2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782679

Histones Classification Based on EGFET Signals

Jeffrey Barahona[1], Hayley Richardson[1], Lina Acosta[1], Geet Khatri[1], Francis Miller[2,3], Spyridon Pavlidis[1], *Member, IEEE*, and Edgar Lobaton[1], *Senior Member, IEEE*

***Abstract*—Dysregulation of histones has been implicated in several medical conditions, including various cancers and neurodegenerative disorders. Histone-specific biosensors are key in detecting and quantifying them, advancing our understanding of chromatin dynamics and epigenetic regulation for potential breakthroughs in cancer research and personalized medicine. The focus of this paper is on quantifying a biosensor's ability to distinguish between Human Histones (H4) and non-target analytes. Classification methods are used to provide complementary analysis to biosensor data derived from sensor manufactured using a KU7 RNA aptamer bonded to a gold electrode. The features found provide high classification performance (F1 score over 0.99) and suggest physical insights to the operation of the sensor not provided by typical analysis. Furthermore, machine learning techniques are used in an exploratory analysis to test the effects of faulty manufacturing or differences in testing environments on histone detection accuracy.**

I. INTRODUCTION

Histones play a crucial role in the field of medicine, particularly in understanding the intricacies of gene regulation and chromatin structure. These proteins, around which DNA is tightly wound, contribute significantly to the packaging and organization of genetic material within the cell nucleus. The modifications and interactions of histones are pivotal in regulating gene expression, influencing various cellular processes such as cell differentiation, proliferation, and apoptosis [1]. Dysregulation of histone modifications has been implicated in several diseases, including various cancers and neurodegenerative disorders [2]. Researchers are exploring the therapeutic potential of targeting histones to modulate gene expression patterns and develop novel treatments for conditions where epigenetic regulation is perturbed [3]. Understanding the medical relevance of histones provides valuable insights into the molecular mechanisms underlying diseases and opens avenues for the development of therapeutics.

Biosensors dedicated to histones play a pivotal role in contemporary scientific and technological endeavors, offering a specialized means to detect and quantify these biological molecules[4]. Histone-specific biosensors integrate elements designed for selectively interacting with histones, such as aptamers or proteins, with transducers that translate the biochemical signals into measurable outputs [5]. These biosensors prove essential in unraveling the intricacies of histone modifications, facilitating real-time, precise, and sensitive detection. Their application extends across diverse domains, particularly in biomedical research where they aid in understanding epigenetic mechanisms and unraveling the role of histones in various diseases [1]. The unique capabilities of histone biosensors contribute to advancing our comprehension of chromatin dynamics and epigenetic regulation, fostering breakthroughs in fields like cancer research, [6] and personalized medicine. As these technologies continue to evolve, histone biosensors hold the promise of transforming our ability to scrutinize and manipulate epigenetic processes for therapeutic and diagnostic purposes. Further research in this domain focuses on lowering detection limits and introducing real time detection and quantification of the biomolecule.

The focus of this paper will be on using machine learning to quantify the ability of the sensor to distinguish between the target analyte, Human Histones (H4), and non-target analytes, Calf Thymus Histone (CTH) and Bovine Serum Albumin (BSA). This work focuses on using machine learning methods to determine relationships between features of the signal and model performance, providing indicators of what features provide the most information about the target analyte. The focus here is not on maximizing model performance, rather on using relative model performances to provide insight on how to better characterize sensor specificity and provide a complementary analysis for finding features that delineate the differences between analytes and sensors, providing insight in greater detail than traditional analysis alone.

II. METHODOLOGY

A. Data Collection

The data was collected using an extended-gate field effect transistor (EGFET). The experimental setup involved a gold

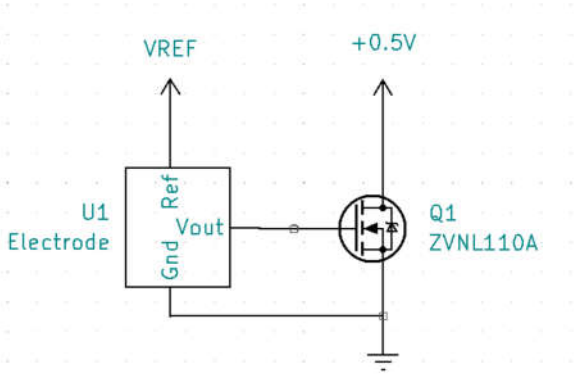

Fig. 1: Diagram of EGFET Data Collection Apparatus

*This work was supported by National Science Foundation (NSF) under awards ECCS-1936772 ECCS-2210335.

