ACS SENSORS

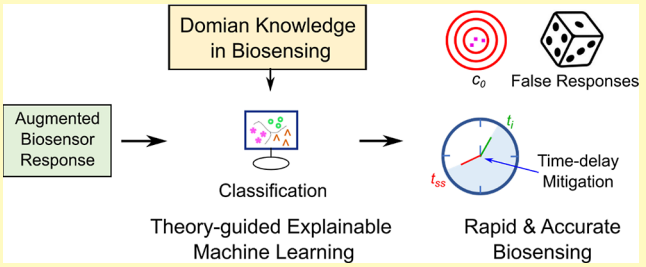
pubs.acs.org/acssensors
Article

Reduction of Biosensor False Responses and Time Delay Using Dynamic Response and Theory-Guided Machine Learning

Junru Zhang, Purna Srivatsa, Fazel Haq Ahmadzai, Yang Liu, Xuerui Song, Anuj Karpatne, Zhenyu Kong, and Blake N. Johnson*

Cite This: ACS Sens. 2023, 8, 4079−4090
Read Online

ABSTRACT: Here, we provide a new methodology for reducing false results and time delay of biosensors, which are barriers to industrial, healthcare, military, and consumer applications. We show that integrating machine learning with domain knowledge in biosensing can complement and improve the biosensor accuracy and speed relative to the performance achieved by traditional regression analysis of a standard curve based on the biosensor steady-state response. The methodology was validated by rapid and accurate quantification of microRNA across the nanomolar to femtomolar range using the dynamic response of cantilever biosensors. Theory-guided feature engineering improved the performance and efficiency of several classification models relative to the performance achieved using traditional feature engineering methods (TSFRESH). In addition to the entire dynamic response, the technique enabled rapid and accurate quantification of the target analyte concentration and false-positive and false-negative results using the initial transient response, thereby reducing the required data acquisition time (i.e., time delay). We show that model explainability can be achieved by combining theory-guided feature engineering and feature importance analysis. The performance of multiple classifiers using both TSFRESH- and theory-based features from the biosensor's initial transient response was similar to that achieved using the entire dynamic response with data augmentation. We also show that the methodology can guide design of experiments for high-performance biosensing applications, specifically, the selection of data acquisition parameters (e.g., time) based on potential application-dependent performance thresholds. This work provides an example of the opportunities for improving biosensor performance, such as reducing biosensor false results and time delay, using explainable machine learning models supervised by domain knowledge in biosensing.

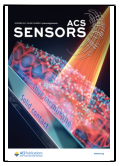


KEYWORDS: *biosensing, chemical sensors, artificial intelligence, biosensor performance, classification, false negative, false positive, measurement confidence*

Biosensor accuracy and speed remain barriers to industrial, healthcare, military, and consumer applications, such as bioprocess control, medical diagnostics, and biosurveillance.[1−3] For example, false-positive and false-negative results associated with rapid tests for SARS-CoV-2 were prevalent during the COVID-19 pandemic.[4] Although polymerase chain reaction (PCR) is a gold-standard bioanalytical method for medical diagnostics, trained dogs performed similar to PCR tests for detecting SARS-CoV-2.[5] Thus, creating reliable, high-performance biosensors for real-world applications (i.e., use in practical environments) may require advances in transducer and biorecognition technologies as well as artificial intelligence (AI)-guided biosensing.

Long-standing challenges and goals in the chemical sensing field, such as sensor drift,[6] reusability, high-throughput sensing,[7] calibration, and mixture analysis,[8] have led to the application of data science models for analyzing chemical sensor response, such as for electronic nose and wearable sensing applications.[9] Recent advances in biosensor design (e.g., paper-based biosensors), manufacturing (e.g., printing[10]), and applications[11,12] (e.g., Internet-of-Things) have driven the use of machine learning in biosensing.[13−17] Data-driven biosensing studies can be classified by the improved biosensor performance metric (e.g., accuracy or speed), the structure of the biosensor data analyzed (e.g., univariate or multivariate time-series (TS) data), or the learning objective (e.g., regression or classification). Relatively accurate, simple, and explainable learning models can often be constructed for analyzing complex data.[18] For example, random forest has been shown to be among the best of over 100 models for the University of California-Irvine Machine Learning Repository.[19]

Received: June 22, 2023
Accepted: September 29, 2023
Published: November 6, 2023



Support vector machine (SVM), linear discriminant analysis, and principal component analysis have been widely used in biosensing applications, such as for response prediction, pathogen classification, analyte detection, and feature extraction.[13,16,20−24] However, although such studies have established the value of using data science to improve biosensor performance, several challenges impede the widespread use of machine learning in biosensing applications. For example, biosensor data associated with practical applications based on traditional experimental design (e.g., calibration protocols) are often sparse and imbalanced. Data labels may also exhibit different uncertainty levels. To date, data science models in biosensing have often not used biosensor theory. Our critical review of the data-driven biosensing literature has revealed a significant opportunity to establish new AI-guided biosensing methodologies that are supported by data augmentation (e.g., generative AI) and supervised by domain knowledge.

Here, we establish a methodology for AI-guided (i.e., data-driven) biosensing that is founded on data augmentation and theory-guided supervised machine learning. The utility and impact of the method are demonstrated by the rapid and accurate quantification of microRNA (miRNA) across the nanomolar to femtomolar range using univariate TS data (i.e., dynamic signal change) obtained from cantilever biosensors. Data augmentation is utilized to address the data sparsity and class imbalance challenges of experimental calibration data. We show that considering the target analyte concentration as a categorical variable enables accurate classification of the dynamic biosensor response and quantification of biosensor false-negative and false-positive responses. We also show that the theory of surface-based biosensors can further improve model accuracy via feature engineering. The methodology also enables rapid and accurate miRNA detection using the biosensor's initial transient response, thereby reducing biosensor time delay, which is a significant barrier in several biosensing applications. Thus, this work helps address critical barriers to the adoption of the biosensor technology associated with false responses and time delay.

MATERIALS AND METHODS

Description of Biosensor Type, Target Analyte, and Biosensor Data. The dynamic response of DNA-functionalized piezoelectric cantilever biosensors associated with selective detection of miRNA let-7a in a continuous-flow format, specifically, resonant frequency (Δf) vs time (t), served as the univariate TS data for this study.[25]

Biosensor Data Preprocessing. The dynamic biosensor signal change was normalized as $\theta(t) = (f(t) - f_i)/(f_f - f_i)$, where f_i is the initial baseline resonant frequency before target analyte binding, f_f is the final baseline resonant frequency after the target analyte binding process reached steady state, and $f(t)$ is the continuously monitored resonant frequency throughout the binding process. Thus, the normalization process removed the traditional feature used for quantifying the target analyte concentration from the TS data (the net biosensor signal change at steady state), thereby scaling both the deterministic and the random components of the signal by the same value. Additionally, the normalization process accounts for performance variance among different biosensors, which is a common characteristic of sensors fabricated by batch manufacturing. Time $t = 0$ corresponded to the start of the transient (i.e., binding) response. The normalized dynamic biosensor response (i.e., the normalized dynamic biosensor signal change) contained both nonstationary and stationary TS data associated with the binding response and the steady-state response, respectively.

Data Augmentation. Data augmentation of the dynamic response before normalization (i.e., Δf vs t), which allowed the augmentation methods to "see" the signal-to-noise ratio (SNR) characteristics of the raw (unscaled) signal, was achieved using jittering, scaling, magnitude warping, window slicing, time warping, and window warping.[26,27] Random numbers used in data augmentation were obtained from a normal distribution. The standard deviation of the distribution varied across the different augmentation techniques based on the associated experimental data. The number of biosensor responses (i.e., TS) in each class was first increased regardless of the class distribution by using the aforementioned augmentation techniques. This increased the data size (i.e., the number of biosensor responses) from 16 to 112, addressing the data sparsity. The data were then further augmented with respect to the class distribution. This increased the number of biosensor responses from 112 to 268, addressing the class imbalance. Thus, the classes (i.e., the concentration bins associated with the calibration standards) for which more calibration data were available contained relatively less augmented data. Details associated with the data augmentation process are provided as Supporting Information.

Traditional Feature Engineering. Features from the unaugmented and augmented TS data sets (i.e., normalized dynamic biosensor responses) were generated by scalable hypothesis testing (TSFRESH; Python),[28] which we refer to as "TSFRESH-based features," based on previous use in chemical sensing applications.[9,29] Imputation was used for analysis of the initial transient response (i.e., a feature with a value of negative infinity, infinity, or NaN was converted to the minimum, maximum, and median feature values obtained across all TS, respectively) and was not used for analysis of the entire biosensor response. For example, the process generated 511 TSFRESH-based features using the entire biosensor response.

