Emerging Topics in the Next Generation 5G/6G Networks: A practical Overview

Wendel Balbuena
Advisor: Dr. Hisham Kholidy

SUNY POLYTECHNIC INSTITUTE

I. Overview:

5G is the next-generation mobile network, and it is being developed with the purpose of meeting the network requirements of today, and the future. 5G has the goal to make network connectivity more accessible, faster, energy efficient, and reliable, all to pave the way for Enhanced Mobile Broadband, Ultra-Reliable Low Latency, and Massive Machine-Type Communications. In the paper we discuss some of the emerging topics in 5G, presenting an introductory overview of them, their benefits, the challenges these present, and their proposed solutions gathered from recent academic sources.

II. Device-to-Device Communication and Networking in 5G

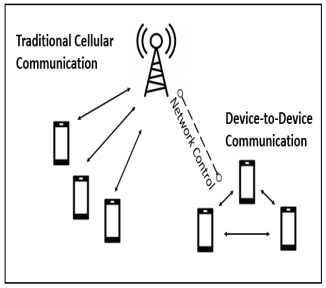
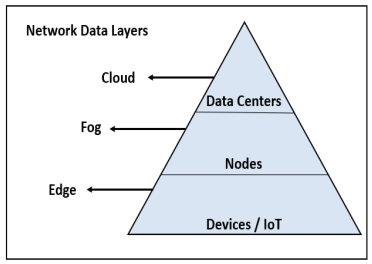


Fig. 1: *Conceptual map that illustrates and contrasts D2D from a traditional cellular network approach.*

Device-to-Device Communication allows for high-speed, more capable, and more power-efficient data transmission by bypassing base stations and traditional signal towers, only using these signal towers for network control. This is achieved by letting devices that are close to each other communicate directly instead of using a base station.

III. Mobile Edge Computing (MEC) in 5G



Multi-Access Edge Computing (MEC) [1] is the implementation of specialized equipment that processes data at the network edge instead of sending it to the cloud [2,3].

Fig. 2: *Graphical concept of Network Edge demonstrating a high-level overview of where equipment exists in the network.*

IV. Multicast/Broadcast for 5G and Beyond

5G Multicast Broadcast Services (MBS) will leverage the use of existing independent multicast broadcast technologies to gain increased area coverage, mobility support, and spectral and resource efficiency.

MBS use cases:

- IPTV
- Vehicle-to-Everything (V2X) services
- IoT communication
- Public safety announcements
- Software updates

V. Multi-Radio Access Networks

This technology leverages the use of multiple access points of available Radio Access Technologies (RAT) simultaneously on the same user equipment. It uses the ATSSS (Access Traffic Steering, Switching, and Splitting) architecture to allow these multiple connections to happen.

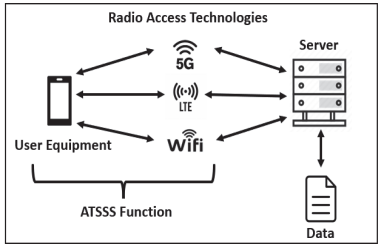


Fig. 3*: Conceptual map showing how multi-radio access networks work by using the ATSSS function, improving data rate, and reliability, and decreasing latency.*

VI. Latency in Convergence of RAN and 5G Core.

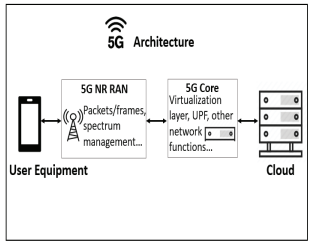


Fig. 4*: High-level concept map of the 5G architecture that illustrates the relation between RAN and 5G Core.*

One of 5G's main requirements is the need for reliable low latency, with the convergence of the core network and the Radio Access Network (RAN) being one of the main contributors to latency.

VII. Practical Experiments: Kubernetes Hardening

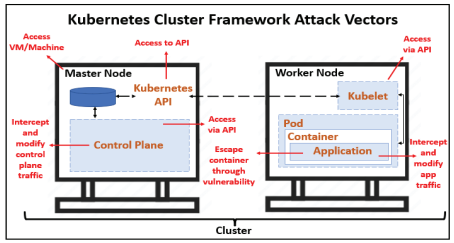


Fig. 5: *High-level concept map of the Kubernetes framework and its attack vectors.*

Tutorials to harden Kubernetes created and implemented:

1. Upgraded Kubernetes to the latest version.
2. Ran containers with a non-root user.
3. Performed image scanning
4. Implemented secure authentication to the Kubernetes API server and enabled RBAC.
5. Set up network policies.
6. Enabled audit logging.