[1]Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695, USA.; Corresponding author: edgar.lobaton@ncsu.edu.

[2]Department of Medicine, Veterans Affairs Tennessee Valley Medical Center, Nashville, USA.

[3]Department of Medicine, Vanderbilt University Medical Center, Nashville, USA

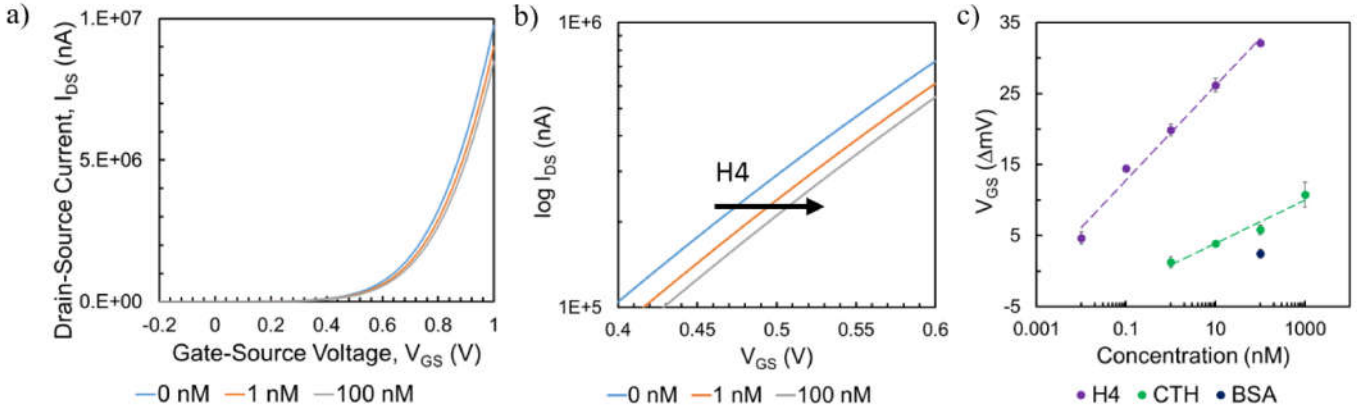

Fig. 2: Typical EGFET Analysis Pipeline. a) I-V transfer characteristics of EGFET sensor from -0.2 V to 1.0 V of 0 nM (HBS buffer), 1 nM, and 100 nM H4. b) Focused window of semi-logarithmic I_{DS} demonstrating shift in V_{GS} due to H4 binding. c) Differential V_{GS} for H4, CTH, and BSA over a range of concentrations. Error bars represent four I-V sweeps.

planar electrode with a KU7 aptamer bonded [7] to the surface, specifying it to promote histone adsorption onto the surface. This sensor configuration has been previously used and is fully described in [8]. Presence of histones was measured by using a reference electrode to induce a potential on the test electrode. The electrode was connected to a ZVNL120A field effect transistor (FET). The drain current of the FET was measured and is related to the density of aptamer-target complex sites on the electrode. A diagram of the set up is shown in Figure 1. The data was collected by testing various electrodes with different preparation methods which are summarized in Table I. Figure 2 illustrates some of the data and a typical EGFET analysis pipeline.

B. Dataset Description

The dataset is composed of experiments done using target analyte, H4, and non-target analytes, CTH and BSA, at 10pM, 100pM, 1nM, 10nM, and 100nM with at least two repetitions at each concentration. Each trial is composed of 17-57 cycles, where each cycle is defined as $c_i : \mathbb{R} \to \mathbb{R}$ mapping input voltages to output current response values. Individual cycles are used for feature extraction and training. The dataset is denoted as $C = \{c_i\}_{i \in I}$ and contains a total of 117 trials and 5121 cycles. The number of cycles per analyte is provided in Table II[1].