Theory-Guided Feature Engineering. Features from the unaugmented and augmented TS data sets were generated from the theory of surface-based affinity biosensors,[30] which we refer to as "theory-based features." For example, the process generated 14 theory-based features (see Table S1), including the rate of change of the normalized dynamic biosensor signal during the initial transient period. Details associated with the theory of surface-based affinity biosensors and model assumptions associated with the biosensor used in this study are provided in the Supporting Information. We note that characteristics of the dynamic signal change have been previously investigated as features for biosensing applications.[21]

Classification of Dynamic Biosensor Response Using Theory-Guided Supervised Machine Learning for Quantification of the Target Analyte Concentration and Probability of False Responses. The TSFRESH- and theory-based features served as the inputs to the supervised learning models for classifying the dynamic biosensor normalized signal change, which we refer to hereinafter as the "dynamic response." The dynamic responses were classified based on the target analyte concentration, specifically, the concentration bin label. The label (y) for each response (i.e., TS) was obtained by logarithmic binning of the target analyte concentration (c) as $y = [-\log_{10}(c/(1\ M))]$, where M refers to molarity, and brackets indicate rounding to the nearest integer. Additional details on the binning method, including inputs and constraints associated with the biosensor performance characteristics, design of experiments (e.g., calibration study design), and the biosensing application, are provided in the Supporting Information. The model performance was evaluated by the average classification scores (i.e., the F1 score, precision, and recall) obtained by cross-validation. The training and test data sets were obtained using stratified k-fold cross-validation (k = 5; i.e., 80 and 20% of the data were used for training and test data, respectively). The fold preserved the class distribution. Cross-validation with a fixed random state ensured that the model performance using TSFRESH, theory, or both features could be compared. The presented F1 score, precision, and recall are the average values for all classes and splits. Grid search was used for hyperparameter tuning. The macro F1 score was used as the tuning metric. Details associated with the search space are provided as Supporting Information.

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

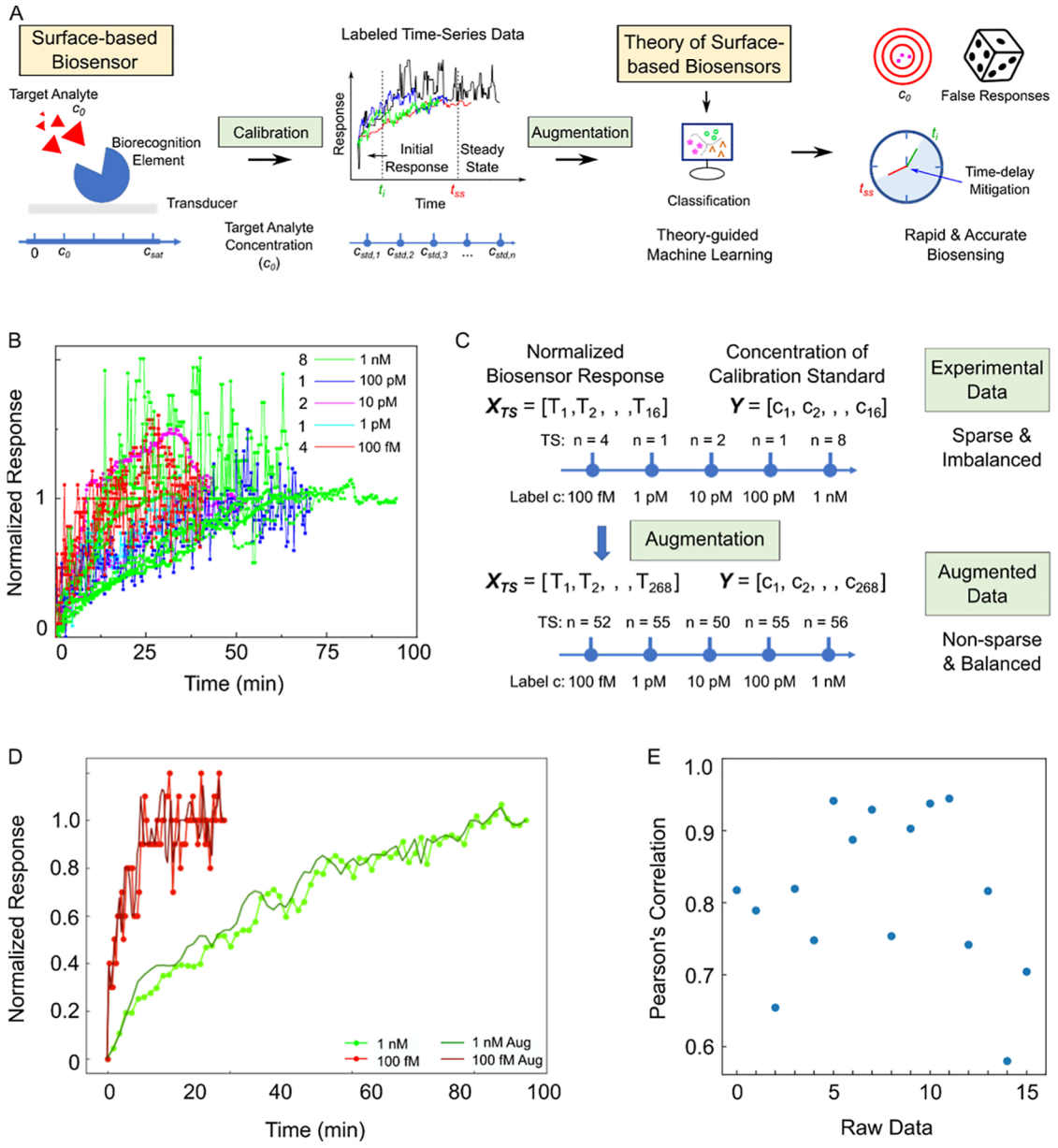


Figure 1. (A) Illustration of rapid and accurate biosensing via classification of the dynamic biosensor response (i.e., normalized dynamic biosensor signal change) using theory-guided machine learning. (B) Representative normalized dynamic biosensor signal change based on the net signal change obtained from the steady-state response. (C) Description of the data structure associated with unaugmented (i.e., experimental) and augmented calibration data sets, as well as the effect of data augmentation on sparsity and class imbalance of the data set. (D) Comparison of the unaugmented and augmented (window warp) normalized dynamic responses. (E) Pearson correlation coefficient associated with experimental and augmented biosensor calibration data.

RESULTS AND DISCUSSION

Quantification of the Target Analyte Concentration and Probability of False-Positive and False-Negative Results via Classification of Dynamic Biosensor Response. As shown in Figure 1A, the goal of this study was to improve the accuracy and speed of surface-based affinity biosensors by integrating domain knowledge in biosensing with data science models using dynamic response. The cantilever biosensor miRNA detection case study involved 16 calibration standards (Figure S1A). The concentration of the calibration standards ranged from 1 nM to 100 fM. Representative normalized dynamic responses for the most dilute and concentrated standards are shown in Figure S1B. While the target analyte concentration is a continuous variable, which is the rationale for regression analysis of standard curve data, the discretized data structure associated with calibration standards and test samples in many biosensing applications suggests that concentration may also be considered as a categorical variable, which enables the use of classification (see Figure 1A and Supporting Information for further details). One motivation for using classification models to complement regression analysis in biosensing is the utility of the classification scores for decision-making, such as in bioprocess control, medical diagnostics, and biosurveillance applications. For example, decision-making in such applications is often based on classifying the concentration of a target analyte, such as a pathogen or biomarker, as present vs absent or in a particular concentration range (i.e., bin), such as below or

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

above an infectious dose or within a diagnostic threshold. Classification models may also consider the possibility of misclassification, such as that caused by sensor drift, nonspecific binding, or user error, which is an important aspect of creating reliable and resilient biosensing tools and methods. As shown in Figure S1C, the steady-state response (i.e., net resonant frequency shift) associated with each calibration standard in the miRNA detection case study exhibited a statistically significant difference, which further supports consideration of the binned miRNA concentration as a categorical variable.

