APPLICATION OF DEEP NEURAL NETWORKS FOR THE ANALYSIS OF COMPLEX EXCITATION-EMISSION SENSOR DATA

PROJECT DESCRIPTION

The identification and treatment of infectious diseases is crucial to public health by preventing outbreaks and limiting the spread of such sicknesses. Timely and accurate diagnosis of these diseases can be especially difficult in resource-limited environments due to the necessity of expensive equipment, skilled people, and time required. To overcome these limitations, several molecular sensors have been developed to make the identification of bacterial pathogens faster and simpler. Many of these sensors, however, rely on the use of antibodies, which can be expensive to produce and can have limited shelf life. We have previously had success developing a bacterial sensor by using multi-dimensional fluorescence spectroscopy of environment-sensitive organic fluorescent dyes and machine learning to classify bacteria by analyzing fluorescence spectra[1]. These dyes are significantly cheaper to produce and are stable in storage for multiple months, therefore using them as sensors is more accessible for low resource environments.

Computational tools used to analyze data patterns and process sensor data evolved significantly over the past two decades. While it was sufficient to use simple classification and/or dimension reduction techniques such as principal component analysis (PCA) or linear discriminant analysis (LDA) to process the readouts of simpler sensors and sensor arrays,[2] the increasing complexity of the experimental data led to adoption of more complicated processing methods. Various algorithms such as support vector machines (SVM)[3] and k-nearest neighbors[4] became widely used, with neural networks quickly growing in popularity in the past 5-7 years,[5] eventually becoming one of the most widely used tools. In analysis of multidimensional data such as excitation-emission spectra, parallel factor analysis (PARAFAC, essentially a realization of PCA applied to higher order arrays) has long been an industry standard.[6] While deep neural networks are powerful tools widely adopted for image analysis and object recognition, to this day they have seen very little application in analysis of multidimensional spectral data.[7]

The environment-sensitive dyes used in our sensors interact with the cell envelopes of bacterial pathogens. These interactions lead to slightly different spectral responses from those sensors, which can then be analyzed using machine learning to identify which bacterial pathogens are present[1]. We hypothesize that when using multiple dyes simultaneously, the interaction between the dyes can provide even more information without creating spectral overlap due to the nature of excitation-emission spectroscopy. This project seeks to examine how the synergistic relationship of a mixture of dyes impacts deep learning classification when compared to the analysis of dye fluorescence patterns individually.

METHODOLOGY

Our lab has generated a set of data on spectral responses of two environment-sensitive dyes, 2-(4'-*N,N*-dimethylaminophenyl)-3-hydroxychromone (DMAF) and Nile Red, upon their interaction with four representative pathogenic bacteria: Gram-positive *Staphylococcus aureus* and *Bacillus subtilis* and Gram-negative *Escherichia coli* and *Klebsiella pneumoniae.* This data will be organized into three subsets: all DMAF spectra, all Nile Red spectra, and all DMAF/Nile Red mixture spectra.

Model training. The ResNet-18 model, an open-source image classification model,[8] will be used in this project because of its smaller size. It is expected that a less "deep" model will be less prone to overfitting when trained on limited data, and thus provide higher classification accuracy. The model's architecture will be left intact, but the input and output structure will be modified to adapt to the format of our spectral data dimensions and the number of classes. This means that we will not be able to use transfer learning, leveraging the existing weights and biases. Three models will be trained, one for each data subset (individual dyes and the dye mixture). We will train our models using the "adam" solver with a batch size of 11, and 100 epochs max, but training will be automatically stopped if the validation loss falls below 0.02. The data will be randomly split into training (80%)

and validation (20%) subsets. After training has ended, classification accuracy will be used as the main performance criterion. The confusion matrices will be used to analyze possible misclassification patterns.

Model testing and performance comparison. Our lab has collected and organized a separate data set specifically for testing, so the models will never be exposed to this data during training. After the models are successfully trained and saved, they will be tested against this independent data set. The testing accuracy will be compared to the classification accuracy obtained during training, and conclusions regarding the overall performance of the sensor, and the possible model overfitting, will be made.

EXPECTED OUTCOMES AND ALTERNATIVE STRATEGIES

It is expected that the models trained on the DMAF and Nile Red data will perform similarly due to the internal consistency of the spectral responses of the dyes. We are also expecting that the model trained on the dye mixture data will perform better than both single-dye models due to the hypothesized synergistic effect of both dyes in the mixture.

Exceptional comparative performance of the dye-mixture model may bring about further experimentation into dye-mixture relationships. Mixtures of other dyes or the addition of a third dye as an internal standard are possibilities for future experiments.

Our lab may pursue alternative strategies if the model comparison is unsatisfactory, such as altering training parameters or testing other model architectures. Certain parameters, such as the mini batch size or choice of solver, could be pursued to optimize training, though the extent of this path is rather limited. Adjacent ResNet models or larger structures may also be tested. Viable options include ResNet-50 and ResNet-100, which are similar in size, or a larger structure such as NasNet-Large, although a model of this size is more susceptible to overfitting when trained in smaller data sets, as was mentioned previously.

PROJECT TIMELINE

Timeline	April	May	June	July	August
Experiment design					
Single dye model training					
Dye mixture model training					
Comparison of models and consideration of future steps					

Higher focus Lower focus

STUDENT/FACULTY MENTOR ROLES

As the student, my role will be to assist in the training and testing of the models. I will use my skills in code development, neural network training/testing, and experiment design, and work on expanding my knowledge about deep learning and fluorescence spectroscopy. The mentor, Dr. Denis Svechkarev, will facilitate learning during this opportunity, aid in the training and testing of models, and provide help and support in learning new skills and techniques.