[1]The code and data can be found at https://github.com/ARoS-NCSU/Histones-EGFET-Classification

TABLE I: Experiment Abbreviations and Descriptions

Abbreviation	Experiment Description
PEG+KU7	Electrode is cleaned and a HBS buffer
PEG+KU7-AS	Electrode is cleaned and an artificial sweat buffer is used
PEG+KU7-Polished	Electrode is polished before binding aptamer and a HBS buffer solution is used.
PEG+KU7-24Hr	Electrode is cleaned before binding aptamer and a HBS buffer solution is used. The electrode is soaked in the buffer solution for 24 hours before testing.

C. Model Description

The main goal of our modeling efforts is to differentiate between the different analytes. The first set of analyses were done using a Random Forest model using 1000 trees and a maximum depth of 5. The model was selected by random hyperparameter optimization, selecting among multiple model types including Adaboost, Logistic Regression, Quadratic Discriminant Analysis, K-Neighbors classification, and Gaussian Processes. Using the catch22 [9], tsfel [10] and tsfresh [11] libraries, a set of over 1400 features was extracted[2]. This set is composed of spectral, wavelet, temporal, and statistical features. Spectral features include FFT, LPCC, and MFCC coefficients along with entropy measures of power spectral density. Temporal features include autocorrelation values at different lags, Langevin fixed points, subsequence similarity scores, among others. Statistical features include absolute energy, kurtosis, and other standard statistical features. Wavelet features include coefficient values of wavelet transforms using a Ricker wavelet, the number of peaks of the signal obtained after smoothing with Ricker wavelets at various wavelet widths, and others. A full list of these features can be obtained from the library documentation. These features, denoted by $f_j : \mathbb{R}^n \to \mathbb{R}$ where j is the index corresponding to the type of feature and n is the number of samples in a cycle, were extracted from the EGFET data and the time derivative of the EGFET data using second order central differences. We define the matrix $F = [f_{ij}]$ where $f_{ij} = f_j(s_i)$ and $s_i = [c_i(v(1)), c_i(v(2)), \ldots, c_i(v(n))]$ with

[2]Code will be made public if accepted

TABLE II: Overall Dataset Distribution

Analyte	Number of Cycles
CTH	974
BSA	1180
H4	2967
Total	5121

$v(k) = 0.02k - 0.2$ as the k-th value of the input voltage ramp from -0.2 to 1 V applied to the electrode. We set $n = 60$. We define the matrix $F' = [f'_{ij}]$ where $f'_{ij} = f_j(s'_i)$ and s'_i is the numerical derivative of s_i. The resulting feature matrix is defined as $G = [F \mid F']$ where G is a concatenation of the two matrices.

Colinear features were filtered out by computing absolute pairwise Pearson correlation coefficients between feature pairs on the matrix $G = [g_{ij}]$, as shown below:

$$\|cor(g_{ij}, g_{ik})\| = \left| \frac{\sum_{i \in I}(g_{ij} - \bar{g_j})(g_{ik} - \bar{g_k})}{\sqrt{\sum_{i \in I}(g_{ij} - \bar{g_j})^2 \sum_{i \in I}(g_i k - \bar{g_k})^2}} \right|.$$

A standard correlation threshold of 0.95 was used for filtering by column on the correlation matrix. Features that had correlation coefficients of higher than 0.95 were dropped from the feature matrix G. This was done to reduce overfitting. No significant differences in performance were observed when using other correlation computation methods.

Remaining features were filtered by using a variance threshold of 0.01 to remove nearly constant features. The remaining features were then standardized by the following rule:

$$h_{ij} = \frac{g_{ij} - \tilde{g_j}}{q_{75}(\{g_{ij}\}_i) - q_{25}(\{g_{ij}\}_i)}$$

where $\tilde{g_j}$ is the median of the features $\{g_{ij} | i \in I\}$ and q_m is the value of the m-th quantile of the features. This method assumes no distribution on the features and helps mitigates the effects of outliers. The resulting features set contained 384 features with a ratio of 3:2:2 of spectral, temporal, and statistical features, respectively.