Addressing Challenges of Data Sparsity and Class Imbalance in Biosensor Calibration Data via Data Augmentation. While one-way ANOVA based on the biosensor steady-state response (Figure S1C) suggests that the binned miRNA concentration may be modeled as a categorical variable, the application of classification models is constrained by sparsity and class imbalance of the calibration data. Sparsity and class imbalance of biosensor data are common challenges to the use of data-driven biosensing methods in several applications and may arise from the design of experiments (e.g., calibration study design) or challenges with data management. We first normalized the dynamic biosensor response using the net signal change obtained from the steady-state response, which is the output associated with the traditional standard curve (Figure 1B). While efforts have been made to apply machine learning in biosensing and other bioanalytical applications (see Table S2), they have often employed data augmentation, such as of image and TS data.[31,32] As shown in Figure 1C, the calibration data associated with the miRNA detection case study were sparse and imbalanced. For example, the calibration study generated 16 dynamic biosensor responses with 8, 1, 2, 1, and 4 responses associated with the concentrations of 1 nM, 100 pM, 10 pM, 1 pM, and 100 fM, respectively. The quantity and distribution of calibration standards used in the case study arose from the goal of synergizing the design of experiments with a clinically relevant miRNA-based diagnostic application. Thus, the data set contained an imbalanced class distribution with more samples near the upper and lower limits of the biosensor dynamic range to reflect the variation in let-7a concentration across different diagnostic and therapeutic applications, which varies across several orders of magnitude from the nanomolar to subfemtomolar range.[33−37]

Fluctuations in the biosensor signal, such as those associated with noise, were considered as a characteristic of the normalized signal that may have predictive power. For example, it is not yet established if the processes associated with noise generation in biosensors are dependent on the target analyte concentration, but several studies have shown that sensor noise can have predictive power in chemical sensing and biosensing applications.[38−40] Thus, we did not remove sensor noise from the signal prior to data augmentation or classification. The impact of the sensor noise level (N) is also a potential input and constraint to the binning method, as described in the Supporting Information. Analysis of the SNR among the normalized biosensor responses, which is equivalent to analysis of N since all normalized responses exhibit a total signal change of one, showed no significant difference in N among the different concentrations (classes) examined in this case study. For example, considering N as the standard deviation of the steady-state normalized response, we found that the unaugmented (experimental) data in the classes of 1 nM, 10 pM, and 100 fM exhibited N = 0.090 ± 0.131, 0.109 ± 0.087, and 0.168 ± 0.064, respectively (class sparsity of 100 and 1 pM did not enable analysis of the variance of N at those concentration levels). Thus, no clear trends in the unaugmented data were observed among N and the sensor response times (i.e., time to steady state) or the net signal change at steady state, which were both concentration-dependent. As shown in Figure 1C, several TS data augmentation methods were used to address the data sparsity and class imbalance in the experimental biosensor calibration data, expanding the data set from 16 to 268 TS. Previous studies that have leveraged data augmentation for classification studies involving non-biosensing miRNA applications[41] have generated results from similar-sized data sets containing 40−200 samples.[42,43] The class distribution of the augmented calibration data for concentrations of 1 nM, 100 pM, 10 pM, 1 pM, and 100 fM was 56, 55, 50, 55, and 52, respectively. A comparison of the unaugmented and augmented normalized biosensor responses at 1 nM and 100 fM is shown in Figure 1D. The mean Pearson's correlation coefficient associated with the augmented data is shown in Figure 1E. The correlation score was greater than 0.7 for the majority of samples, indicating that the augmented data were representative of the experimental data. To characterize any potential effects of data augmentation on the SNR, we assessed the noise level among responses in the augmented data sets. We found that the classes 1 nM, 10 pM, and 100 fM for augmented data (window warping) exhibited N = 0.083 ± 0.120, 0.108 ± 0.092, and 0.142 ± 0.038, respectively, which were comparable to those of unaugmented data. Thus, data augmentation did not have a clear impact on SNR based on this analysis.

Classification of Dynamic Biosensor Response Using Traditional Machine Learning Models and Unaugmented Calibration Data. Having motivated the use of classification in biosensing and addressed the challenges of sparsity and class imbalance in experimental biosensor data by augmentation, we next classified the unaugmented normalized dynamic biosensor responses using several established classification models. Figure S2 validates the step size uniformity of the TS, which is assumed in the TSFRESH analysis. As shown in Figure S3, TSFRESH-based features were obtained from the biosensor's entire dynamic response and the initial transient response. The nonstationary and stationary portions of the normalized dynamic response are shown in Supporting Information (see Figure S4). TSFRESH generates a number of features from a TS, which depend on its length and sampling frequency. Therefore, the entire biosensor responses and the initial transient responses associated with the experimental and augmented data sets generate unique TSFRESH features, some of which may be common features. Detailed descriptions of TSFRESH-based features are provided in the Supporting Information. The F1 score is a weighted measure of model classification capability that accounts for the precision and recall (i.e., the extent of false positives and false negatives) and is widely used to evaluate the performance of machine learning models for biosensing applications.[44,45] Precision is a measure of positive predictive value (i.e., false positives), and recall, also known as a true positive rate or sensitivity, is a measure of false negatives. High precision indicates a low number of false positives, and high recall indicates a low number of false negatives. Precision and recall have also been used in the analysis of machine learning models

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

Table 1. Classification Results Using TSFRESH-Based Features, Theory-Based Features, and Both Feature Types Obtained from Unaugmented Data[a]

	TSFRESH			theory			TSFRESH & theory		
model	F1 score	precision	recall	F1 score	precision	recall	F1 score	precision	recall
SVM	0.427	0.376	0.532	***0.689***	***0.684***	***0.75***	0.427	0.376	0.532
logistic regression	0.236	0.2	0.298	*0.491*	*0.472*	*0.582*	0.224	0.188	0.298
decision tree	0.413	0.41	0.482	0.333	0.352	0.398	0.378	0.466	0.434
random forest	0.533	0.54	0.58	0.4	0.444	0.482	0.533	**0.54**	0.58
gradient boost	**0.544**	**0.572**	**0.6**	*0.6*	*0.594*	*0.682*	**0.547**	0.51	**0.632**
XGBoost	0.444	0.456	0.484	*0.6*	*0.594*	*0.682*	0.444	0.456	0.484

[a]Bold black font indicates the best-performing models using TSFRESH-based features, theory-based features, and both features. Italic font indicates the models that exhibited improved performance using theory-vs TSFRESH-based features and both features. Underline font indicates the best performance across all features and combinations for each model.

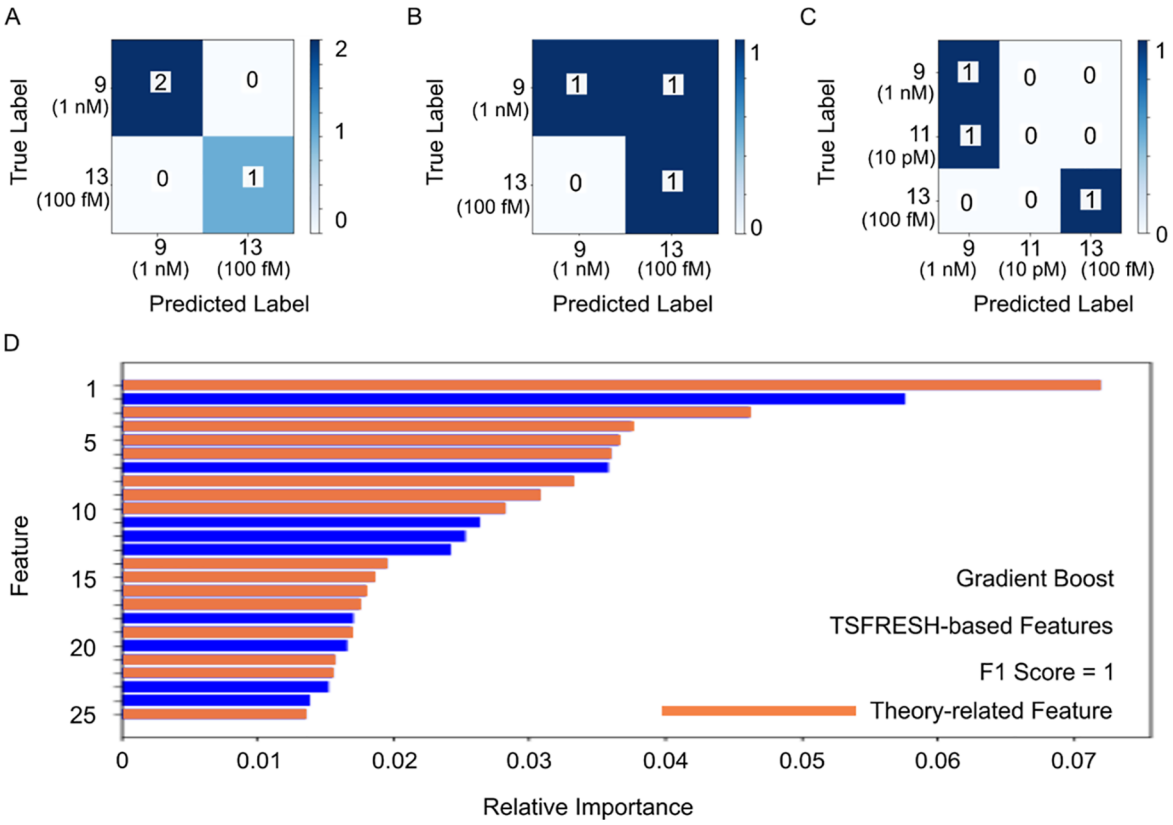


Figure 2. Confusion matrices associated with classification results using TSFRESH-based features (A; gradient boost), theory-based features (B; SVM), and both features (C; XGBoost) (also see Table 1). (D) Feature importance plot of the gradient boost classifier using TSFRESH-based features for the highest performing split (F1 score = 1). The relative importance of the top 25 TSFRESH-based features is shown, and orange color indicates a TSFRESH-based feature that is related to domain knowledge in biosensing (feature descriptions provided in the Supporting Information).

