, J.; Osadchy, M.; Ashton, L.; Foster, M.; Solomon, C. J.; Gibson, S. J. Deep convolutional neural networks for Raman spectrum recognition: a unified solution. *Analyst* **2017**, *142* (21), 4067–4074.

(155) Acquarelli, J.; van Laarhoven, T.; Gerretzen, J.; Tran, T. N.; Buydens, L. M. C.; Marchiori, E. Convolutional neural networks for vibrational spectroscopic data analysis. *Anal. Chim. Acta* **2017**, *954*, 22–31.

(156) Lu, W.; Chen, X.; Wang, L.; Li, H.; Fu, Y. V. Combination of an Artificial Intelligence Approach and Laser Tweezers Raman Spectroscopy for Microbial Identification. *Anal. Chem.* **2020**, *92* (9), 6288–6296.

(157) Cao, L.; Cui, X.; Hu, J.; Li, Z.; Choi, J. R.; Yang, Q.; Lin, M.; Hui, L. Y.; Xu, F. Advances in digital polymerase chain reaction (dPCR) and its emerging biomedical applications. *Biosens. Bioelectron.* **2017**, *90*, 459–474.

(158) Hu, Z.; Fang, W.; Gou, T.; Wu, W.; Hu, J.; Zhou, S.; Mu, Y. A novel method based on a Mask R-CNN model for processing dPCR images. *Anal. Methods* **2019**, *11* (27), 3410–3418.

(159) Fan, Q.; Chen, G.; Zhou, X.; Li, L. In *Image Threshold Segmentation Based on Auxiliary Individual Oriented Crossover Genetic Algorithm*; 2019 IEEE International Conference on Industrial Internet (ICII), IEEE: 2019; pp 411–416.

(160) He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. In *Mask R-CNN*; 2017 IEEE International Conference on Computer Vision (ICCV), 22–29 Oct. 2017; 2017; pp 2980–2988.

(161) Rissin, D. M.; Kan, C. W.; Campbell, T. G.; Howes, S. C.; Fournier, D. R.; Song, L.; Piech, T.; Patel, P. P.; Chang, L.; Rivnak, A. J. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* **2010**, *28* (6), 595.

(162) Liu, Y.; Li, F.; Qiu, L.; Yang, K.; Li, Q.; Zheng, X.; Hu, H.; Guo, T.; Wu, C.; Kim, T. W. Fluorescent Microarrays of in Situ Crystallized Perovskite Nanocomposites Fabricated for Patterned Applications by Using Inkjet Printing. *ACS Nano* **2019**, *13* (2), 2042–2049.

(163) Xiao, M.; Liu, Z.; Xu, N.; Jiang, L.; Yang, M.; Yi, C. A smartphone-based sensing system for on-site quantitation of multiple heavy metal ions using fluorescent carbon nanodots-based microarrays. *ACS sensors* **2020**, *5* (3), 870–878.

(164) Wang, Z.; Huang, X.; Cheng, Z. Automatic Spot Identification Method for High Throughput Surface Plasmon Resonance Imaging Analysis. *Biosensors* **2018**, *8* (3), 85.

(165) Solmaz, M. E.; Mutlu, A. Y.; Alankus, G.; Kiliç, V.; Bayram, A.; Horzum, N. Quantifying colorimetric tests using a smartphone app based on machine learning classifiers. *Sens. Actuators, B* **2018**, *255*, 1967–1973.

(166) Yetisen, A. K.; Martinez-Hurtado, J.; Garcia-Melendrez, A.; da Cruz Vasconcellos, F.; Lowe, C. R. A smartphone algorithm with inter-phone repeatability for the analysis of colorimetric tests. *Sens. Actuators, B* **2014**, *196*, 156–160.

(167) Gunda, N. S. K.; Gautam, S. H.; Mitra, S. K. Artificial Intelligence Based Mobile Application for Water Quality Monitoring. *J. Electrochem. Soc.* **2019**, *166* (9), B3031.

(168) Kim, H.; Awofeso, O.; Choi, S.; Jung, Y.; Bae, E. Colorimetric analysis of saliva–alcohol test strips by smartphone-based instruments using machine-learning algorithms. *Appl. Opt.* **2017**, *56* (1), 84–92.