A thousand repetitions of 5-Fold Cross Validation were used for training and validation for each of the experiments where the data in each fold was sampled without replacement from the dataset. No additional hyperparameter selection was performed. Furthermore, the data used for training and validation were stratified by the experiment set, concentration of analyte, and trial number. Effectively, no data from any single trial performed at a given concentration could be shared between training and validation folds. For each training fold, a separate Random Forest model was trained and a feature importance threshold set at the median importance value derived from the model was used to further prune features. The pruned features and their importance score were recorded. The cross validation metrics were obtained by training and validating Random Forest models using only the pruned features.

D. Magnitude Controlled Experiments

For these experiments, the model is trained on data where each cycle is independently scaled to have values between 0 and 1 before feature extraction and filtering. Effectively, we are eliminating the magnitude differences between different analytes and purely looking at transient and morphological differences between the signal. The goal is to discern potential features of the signal that can be used for specificity comparison outside of the traditional methods.

E. Sampling Frequency Controlled Experiments

For these experiments, we downsample the signals from the data collection, yielding applied voltage intervals of 40mV instead of 20mV. The goal of this is to change the amount of transient information available in the model, providing insight on the role of (relatively) higher frequency features on the model. We limit the analysis to a downsampling factor of 2 to keep the statistical features extracted from this data accurate.

F. Cross Surface Analysis

The purpose of these sensors is to detect the presence of histones. Traditionally, the specificity of these sensors is determined through the magnitude of the responses of the target analyte relative to other non-target analytes. Our non-target analytes are BSA and CTH, which are treated in the analysis as a single category. We hypothesize that morphological and transient features of the EGFET response can also be used to discriminate between target and non-target analytes, providing additional tools to characterize sensor performance. We consider two variants of our analysis:

1) One to Many: For these experiments, we train the model on the data from a single electrode type and gather its performance metrics on other surface types. The goal is to provide a heuristic of how similar these surfaces are. In other words, is the difference in manufacturing and testing environments between these surfaces meaningful in terms of histone response? Large drops in performance would indicate that the surfaces yield significantly different responses to the target analyte whereas similar model performance would indicate an insignificant difference. In terms of modeling, we are examining the generalizability of models trained on data using the manufacturing and testing process specified, i.e. robustness of data derived from this sensor.

2) Many to One: For these experiments, we train the model on the data from multiple surfaces and test it on surface data left out of the training dataset. This is used to determine surface outliers, i.e., surfaces that are vastly different from all other surface data, providing insight on anomalies caused by manufacturing processes.

Using data from multiple surfaces causes a class imbalance during training which was mitigated using SMOTE[12].

III. RESULTS

The main metrics used to measure performance of these models are sensitivity, specificity, and F1 score where F1 score is defined as the geometric mean of precision and recall. Sensitivity and specificity are used to provide additional context for model performance. We define sensitivity as

$$\frac{TP}{TP + FN},$$

and we define specificity as

$$\frac{TN}{TN + FP}$$

where TP, FN, TN, and FP refer to true positive, false negative, true negative, and false positive predictions, respectively.

TABLE III: Cross-Surface Baseline Case

	One-To-Many			Many-To-One		
	Sens.	Spec.	F1	Sens.	Spec.	F1
PEG+KU7-AS	0.78	0.91	0.69	0.98	0.96	0.97
PEG+KU7	0.94	0.96	0.92	0.80	0.96	0.80
PEG+KU7-Polished	0.50	1.0	0.20	0.73	0.96	0.74
PEG+KU7-24Hr	0.63	0.96	0.48	0.98	0.96	0.98
Average	0.71	0.96	0.57	0.87	0.96	0.87

A. Single Surface

These experiments serve as a check for sanity for our approach. We train models for each individual surface and test it to verify that it can identify the target and non-target analytes for that same surface. This validates our choice of features as descriptive enough to distinguish the responses of the different analytes. For each of the surfaces and buffers, using 1000×5-fold cross-validation, we find that we have average F1 scores > 0.99 for each of the cases examined: baseline, magnitude controlled, and down sampled input data.