for biosensing applications.[3,46,47] The ability to quantify and reduce false results, particularly false negatives, is critical in biosensing applications, such as medical diagnostics and biosurveillance, given the potential severity of consequences associated with decisions made from false-negative results,[48,49] such as taking no action to a present disease or biothreat. The average classification score for each model among the different splits used for cross-validation is provided in Table 1 (high and low scores among the different splits are shown in Table S3). Cross-validation is a useful method for gauging a model's ability to generalize to unseen examples (i.e., to assess the generalizability of the model to an unseen test set) and is also useful for analysis of small data sets that require direct comparison of results, such as obtained without and with data augmentation. Gradient boost gave the best performance in classifying the experimental dynamic biosensor response (three concentration classes) with TSRESH-based features (see Table 1), which corresponded to an average F1 score, precision, and recall of 0.544, 0.572, and 0.6, respectively. The confusion matrix for the gradient boost classifier for a representative data split is provided in Figure 2A and shows that the biosensor responses in the most concentrated and dilute samples [class 9 (1 nM) and class 13 (100 fM)] were classified with 100% accuracy. We note that the data splits associated with cross-validation for the unaugmented data set may not contain three classes in the test set as they are randomly generated.

Classification of Dynamic Biosensor Response Using Theory-Guided Machine Learning Models and Unaug-

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

Table 2. Model Performance Using TSFRESH-Based Features, Theory-Based Features, and Both Feature Types Obtained with Data Augmentation (i.e., Augmented Biosensor Calibration Data)[a]

	TSFRESH and aug			theory and aug			TSFRESH, theory, and aug		
model	F1 score	precision	recall	F1 score	precision	recall	F1 score	precision	recall
SVM	0.764	0.79	0.76	0.671	0.7	0.67	*0.77*	*0.796*	*0.768*
logistic regression	0.767	0.774	0.768	0.638	0.652	0.638	*0.775*	*0.784*	*0.774*
decision tree	0.853	0.862	0.854	0.707	0.726	0.706	*0.874*	*0.888*	*0.872*
random forest	**0.933**	**0.938**	**0.934**	0.777	0.79	0.776	**0.933**	0.938	0.932
gradient boost	**0.933**	**0.938**	0.932	**0.78**	**0.794**	**0.78**	***0.933***	***0.94***	***0.934***
XGBoost	0.901	0.912	0.902	0.759	0.784	0.758	0.898	0.904	0.898

[a]Bold black font indicates the best-performing models. Italic font indicates models that exhibited improved performance using both features vs only TSFRESH- or theory-based features (i.e., a single feature type). Italic, bold font indicates the best performance among all features and combinations. Abbreviations: augmentation (aug).

mented Calibration Data. We next demonstrated that domain knowledge in biosensing, specifically the theory of surface-based biosensors, can be leveraged to further improve the model performance by theory-guided feature engineering. As shown in Figures 1B and S1B, the rate of change of the initial transient response appeared to be dependent on the target analyte concentration, which motivates investigating the model performance using theory-based features from both the entire response and the initial transient response (Figure S3). As shown in eq 1, the target analyte concentration (c_0), which is also the output of the machine learning model ($Y = c_0$), is proportional to the rate of change of the dimensionless dynamic signal change with respect to time in the initial transient period ($\frac{\partial\left(\frac{\Delta f_n}{f_n}\right)}{\partial t}|_{t\sim 0}$) as

$$c_0 \sim \frac{\partial\left(\frac{\Delta f_n}{f_n}\right)}{\partial t}|_{t\sim 0} \quad (1)$$

A detailed discussion of the theory of surface-based affinity biosensors, assumptions for the biosensor type used in this case study (milli-cantilever biosensors), and derivation of eq 1 are provided in the Supporting Information. Equation 1 may provide insights into the potential for using nonequilibrium methods for analysis of biosensor data, which is an emerging area.[50]

SVM exhibited the best model performance with F1 score, precision, and recall of 0.689, 0.684, and 0.75, respectively, using theory-based features (see Table 1). We also found that SVM, logistic regression, gradient boost, and XGBoost exhibited improved performance using theory-based features relative to the performance using TSFRESH features. When comparing the model performance using TSFRESH- vs theory-based features, one should consider the number of features (511 TSFRESH- vs 14 theory-based features). The comparable performance achieved using over an order of magnitude fewer features highlights the value of leveraging domain knowledge to improve the efficiency (e.g., reduced computing power) and explainability of machine learning models for AI-guided biosensing applications. As shown by the confusion matrix in Figure 2B, the SVM model misclassified one response in class 9 (1 nM) as class 13 (100 fM), which was potentially caused by limited available training data in the unaugmented data set due to sparsity.

Table 1 also summarizes the model performance using both feature types. Gradient boost exhibited the best performance based on the F1 score and recall, while random forest exhibited the highest precision. The confusion matrix associated with XGBoost is highlighted in Figure 2C, given that the data split included all three classes (i.e., concentrations) in the test data set (the F1 score was also the highest among the other models with such a data split). The biosensor responses associated with 1 nM (class 9) and 100 fM (class 13) were successfully classified by XGBoost, while biosensor responses associated with the 10 pM (class 11) class were misclassified as class 9 (1 nM). Among the use of TSFRESH-based, theory-based, and both feature types, SVM exhibited the best performance by using only theory-based features. This result demonstrated the value of leveraging domain knowledge in AI-guided biosensing when using sparse data sets. The performance obtained using the unaugmented data agreed well with reported values, such as for virus classification (~70% accuracy),[22] classification of wine quality (67.3 and 60.5% accuracy, respectively),[51] and disease classification (~50% accuracy).[52]

Feature importance analysis, which can be performed on tree-based models (e.g., decision tree, random forest, and gradient boost), is a useful method for explaining the model performance using both feature types. For example, Figure 2D highlights the 25 most important TSFRESH-based features of the gradient boost classifier, which exhibited an F1 score of 1. While TSFRESH-based features are generated by statistical hypothesis testing, several TSFRESH-based features are indirectly related to biosensor theory, which we refer to as "theory-related features." For example, theory-related TSFRESH-based features, including the standard deviation of linear trend (feature 1) and the aggregate variance of the 0.2 to 0.4 quantile (feature 3), are similar to the theory-based features of linear trend (see eq 1) and dynamic SNR, which can be identified through first principles. One of the most important TSFRESH-based features was the variance of the 0.2 to 0.4 quantile of normalized dynamic biosensor response, which corresponds to the biosensor initial transient response, suggesting that the initial transient response contains features that significantly contribute to classification accuracy in biosensing applications. The feature importance analysis of the gradient boost model using only TSFRESH-based features provides insights into the value of using both theory-guided feature engineering and the biosensor transient response for classification tasks in biosensing applications. Based on these insights, we examined the performance of the random forest model, which exhibited high precision with and without the steady-state response. Interestingly, the performance without the steady-state response was relatively superior, indicating that it may not be necessary to utilize the entire biosensor response for classification of the target analyte concentration and