- (169) Ballard, Z. S.; Joung, H.-A.; Goncharov, A.; Liang, J.; Nugroho, K.; Di Carlo, D.; Garner, O. B.; Ozcan, A. Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors. *NPJ. digital medicine* **2020**, *3* (1), 1–8.
- (170) Luo, R.; Ma, G.; Bi, S.; Duan, Q.; Chen, J.; Feng, Y.; Liu, F.; Lee, J. Machine learning for total organic carbon analysis of environmental water samples using high-throughput colorimetric sensors. *Analyst* **2020**, *145* (6), 2197–2203.
- (171) Koydemir, H. C.; Gorocs, Z.; Tseng, D.; Cortazar, B.; Feng, S.; Chan, R. Y. L.; Burbano, J.; McLeod, E.; Ozcan, A. Rapid imaging, detection and quantification of *Giardia lamblia* cysts using mobile-phone based fluorescent microscopy and machine learning. *Lab Chip* **2015**, *15* (5), 1284–1293.
- (172) Duan, Q.; Lee, J.; Zheng, S.; Chen, J.; Luo, R.; Feng, Y.; Xu, Z. A color-spectral machine learning path for analysis of five mixed amino acids. *Chem. Commun.* **2020**, *56* (7), 1058–1061.
- (173) Lim, H. J.; Saha, T.; Tey, B. T.; Tan, W. S.; Ooi, C. W. Quartz crystal microbalance-based biosensors as rapid diagnostic devices for infectious diseases. *Biosens. Bioelectron.* **2020**, *168*, 112513.
- (174) Tatarko, M.; Muckley, E. S.; Subjakova, V.; Goswami, M.; Sumpter, B. G.; Hianik, T.; Ivanov, I. N. Machine learning enabled acoustic detection of sub-nanomolar concentration of trypsin and plasmin in solution. *Sens. Actuators, B* **2018**, *272*, 282–288.
- (175) Adak, M. F.; Lieberzeit, P.; Jarujamrus, P.; Yumusak, N. Classification of alcohols obtained by QCM sensors with different characteristics using ABC based neural network. *Engineering Science and Technology, an International Journal* **2020**, *23* (3), 463–469.
- (176) van Berkel, C.; Gwyer, J. D.; Deane, S.; Green, N.; Holloway, J.; Hollis, V.; Morgan, H. Integrated systems for rapid point of care (PoC) blood cell analysis. *Lab Chip* **2011**, *11* (7), 1249–1255.
- (177) Huang, X.; Jiang, Y.; Liu, X.; Xu, H.; Han, Z.; Rong, H.; Yang, H.; Yan, M.; Yu, H. Machine learning based single-frame super-resolution processing for lensless blood cell counting. *Sensors* **2016**, *16* (11), 1836.
- (178) Huang, X.; Guo, J.; Wang, X.; Yan, M.; Kang, Y.; Yu, H. A contact-imaging based microfluidic cytometer with machine-learning for single-frame super-resolution processing. *PLoS One* **2014**, *9* (8), e104539.
- (179) Huang, X.; Wang, X.; Yan, M.; Yu, H. A robust recognition error recovery for micro-flow cytometer by machine-learning enhanced single-frame super-resolution processing. *Integration, the VLSI Journal* **2015**, *51*, 208–218.
- (180) Oliver, C. R.; Altemus, M. A.; Westerhof, T. M.; Cheriyan, H.; Cheng, X.; Dziubinski, M.; Wu, Z.; Yates, J.; Morikawa, A.; Heth, J. A platform for artificial intelligence based identification of the extravasation potential of cancer cells into the brain metastatic niche. *Lab Chip* **2019**, *19* (7), 1162–1173.
- (181) Faura, G.; González-Calabuig, A.; del Valle, M. Analysis of amino acid mixtures by voltammetric electronic tongues and artificial neural networks. *Electroanalysis* **2016**, *28* (8), 1894–1900.
- (182) Ccopa Rivera, E.; Swerdlow, J. J.; Summerscales, R. L.; Uppala, P. P. T.; Maciel Filho, R.; Neto, M. R.; Kwon, H. J. Data-Driven Modeling of Smartphone-Based Electrochemiluminescence Sensor Data Using Artificial Intelligence. *Sensors* **2020**, *20* (3), 625.
- (183) Tan, A.; Zhao, Y.; Sivashanmugan, K.; Squire, K.; Wang, A. X. Quantitative TLC-SERS detection of histamine in seafood with support vector machine analysis. *Food Control* **2019**, *103*, 111–118.
- (184) Nguyen, C. Q. Machine Learning for SERS Quantitative Detection of Pyocyanin; UC Irvine, 2018.
- (185) Shi, H.; Wang, H.; Meng, X.; Chen, R.; Zhang, Y.; Su, Y.; He, Y. Setting up a surface-enhanced Raman scattering database for artificial-intelligence-based label-free discrimination of tumor suppressor genes. *Anal. Chem.* **2018**, *90* (24), 14216–14221.
- (186) Mendes, J. J. A., Jr; Vieira, M. E. M.; Pires, M. B.; Stevan, S. L., Jr Sensor fusion and smart sensor in sports and biomedical applications. *Sensors* **2016**, *16* (10), 1569.
- (187) Liggins, M.; Hall, D.; Llinas, J. *Handbook of multisensor data fusion: theory and practice*; CRC Press: 2017.
- (188) Dasarathy, B. V. Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proc. IEEE* **1997**, *85* (1), 24–38.
- (189) Gravina, R.; Alinia, P.; Ghasemzadeh, H.; Fortino, G. Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion* **2017**, *35*, 68–80.
- (190) Bahrepour, M.; Meratnia, N.; Taghikhaki, Z.; Havinga, P. J. Sensor fusion-based activity recognition for Parkinson patients. *Sensor Fusion-Foundation and Applications* **2011**, 171–190.
- (191) Li, C.; Heinemann, P.; Sherry, R. Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection. *Sens. Actuators, B* **2007**, *125* (1), 301–310.