PREVIOUS FUNDING

No previous funding has been awarded.

BUDGET AND BUDGET JUSTIFICATION

Budget Item	Cost ($)
Student stipend (160 hrs @ $12.50/hr)	2,000
Commercial dyes for reference data acquisition	400
Solvents for spectroscopy	100
Total of Request	**2,500**

The student will be supported by stipend for their time and effort as they carry out the experiments described in this proposal. The commercial dyes and solvents for fluorescence spectroscopy will be used to acquire reference data to study the spectral behavior of the dyes in true solutions in the absence of bacteria. This will help evaluating the possible synergistic relationship of the dyes in their mixture, and how it may impact deep learning classification. Other consumables that will be used in the project (such as cuvettes) will be provided using other costs that cover lab operational expenses.

REFERENCES

(1) Henry, J.; Endres, J. L.; Sadykov, M. R.; Bayles, K. W.; Svechkarev, D. Fast and Accurate Identification of Pathogenic Bacteria Using Excitation–Emission Spectroscopy and Machine Learning. *Sens. Diagn.* **2024**, *3* (8), 1253–1262. https://doi.org/10.1039/D4SD00070F.

(2) Skrobot, V. L.; Castro, E. V. R.; Pereira, R. C. C.; Pasa, V. M. D.; Fortes, I. C. P. Use of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in Gas Chromatographic (GC) Data in the Investigation of Gasoline Adulteration. *Energy Fuels* **2007**, *21* (6), 3394–3400. https://doi.org/10.1021/ef0701337.

(3) Li, H.; Liang, Y.; Xu, Q. Support Vector Machines and Its Applications in Chemistry. *Chemom. Intell. Lab. Syst.* **2009**, *95* (2), 188–198. https://doi.org/10.1016/j.chemolab.2008.10.007.

(4) Schroeder, V.; Evans, E. D.; Wu, Y.-C. M.; Voll, C.-C. A.; McDonald, B. R.; Savagatrup, S.; Swager, T. M. Chemiresistive Sensor Array and Machine Learning Classification of Food. *ACS Sens.* **2019**, *4* (8), 2101–2108. https://doi.org/10.1021/acssensors.9b00825.

(5) Namuduri, S.; Narayanan, B. N.; Davuluru, V. S. P.; Burton, L.; Bhansali, S. Review—Deep Learning Methods for Sensor Based Predictive Maintenance and Future Perspectives for Electrochemical Sensors. *J. Electrochem. Soc.* **2020**, *167* (3), 037552. https://doi.org/10.1149/1945-7111/ab67a8.

(6) Bro, R. PARAFAC. Tutorial and Applications. *Chemom. Intell. Lab. Syst.* **1997**, *38* (2), 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4.

(7) Schroeder, E. B.; Cheatum, C. M. Classification of Samples via Neural-Network Augmented Two-Dimensional Infrared Spectroscopy. *J. Phys. Chem. B* **2025**, *129* (19), 4738–4746. https://doi.org/10.1021/acs.jpcb.4c08573.

(8) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Las Vegas, NV, USA, 2016; pp 770–778. https://doi.org/10.1109/CVPR.2016.90.

Omaha, 30 December 2025

FUSE Selection Committee

Dear colleagues,

It is my great pleasure to support Cody Schappert's FUSE application titled "Application of deep neural networks for the analysis of complex excitation-emission sensor data" and to write this letter of support on his behalf.

Cody is a junior majoring in chemistry and computer science. As a student in my General Chemistry II class in fall 2024, he quickly attracted my attention by his high performance, insightful questions, broad scientific interests leading to him choosing to add a second major to his portfolio, and ambitious plans to pursue a graduate degree in the future. Cody joined my lab in early fall 2024, where he continues to actively contribute to our research. Physical organic chemistry, fluorescence spectroscopy, chemical sensors, and computational chemistry are among my main areas of expertise, which allows me to serve as a mentor for the proposed project.

Cody is a devoted and goal-oriented student, and a talented young researcher with an exceptional ability to think outside the box and generate original ideas and creative solutions. His unique ability to see a big picture and find trends in data made him a perfect fit for the project that investigates the impact of the amount and quality of sensor data on its performance. Working on this project, Cody will keep perfecting his experiment design, data analysis and interpretation skills.

The proposed project stems from my work on multidimensional spectroscopy applied to identification of pathogenic bacteria that we perform in our lab at UNO for the past 3 years. Development of fast, reliable, cheap, and stable sensors for infectious disease diagnostics is of high urgence and importance. This project will give Cody a unique opportunity to work in a multidisciplinary team, and apply his design and analysis skills at the intersection of chemistry and computer science. I am confident that he will be able to complete the proposed work at the highest level of quality and rigor, and I am committed to provide all necessary help and support throughout the duration of the project. I am excited to see Cody working on this project. I fully support Cody Schappert's FUSE application. Please do not hesitate to contact me if you have any questions.

Sincerely,

Denis Svechkarev, Ph.D.
Assistant Professor,
Department of Chemistry

University of Nebraska Omaha
Department of Chemistry | College of Arts and Sciences
6001 Dodge Street, DSC 350 | Omaha, NE 68182 | 402.554.6126 | dsvechkarev@unomaha.edu
unomaha.edu