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

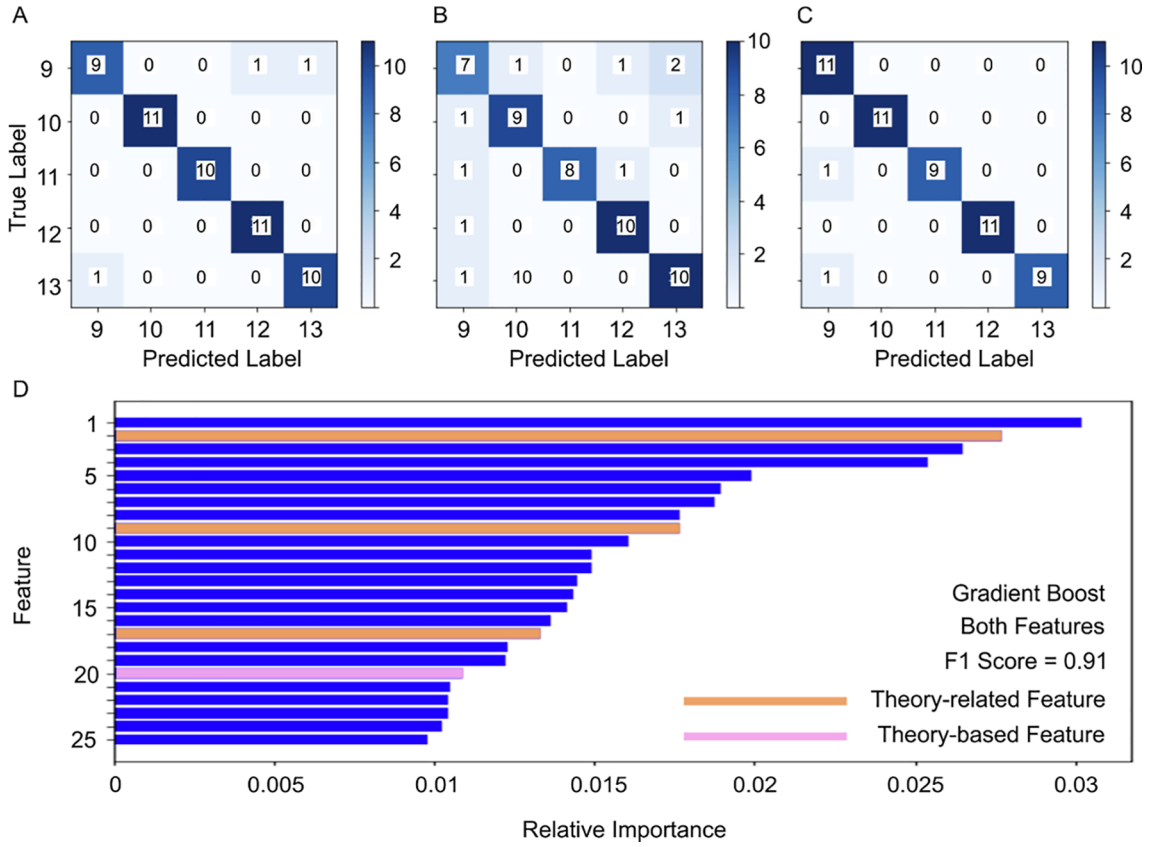


Figure 3. Confusion matrices associated with the classification results with augmented data using TSFRESH-based features (A; random forest), theory-based features (B; gradient boost), and both feature types (C; gradient boost). (D) Feature importance plot of the gradient boost classifier for augmented data for a representative split using both features. Orange indicates a theory-related feature, and pink indicates a theory-based feature (feature descriptions provided in the Supporting Information).

quantification of false-positive and false-negative responses. Overall, the results shown in Table 1 and Figure 2 highlight the value of incorporating domain knowledge in biosensing for classification of unaugmented calibration data and suggest an opportunity for using data augmentation and data from the initial transient response to improve the model performance. To summarize, the use of features obtained from theory-guided feature engineering (i.e., theory-based features) improved the performance and efficiency of SVM, logistic regression, gradient boost, and XGBoost classifiers. Specifically, the F1 score, precision, and recall of all models investigated using unaugmented calibration data increased by an average of 18.2, 18.5, and 19.6%, respectively, relative to the performance using features obtained by traditional feature engineering methods (e.g., TSFRESH-based features). Feature importance analysis also revealed that the theory of surface-based biosensors can help identify critical features for the accurate classification of biosensor data.

Effect of Data Augmentation on Classification of Dynamic Biosensor Response Using Traditional and Theory-Guided Machine Learning. We next examined if biosensor performance could be further improved by addressing the sparsity and class imbalance challenges in the experimental data set using data augmentation. As shown by a comparison of Tables 1 and 2, the model performance using all features improved significantly with data augmentation. Random forest exhibited the best performance using TSFRESH-based features from the augmented data with an average F1 score, precision, and recall of 0.933, 0.938, and 0.934, respectively. Figure 3A presents the confusion matrix of the random forest classifier using TSFRESH-based features, which shows that two responses in class 9 (1 nM) were misclassified as classes 12 (1 pM) and 13 (100 fM). Additionally, one response in class 13 (100 fM) was misclassified as class 9 (1 nM). All 51 other responses were classified correctly. Gradient boost exhibited the best performance using theory-based features from the augmented data set, yielding an average F1 score, precision, and recall of 0.78, 0.794, and 0.78, respectively. Figure 3B shows the confusion matrix of the gradient boost classifier corresponding to F1 = 0.82 (the highest performing split) using theory-based features, which provided misclassifications different from those of the random forest classifier. As shown in Table 2, the performance of four models improved with data augmentation when using both feature types (i.e., TSFRESH- and theory-based features). Gradient boost exhibited the best performance using both features from the augmented data set, yielding an average F1 score, precision, and recall of 0.933, 0.94, and 0.934, respectively. Figure 3C shows the confusion matrix of the gradient boost classifier corresponding to F1 = 0.96 (the highest performing split) using both feature types, which accurately classified all 53 responses except two. For example, one response in class 11 (10 pM) and one response in class 13 (100 fM) were misclassified as class 9 (1 nM). Figure 3D shows the feature importance plot for the gradient boost classifier corresponding to F1 = 0.91 obtained using both feature types. A theory-guided feature was found to be the most important of the top 25 features used in the classification

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

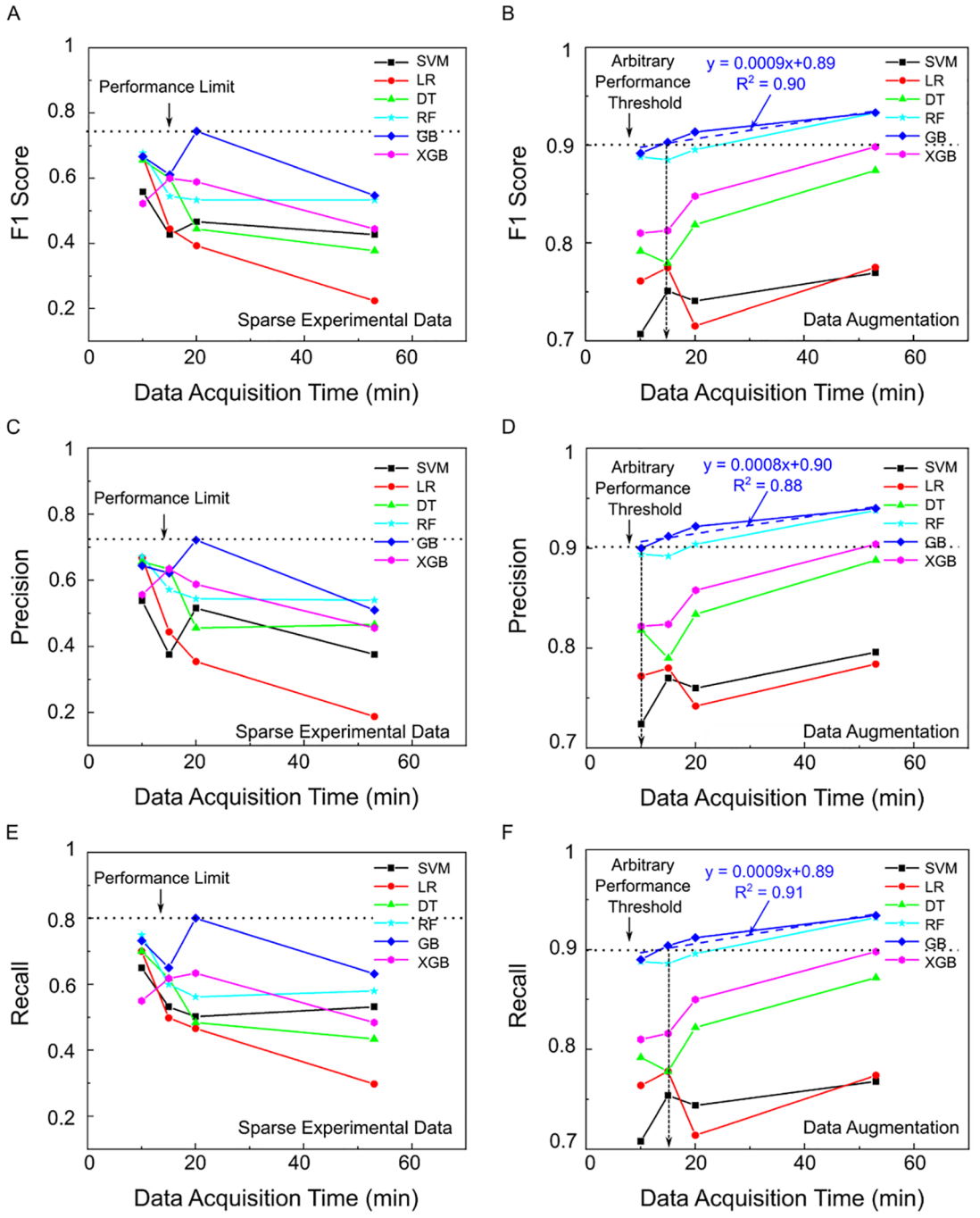


Figure 4. Model performance comparison using both TSFRESH- and theory-based features via F1 score, precision, and recall with sparse experimental data (A,C,E) and augmented data (B,D,F), respectively (data at t = 53 min represent the average performance obtained using the entire biosensor response for all classes, which exhibited a range across the different concentrations examined). Arbitrary thresholds for the biosensor performance based on metrics of F1 score, precision, and recall are represented as dotted black lines. Abbreviations: support vector machine (SVM), logistic regression (LR), decision tree (DT), random forest (RF), gradient boost (GB), and XGBoost (XGB).