For our baseline case (i.e., without magnitude normalization or down sampling), the top features selected by the model were the variance of the difference between the 60th and 10th quantile of the binned signal, mean FFT coefficients at 0.5Hz, and the absolute energy of the signal gradient. When controlling for magnitude, the most relevant features were energy ratios by signal segment, and the standard errors of a fitted linear model measuring the change in spread of the signal.

B. Cross Surface Comparisons

Tables III, IV and V show the results for the one-to-many and many-to-one variants considered. Overall, we see that performance in these cases is much lower than the ideal single surface scenario. Values of specificity remain relatively high; however, sensitivity drops significantly leading to F1 scores ranging from 0.43 to 0.87. The one-to-many variant yield the lowest performance (F1 score of 0.57 for the baseline case) as expected since we used the least amount of data for training. Meanwhile, the many-to-one variant did much better (F1 score 0.87 for the baseline case). This variant is more representative of what we would expect for testing of generalization on a new surface. Out of the four surfaces, the PEG+KU7-Polished produced the lowest performance on the one-to-many and many-to-one variants which indicates that this surface has some different properties.

As expected, the magnitude controlled cases (when the response were normalized) led to a decrease on detection performance. That was a nearly 30% decrease for the F1 score of the many-to-one variant. On the other hand, down-sampling only lead to a 1% decrease for the F1 score of the many-to-one variant. This indicates that information about the response can be captured appropriately at lower frequencies, while the magnitudes are more essential at differentiating the analytes.

IV. DISCUSSION

For the single surface experiments, the baseline case makes use of variance of quantile changes in the signal gradient, FFT coefficients at the high frequency end of what can be measured at the sampling frequency, and the energy carried in the signal gradient to make classifications. In our magnitude controlled case, energy density ratios of the first two segments of the EGFET and the spread of binned segments of the signal provide most of the information needed to discriminate between H4 and the control analytes. We see that, despite removing magnitude entirely from the signal, we still achieve F1 scores above 0.99, indicating that the most valuable features for classification do not depend on differences in magnitude between the analytes.

According to Pullano et. al. [5], an EGFET's drain current can be described, in the linear region, by

$$I_{DS} = \mu C_{ox} \frac{W}{L} \left[\left(V_{ref} - V_{th}^{*} \right) V_{DS} - \frac{1}{2} V_{DS}^{2} \right]$$

where the threshold voltage V_{th}^{*}, dependent on analyte behavior, can be described by

$$V_{th}^{*} = V_{th} + E_{ref} + \chi_{sol} - \frac{W_m}{q} - \phi$$

where E_{ref} is the reference electrode potential, χ_{sol} is the superficial dipole potential of electrolytes in the buffer solution, W_m is the work function of the reference electrode, ϕ is the is the potential of the surface at the sensing membrane interface, and q is the charge. Given that differences in the I_{DS} signal's morphological characteristics – and not only aggregate signal differences – are indicative of different analytes implies that V_{th}^{*} is not constant and changes with V_{ref}.

Across the board, we're able to differentiate between H4 and the non-target analytes fairly easily. However, to better determine the robustness of the sensor, another indicator of the generalizability and robustness of the data generated by

TABLE IV: Cross-Surface Magnitude Controlled Case

	One-To-Many			Many-To-One		
	Sens.	Spec.	F1	Sens.	Spec.	F1
PEG+KU7-AS	0.59	0.93	0.42	0.83	0.83	0.81
PEG+KU7	0.76	0.96	0.64	0.50	0.95	0.40
PEG+KU7-Polished	0.50	1.0	0.19	0.44	0.84	0.36
PEG+KU7-24Hr	0.58	0.81	0.46	0.87	0.75	0.87
Average	0.61	0.93	0.43	0.66	0.84	0.61