model. As shown by comparison of Figures 2D and 3D, data augmentation decreased the magnitude of the relative feature importance of top features by approximately a factor of 2, indicating that more features played a role in the classification model as the data size increased. Thus, no trend was observed in the misclassification results toward under- or over-prediction of the concentration, both of which occurred. In some applications, under- and over-prediction of concentration via misclassification may be viewed as a type of false negative and false positive, respectively. The gradient boost classifier misclassifications exhibited the lowest relative errors (i.e., the difference between predicted and true binned values). We also found that ensemble models, including random forest, gradient boost, and XGBoost, outperformed the relatively simpler models when using both feature types. The small differences observed between the lowest and highest F1 scores among different data splits of the ensemble models (see Table S4) indicated that the models were not overfitting and provided confidence that this methodology could be extended to other biosensing applications. The results obtained using data

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

Table 3. Model Performance Using TSFRESH-Based (A) and Both TSFRESH- and Theory-Based Features (B) from the First 20, 15, and 10 min of the Initial Biosensor Transient Response Obtained with Data Augmentation[a]

A	20 min TSFRESH and aug			15 min TSFRESH and aug			10 min TSFRESH and aug		
model	F1 score	precision	recall	F1 score	precision	recall	F1 score	precision	recall
SVM	0.737	0.758	0.74	0.758	0.764	0.75	0.717	0.736	0.72
logistic regression	0.718	0.738	0.722	0.788	0.802	0.792	0.781	0.79	0.78
decision tree	0.797	0.806	0.8	0.783	0.804	0.784	0.769	0.78	0.766
random forest	0.891	0.896	0.89	0.886	0.894	0.884	0.879	0.886	0.88
gradient boost	***0.922***	***0.93***	***0.92***	**0.899**	**0.904**	**0.9**	**0.891**	**0.896**	***0.892***
XGBoost	0.85	0.862	0.854	0.815	0.824	0.816	0.797	0.81	0.796
B	20 min TSFRESH, theory, and aug			15 min TSFRESH, theory, and aug			10 min TSFRESH, theory, aug		
model	F1 score	precision	recall	F1 score	precision	recall	F1 score	precision	recall
SVM	0.741	0.76	0.744	0.751	0.77	0.754	0.707	0.724	0.708
logistic regression	0.715	0.742	0.714	0.775	0.78	0.778	0.761	0.772	0.764
decision tree	0.819	0.834	0.822	0.779	0.79	0.778	0.792	0.818	0.792
random forest	0.896	0.904	0.896	0.885	0.892	0.886	0.888	0.894	0.888
gradient boost	**0.914**	**0.922**	**0.912**	***0.903***	***0.912***	***0.904***	***0.892***	***0.9***	**0.89**
XGBoost	0.848	0.858	0.85	0.813	0.824	0.816	0.81	0.822	0.81

[a]Bold font indicates the best-performing model with TSFRESH-based features (A) and both features (B). Italic, bold font indicates the best model performance of all models and all combinations of features. Abbreviations: augmentation (Aug).

augmentation suggest that (1) the methodology was resilient to noise generated by piezoelectric cantilever biosensors, and (2) noise from piezoelectric cantilever biosensors may have predictive power. For example, all the data augmentation techniques (e.g., jittering) can be summarized as techniques that capture trends with added noise. The results with data augmentation show that increasing the data size improved the classification scores. This suggests that the method is resilient to sensor noise since the random processes associated with noise generation in augmented data differ from the unaugmented (experimental) TS data. The improvements in F1 score, precision, and recall achieved by data augmentation using both feature types demonstrate the importance of data size and class distribution in AI-guided biosensing applications. In summary, the use of both features (i.e., TSFRESH- and theory-based features) with data augmentation improved the F1 score, precision, and recall of all classifiers by an average of 7.4, 7, and 7.3%, respectively, relative to the performance using only TSFRESH- or theory-based features.

Rapid Classification of the Initial Transient Response via Theory-Guided Machine Learning with Unaugmented Data. Having demonstrated the opportunity for improving biosensor accuracy and reducing false results of biosensors through classification scores using the entire biosensor response, we next examined the ability to reduce the measurement time (i.e., biosensor time delay) required to characterize the target analyte concentration, which is typically constrained by the target binding process, by using the initial transient response. As shown in Table S5, all models investigated enabled quantification of the target analyte concentration by classification of the biosensor initial transient response with comparable performance to that obtained from the entire response. While the imputation function was used for analysis of the initial transient response, it did not significantly impact model performance. Table S5A shows the model performance using the first 20 min of the initial transient response with unaugmented data. The gradient boost and XGBoost classifiers performed similarly using only TSFRESH-based features and both feature types, which was better than that obtained using only theory-based features. Among all models, gradient boost exhibited the best performance with an average F1 score, precision, and recall of 0.744, 0.722, and 0.8, respectively. As shown in Table S5B, SVM using theory-based features from the first 15 min of the initial transient response gave the best performance among all the models with an average F1 score, precision, and recall of 0.702, 0.666, and 0.768, respectively. Table S5C shows the model performance using the first 10 min of the initial transient response. Random forest gave the best performance using the first 10 min of the initial transient response with an average F1 score, precision, and recall of 0.678, 0.672, and 0.75, respectively. In comparison, the best performance using the entire response with unaugmented data was characterized by an F1 score = 0.689 (see Table 1), whereas the best performance using the first 20, 15, and 10 min of the initial transient response was characterized by F1 score = 0.744, 0.702, and 0.678, respectively. In summary, use of the initial transient response improved the performance of several classifiers relative to that achieved using the entire dynamic response by an average of 19.0, 20.1, and 18.7% (10 min), 11.2, 13.0, and 10.7% (15 min), and 8.1, 9.6, and 7.1% (20 min) for the unaugmented data set (see Figure 4.A,C,E). Theory-based features from the first 15 and 20 min and the first 10 and 15 min of the initial transient response were found to exhibit the best performance for logistic regression and SVM, respectively.

Classification of the Initial Transient Response via Theory-Guided Machine Learning with Data Augmentation. We next examined the effect of data augmentation on the model performance using the initial transient response. The model performance using TSFRESH-based features and both feature types from the first 20, 15, and 10 min of the initial transient response is shown in Table 3 (results obtained with only theory-based features are provided in Table S6). Tables 3A and B highlight the effect of data augmentation on the performance of the gradient boost classifier using TSFRESH-based features and both feature types from the initial transient response. For example, the gradient boost classifier provided an average F1 score, precision, and recall of 0.922, 0.93, and 0.92, as well as 0.903, 0.912, and 0.904, using

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090

features from the first 20 and 15 min of the initial transient response, respectively. Comparison of Tables 3 and S6 shows that data augmentation resulted in an average F1 score improvement of 7.2, 6.6, and 5.5% using both feature types vs. only TSFRESH- or theory-based features for the first 20, 15, and 10 min of the initial transient response, respectively.

As shown in Figure 4A, the six common classifiers investigated showed a trend of improving performance with decreasing data acquisition time using experimental data, which was sparse and imbalanced. Four of the six models (SVM, logistics regression, decision tree, and random forest) showed a trend of increasing F1 score with decreasing data acquisition time. Gradient boost and XGBoost exhibited a nonmonotonic relationship between F1 score and data acquisition time. However, both models gave better performance using the initial transient response relative to that using the entire biosensor response. As shown in Figure 4B, the use of data augmentation affected the relationship between model performance and data acquisition time. Specifically, with data augmentation, the F1 score of the six models was positively correlated with the data acquisition time but with a reduced slope (see Figure 4B). As also shown in Figure 4B, the models performed comparably well using the initial transient response and the entire biosensor response (i.e., data augmentation reduced the sensitivity of the model performance to the duration of the biosensor response utilized). Logistic regression using the first 15 min of the initial transient response gave a slightly better performance than using the entire response or the first 20 and 10 min of the initial transient response. The precision and recall exhibited similar trends as F1 score for both the unaugmented and augmented data, as shown in Figure 4C−F. Overall, the results show that the integration of theory-guided machine learning (i.e., use of only theory-based features or both feature types) and data augmentation can significantly improve the biosensor performance, specifically, improving accuracy, speed, and reliability (i.e., reducing the probability of false results). This work also illustrates AI-guided design of experiments for biosensing applications (e.g., required data acquisition time) based on potential predefined application-dependent performance thresholds, such as thresholds in accuracy or false results (see Figure 4).

In addition to improving the biosensor performance, this work may help provide new insights into the predictive power of biosensing methods based on nonequilibrium signal analysis, particularly based on initial velocity (i.e., "slope" or rate of change of the sensor signal). For example, while several approaches based on transient signal analysis have demonstrated promise for reducing biosensor measurement time,[53] they are often based on extracting a single feature or a small number of features from TS data, such as slope (i.e., velocity). Our work extends that principle by highlighting the availability of many features in biosensor TS data that exhibit predictive power through feature importance analysis, including the rate of change of the normalized signal (i.e., a velocity-related feature) and the signal variance in initial time windows (see Tables S1 and S9 for the detailed feature descriptions). Thus, this work provides support for the predictive power of the initial velocity in biosensing applications by leveraging the power of theory-guided data science models. One significant difference between established biosensing methods based on nonequilibrium signal analysis, such as based on initial velocity using regression analysis, and this work is the use of a labeled TS data structure in this study, which enables quantification of the target analyte concentration with a quantifiable probability of false results using classification.

CONCLUSIONS

Here, we show that theory-guided machine learning facilitates rapid and accurate biosensing with quantifiable false-negative and false-positive results based on the classification of dynamic biosensor responses, including the initial transient response. Results were obtained using experimental biosensor data with and without data augmentation. We found that the integration of domain knowledge in biosensing (e.g., multiphysics models of biosensor function) improved the performance of machine learning models, including accuracy and efficiency. Future work will investigate the dependence of the model performance (e.g., accuracy) and minimum time delay on the number of theory-guided features utilized as well as examine the effect of including labeled false responses in the classification task (e.g., caused by drift or nonspecific binding). The reported methodology may also provide a path to help redefine and improve the performance of other types of biosensors and other performance attributes beyond speed and accuracy (e.g., sensitivity). Overall, this work shows how theory-guided machine learning can help improve the performance of biosensors and potentially the understanding of processes that generate false responses as well as how domain knowledge in biosensing can help improve the performance and explainability of data-driven biosensing strategies.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssensors.3c01258.

Biosensing via theory-guided nondeep learning—SI—ACS sens, additional details associated with the methodology (classification problem and binning process), feature engineering, results with nonaugmented data, and theory of surface-based affinity biosensors (PDF)

AUTHOR INFORMATION

Corresponding Author

Blake N. Johnson − *Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States; School of Neuroscience, Department of Materials Science and Engineering, and Department of Chemical Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0003-4668-2011; Phone: (540) 231-0755; Email: bnj@vt.edu; Fax: (540) 231-3322

Authors

Junru Zhang − *Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Purna Srivatsa − *Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061, United States*

Fazel Haq Ahmadzai − *Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0009-0006-7788-8249

Yang Liu − *Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061,*

United States; School of Neuroscience, Virginia Tech, Blacksburg, Virginia 24061, United States

Xuerui Song − *Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Anuj Karpatne − *Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061, United States*

Zhenyu Kong − *Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssensors.3c01258

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

B.N.J. is grateful for the support of the National Science Foundation (CBET-2144310). This work was also supported by GlycoMIP, a National Science Foundation Materials Innovation Platform funded through Cooperative Agreement DMR-1933525. Z.K. is grateful for the support of the Department of Defense (Grant No. N00014-19-1-2728).

■ REFERENCES

(1) Dervisevic, M.; Alba, M.; Prieto-Simon, B.; Voelcker, N. H. Skin in the diagnostics game: Wearable biosensor nano- and microsystems for medical diagnostics. *Nano Today* **2020**, *30*, 100828.

(2) Tan, M.; Xu, Y.; Gao, Z.; Yuan, T.; Liu, Q.; Yang, R.; Zhang, B.; Peng, L. Recent Advances in Intelligent Wearable Medical Devices Integrating Biosensing and Drug Delivery. *Adv. Mater.* **2022**, *34* (27), 2108491.

(3) Cui, F.; Yue, Y.; Zhang, Y.; Zhang, Z.; Zhou, H. S. Advancing biosensors with machine learning. *ACS Sens.* **2020**, *5* (11), 3346−3364.

(4) Patel, S. K.; Surve, J.; Parmar, J.; Ahmed, K.; Bui, F. M.; Al-Zahrani, F. A. Recent Advances in Biosensors for Detection of COVID-19 and Other Viruses. *IEEE Rev. Biomed. Eng.* **2023**, *16*, 22−37.

(5) Hag-Ali, M.; AlShamsi, A. S.; Boeijen, L.; Mahmmod, Y.; Manzoor, R.; Rutten, H.; Mweu, M. M.; El-Tholoth, M.; AlShamsi, A. A. The detection dogs test is more sensitive than real-time PCR in screening for SARS-CoV-2. *Commun. Biol.* **2021**, *4* (1), 686.

(6) Martinelli, E.; Magna, G.; De Vito, S.; Di Fuccio, R.; Di Francia, G.; Vergara, A.; Di Natale, C. An adaptive classification model based on the Artificial Immune System for chemical sensor drift mitigation. *Sens. Actuators, B* **2013**, *177*, 1017−1026.

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

(8) Eklöv, T.; Lundström, I. Gas mixture analysis using a distributed chemical sensor system. *Sens. Actuators, B* **1999**, *57* (1−3), 274−282.

(9) Fairbairn, C. E.; Kang, D.; Bosch, N. Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory. *Drug Alcohol Depend.* **2020**, *216*, 108205.

(10) Ruan, X.; Wang, Y.; Cheng, N.; Niu, X.; Chang, Y. C.; Li, L.; Du, D.; Lin, Y. Emerging applications of additive manufacturing in biosensors and bioanalytical devices. *Adv. Mater. Technol.* **2020**, *5* (7), 2000171.

(11) Efendić, H.; Bećirović, L. S.; Deumić, A.; Pokvić, L. G. Biosensors in monitoring public health: Industry 4.0 applications-a review. *IFAC-Pap.* **2022**, *55* (4), 38−44.

(12) Mujawar, M. A.; Gohel, H.; Bhardwaj, S. K.; Srinivasan, S.; Hickman, N.; Kaushik, A. Nano-enabled biosensing systems for intelligent healthcare: towards COVID-19 management. *Mater. Today Chem.* **2020**, *17*, 100306.

(13) Zhao, Y.; Ashcroft, B.; Zhang, P.; Liu, H.; Sen, S.; Song, W.; Im, J.; Gyarfas, B.; Manna, S.; Biswas, S.; et al. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotechnol.* **2014**, *9* (6), 466−473.

(14) Taniguchi, M. Combination of Single-Molecule Electrical Measurements and Machine Learning for the Identification of Single Biomolecules. *ACS Omega* **2020**, *5* (2), 959−964.

(15) Massah, J.; Asefpour Vakilian, K. An intelligent portable biosensor for fast and accurate nitrate determination using cyclic voltammetry. *Biosyst. Eng.* **2019**, *177*, 49−58.

(16) 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), 5920.

(17) Schackart, K. E.; Yoon, J.-Y. Machine Learning Enhances the Performance of Bioreceptor-Free Biosensors. *Sensors* **2021**, *21* (16), 5519.

(18) Semenova, L.; Rudin, C.; Parr, R. On the Existence of Simpler Machine Learning Models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*; Association for Computing Machinery: Seoul, Republic of Korea, 2022; pp 1827−1858.

(19) Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, *15* (1), 3133−3181.

(20) Rong, Y.; Padron, A.; Hagerty, K.; Nelson, N.; Chi, S.; Keyhani, N.; Katz, J.; Datta, S.; Gomes, C.; McLamore, E. Post hoc support vector machine learning for impedimetric biosensors based on weak protein-ligand interactions. *Analyst* **2018**, *143* (9), 2066−2075.

(21) 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.

(22) Arima, A.; Tsutsui, M.; Harlisa, I. H.; Yoshida, T.; Tanaka, M.; Yokota, K.; Tonomura, W.; Taniguchi, M.; Okochi, M.; Washio, T.; Kawai, T. Selective detections of single-viruses using solid-state nanopores. *Sci. Rep.* **2018**, *8* (1), 16305.