TABLE V: Cross-Surface 40mV Resampling Case

	One-To-Many			Many-To-One		
	Sens.	Spec.	F1	Sens.	Spec.	F1
PEG+KU7-AS	0.76	0.91	0.66	0.90	0.85	0.88
PEG+KU7	0.92	1.0	0.87	0.80	0.92	0.80
PEG+KU7-Polished	0.50	1.0	0.20	0.76	0.93	0.77
PEG+KU7-24Hr	0.62	0.95	0.47	0.98	0.96	0.98
Average	0.70	0.97	0.55	0.86	0.92	0.86

sensor with respect to manufacturing and usage is needed. Here we take advantage of the fact that models generated with data from one distribution typically performs worse on data from a different distribution. The cases where a simpler media, faulty manufacturing, and sensor degradation are examined, represented by the PEG+KU7-AS, PEG+KU7-Polished, and PEG+KU7-24Hr experiments, respectively. Data from our best case, normal manufacturing and testing, can still reliably be used to train a model with good performance on the faulty sensors and the case with a different, although simpler buffer medium. On the other hand, we see that using the data from the other sensors has a significantly worse performance in both the one-to-many and many-to-one tests.

V. CONCLUSION

Results indicate that data from normal manufacturing and testing can reliably train a model with good performance on faulty sensors and different buffer mediums. Conversely, using data from other sensors shows significantly worse performance in various testing scenarios. It is also demonstrated that machine learning methods lends itself to determining the specificity of sensors when standard methods are inconclusive. Future works will extend this to more complex buffer media and a wider variety of sensor conditions. Furthermore, we aim to use the physical insights gathered to simulate the sensor and use machine learning methods to suggest manufacturing steps.

REFERENCES

[1] S. Martire and L. A. Banaszynski, "The roles of histone variants in fine-tuning chromatin organization and function," *Nature Reviews Molecular Cell Biology*, vol. 21, no. 9, pp. 522–541, Sep. 2020, Number: 9 Publisher: Nature Publishing Group.

[2] Y. Wang, X. Miao, Y. Liu, *et al.*, "Dysregulation of histone acetyltransferases and deacetylases in cardiovascular diseases," *Oxidative Medicine and Cellular Longevity*, vol. 2014, p. 641 979, 2014.

[3] P. A. Lazo, "Targeting histone epigenetic modifications and DNA damage responses in synthetic lethality strategies in cancer?" *Cancers*, vol. 14, no. 16, p. 4050, Aug. 22, 2022.

[4] Y. Chen, Y. Zhou, and H. Yin, "Recent advances in biosensor for histone acetyltransferase detection," *Biosensors and Bioelectronics*, vol. 175, p. 112 880, Mar. 1, 2021.

[5] S. A. Pullano, C. D. Critello, I. Mahbub, *et al.*, "EGFET-based sensors for bioanalytical applications: A review," *Sensors*, vol. 18, no. 11, p. 4042, Nov. 2018, Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[6] B. Liang, Y. Wang, J. Xu, Y. Shao, and D. Xing, "Unlocking the potential of targeting histone-modifying enzymes for treating IBD and CRC," *Clinical Epigenetics*, vol. 15, no. 1, p. 146, Sep. 11, 2023.

[7] K. T. Urak, G. N. Blanco, S. Shubham, *et al.*, "RNA inhibitors of nuclear proteins responsible for multiple organ dysfunction syndrome," *Nature Communications*, vol. 10, no. 1, p. 116, Jan. 10, 2019, Number: 1 Publisher: Nature Publishing Group.

[8] H. Richardson, J. Barahona, G. Carter, F. Miller, E. Lobaton, and S. Pavlidis, "CHARACTERIZATION OF APTAMER FUNCTIONALIZED GOLD ELECTRODES FOR HISTONE DETECTION," in *2022 Solid-State, Actuators, and Microsystems Workshop Technical Digest*, Hilton Head, South Carolina, USA: Transducer Research Foundation, Jun. 5, 2022, pp. 292–295.

[9] *DynamicsAndNeuralSystems/pycatch22*, original-date: 2022-07-05T03:58:34Z, Dec. 21, 2023.

[10] "Welcome to TSFEL documentation! — TSFEL 0.1.6 documentation." (), [Online]. Available: https://tsfel.readthedocs.io/en/latest/ (visited on 01/11/2024).

[11] "Authors — tsfresh 0.20.1.post0.dev7+ge2dfc6f documentation." (), [Online]. Available: https://tsfresh.readthedocs.io/en/latest/authors.html (visited on 01/11/2024).

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority oversampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 1, 2002.