(23) Gonzalez-Navarro, F. F.; Stilianova-Stoytcheva, M.; Renteria-Gutierrez, L.; Belanche-Muñoz, L.; Flores-Rios, B. L.; Ibarra-Esquer, J. E. Glucose oxidase biosensor modeling and predictors optimization by machine learning methods. *Sensors* **2016**, *16* (11), 1483.

(24) Li, Z.; Paul, R.; Ba Tis, T.; Saville, A. C.; Hansel, J. C.; Yu, T.; Ristaino, J. B.; Wei, Q. Non-invasive plant disease diagnostics enabled by smartphone-based fingerprinting of leaf volatiles. *Nat. Plants* **2019**, *5* (8), 856−866.

(25) Johnson, B. N.; Mutharasan, R. Sample Preparation-Free, Real-Time Detection of microRNA in Human Serum Using Piezoelectric Cantilever Biosensors at Attomole Level. *Anal. Chem.* **2012**, *84* (23), 10426−10436.

(26) Iwana, B. K.; Uchida, S. An empirical survey of data augmentation for time series classification with neural networks. *PLoS One* **2021**, *16* (7), No. e0254841.

(27) Um, T. T.; Pfister, F. M.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM international conference on multimodal interaction*; Association for Computing Machinery, 2017; pp 216−220.

(28) Christ, M.; Braun, N.; Neuffer, J.; Kempa-Liehr, A. W. Time series feature extraction on basis of scalable hypothesis tests (tsfresh-a python package). *Neurocomputing* **2018**, *307*, 72−77.

(29) Ariss, T.; Fairbairn, C. E.; Bosch, N. Examining new-generation transdermal alcohol biosensor performance across laboratory and field contexts. *Alcohol: Clin. Exp. Res.* **2023**, *47* (1), 50−59.

(30) Squires, T. M.; Messinger, R. J.; Manalis, S. R. Making it stick: convection, reaction and diffusion in surface-based biosensors. *Nat. Biotechnol.* **2008**, *26* (4), 417−426.

(31) Liu, L.; Zhan, X.; Wu, R.; Guan, X.; Wang, Z.; Zhang, W.; Pilanci, M.; Wang, Y.; Luo, Z.; Li, G. Boost AI Power: Data Augmentation Strategies With Unlabeled Data and Conformal

Prediction, a Case in Alternative Herbal Medicine Discrimination With Electronic Nose. *IEEE Sens. J.* **2021**, *21* (20), 22995−23005.
(32) Forestier, G.; Petitjean, F.; Dau, H. A.; Webb, G. I.; Keogh, E. In Generating synthetic time series to augment sparse datasets; *2017 IEEE international conference on data mining (ICDM)*, IEEE: 2017; pp 865−870.
(33) Ouyang, T.; Liu, Z.; Han, Z.; Ge, Q. MicroRNA Detection Specificity: Recent Advances and Future Perspective. *Anal. Chem.* **2019**, *91* (5), 3179−3186.
(34) Simino, L. A. P.; Panzarin, C.; Fontana, M. F.; de Fante, T.; Geraldo, M. V.; Ignácio-Souza, L. M.; Milanski, M.; Torsoni, M. A.; Ross, M. G.; Desai, M.; Torsoni, A. S. MicroRNA Let-7 targets AMPK and impairs hepatic lipid metabolism in offspring of maternal obese pregnancies. *Sci. Rep.* **2021**, *11* (1), 8980.
(35) Perron, M. P.; Boissonneault, V.; Gobeil, L.-A.; Ouellet, D. L.; Provost, P. Regulatory RNAs. In *Target Discovery and Validation Reviews and Protocols: Vol. 2: Emerging Molecular Targets and Treatment Options*; Sioud, M., Ed.; Humana Press: Totowa, NJ, 2007; pp 311−326.
(36) Zhang, K.; Wang, K.; Zhu, X.; Xie, M. Entropy-driven reactions in living cells for assay let-7a microRNA. *Anal. Chim. Acta* **2017**, *949*, 53−58.
(37) Daneshpour, M.; Karimi, B.; Omidfar, K. Simultaneous detection of gastric cancer-involved miR-106a and let-7a through a dual-signal-marked electrochemical nanobiosensor. *Biosens. Bioelectron.* **2018**, *109*, 197−205.
(38) Kish, L. B.; Vajtai, R.; Granqvist, C. G. Extracting information from noise spectra of chemical sensors: single sensor electronic noses and tongues. *Sens. Actuators, B* **2000**, *71* (1−2), 55−59.
(39) Morati, N.; Contaret, T.; Gomri, S.; Fiorido, T.; Seguin, J.-L.; Bendahan, M. Noise spectroscopy data analysis-based gas identification with a single MOX sensor. *Sens. Actuators, B* **2021**, *334*, 129654.
(40) Smulko, J.; Granqvist, C.-G.; Kish, L. B. On The Statistical Analysis of Noise in Chemical Sensors and Its Application For Sensing. *The Random and Fluctuating World*, 2022; pp. 475−481
(41) Muhamed Ali, A.; Zhuang, H.; Ibrahim, A.; Rehman, O.; Huang, M.; Wu, A. A Machine Learning Approach for the Classification of Kidney Cancer Subtypes Using miRNA Genome Data. *Appl. Sci.* **2018**, *8* (12), 2422.
(42) Duy, J.; Koehler, J. W.; Honko, A. N.; Schoepp, R. J.; Wauquier, N.; Gonzalez, J.-P.; Pitt, M. L.; Mucker, E. M.; Johnson, J. C.; O'Hearn, A.; et al. Circulating microRNA profiles of Ebola virus infection. *Sci. Rep.* **2016**, *6* (1), 24496.
(43) Duffy, F. J.; Thompson, E.; Downing, K.; Suliman, S.; Mayanja-Kizza, H.; Boom, W. H.; Thiel, B.; Weiner Iii, J.; Kaufmann, S. H.; Dover, D.; et al. A serum circulating miRNA signature for short-term risk of progression to active tuberculosis among household contacts. *Front. Immunol.* **2018**, *9*, 661.
(44) Tellechea-Luzardo, J.; Martín Lázaro, H.; Moreno López, R.; Carbonell, P. Sensbio: an online server for biosensor design. *BMC Bioinf.* **2023**, *24* (1), 71.
(45) Bao, G.; Lin, M.; Sang, X.; Hou, Y.; Liu, Y.; Wu, Y. Classification of Dysphonic Voices in Parkinson's Disease with Semi-Supervised Competitive Learning Algorithm. *Biosensors* **2022**, *12* (7), 502.
(46) Amethiya, Y.; Pipariya, P.; Patel, S.; Shah, M. Comparative Analysis of Breast Cancer detection using Machine Learning and Biosensors. *Intell. Med.* **2022**, *2*, 69.
(47) Kim, M.-J. Building a cardiovascular disease prediction model for smartwatch users using machine learning: Based on the Korea national health and nutrition examination survey. *Biosensors* **2021**, *11* (7), 228.
(48) Bansal, R. C.; Chandrasekaran, K.; Ayala, K.; Smith, D. C. Frequency and explanation of false negative diagnosis of aortic dissection by aortography and transesophageal echocardiography. *J. Am. Coll. Cardiol.* **1995**, *25* (6), 1393−1401.
(49) Berkowitz, K.; Baxi, L.; Fox, H. E. False-negative syphilis screening: The Prozone phenomenon, nonimmune hydrops, and diagnosis of syphilis during pregnancy. *Am. J. Obstet. Gynecol.* **1990**, *163* (3), 975−977.
(50) Maganzini, N.; Thompson, I.; Wilson, B.; Soh, H. T. Pre-equilibrium biosensors as an approach towards rapid and continuous molecular measurements. *Nat. Commun.* **2022**, *13* (1), 7072.
(51) Liu, H.; Li, Q.; Yan, B.; Zhang, L.; Gu, Y. Bionic Electronic Nose Based on MOS Sensors Array and Machine Learning Algorithms Used for Wine Properties Detection. *Sensors* **2019**, *19* (1), 45.
(52) Murphy, E. K.; Mahara, A.; Khan, S.; Hyams, E. S.; Schned, A. R.; Pettus, J.; Halter, R. J. Comparative study of separation between ex vivo prostatic malignant and benign tissue using electrical impedance spectroscopy and electrical impedance tomography. *Physiol. Meas.* **2017**, *38* (6), 1242−1261.
(53) Leyden, M. R.; Messinger, R. J.; Schuman, C.; Sharf, T.; Remcho, V. T.; Squires, T. M.; Minot, E. D. Increasing the detection speed of an all-electronic real-time biosensor. *Lab Chip* **2012**, *12* (5), 954−959.

https://doi.org/10.1021/acssensors.3c01258
ACS Sens. 2023, 8, 4079−4090