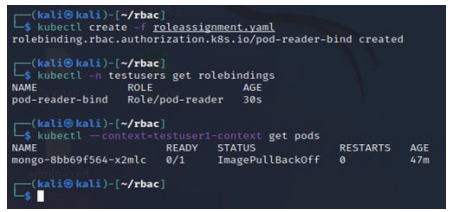


Fig. 6: *Screenshot taken during the implementation of authentication and RBAC.*

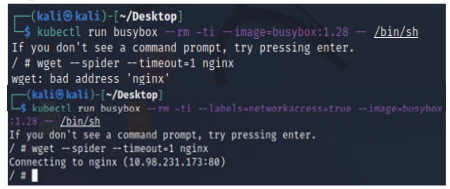


Fig. 7: *Screenshot showing whitelist network policy working.*

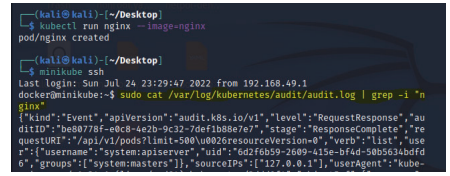


Fig. 8: *Screenshot showing audit logs enabled and recording events.*

VIII. References

[1] N. Hassan, K.-L. A. Yau and C. Wu, "*Edge Computing in 5G: A Review*," IEEE Access, vol. 7, pp. 127276-127289, 2019.

[2] Hisham A. Kholidy, "*Multi-Layer Attack Graph Analysis in the 5G Edge Network Using a Dynamic Hexagonal Fuzzy Method*", Sensor Journal. Sensors 2022, 22, 9.

[3] Hisham A. Kholidy, *Andrew Karam, James L. Sidoran, Mohammad A. Rahman, "5G Core Security in Edge Networks: A Vulnerability Assessment Approach*", the 26th IEEE Symposium on Computers and Communications (IEEE ISCC 2021), Athens, Greece, September 5-8, 2021.

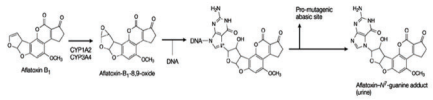
Profiling Yeast DNA Repair Mutants for Sensitivity to Bioactivated Drugs and Toxins

SUNY POLYTECHNIC INSTITUTE

Danielle Huli, Michael Fasullo Ph.D

Introduction:

Many drugs can become DNA damaging agents when bioactivated. Human cytochrome P450 enzymes (CYPs) are a group of enzymes that activate several drugs and toxins, including aflatoxin B_1 (AFB1) (1). CYP3A4 is a major P450 enzyme that's abundant in the human liver. In this study, a plasmid expressing CYP3A4 was introduced into *ccharomyces cerevisiae* (budding yeast) strains defective in DNA repair. We used selective ploidy ablation (SPA) as a method to introduce the expression vector. We also screened DNA repair mutants expressing CYP1A2 for AFB1 sensitivity. We identified the DNA repair genes SGS1and MMS2 as AFB1 resistance genes.



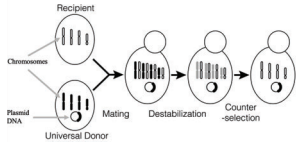
igure 1 : Bioactivation of AFB1

Methods

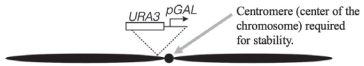
Confirming DNA damage sensitivity: Strains of yeast, each missing a single gene that helps with DNA damage repair, were tested for sensitivity to UV radiation, X-Ray radiation, Methyl Methanesulfonate (MMS), and Hydroxyurea (HU). To test sensitivity to these stimuli, yeast cultures were spotted (5 µL) onto YPD, MMS, and HU plates. YPD plates were then exposed to either UV (30 seconds) or X-Ray radiation (30 minutes), with one YPD plate left unexposed to act as a control.

Introduction of expression plasmids: The SPA method was used to quickly introduce expression plasmids into 30 strains to create haploid strains with the recipient strain DNA and a CYP3A4 plasmid.

Universal Donor strains (UDS) containing the plasmid of interest and chromosomes with a specific URA3 gene and Gal promotor were mated with the DNA repair mutant yeast of interest, creating diploid cells.



igure 2 . Selective ploidy ablation (SPA) method of introducing expression plasmids (2)



igure 3 . Chromosome of Universal Donor Strain (2)

Growing the cells in galactose media activated the Gal promotor, destabilizing the centromere, and spotting on FOA plates rendered the strains that haven't destabilized the centromere inviable.

Performing growth curves: DNA repair mutants containing CYP1A2 were tested for sensitivity to AFB1. Growth in YPD and SC-Ura was recorded over 24 hours in the presence of 0.5% DMSO, 5 µM AFB1, and 50 µM AFB1. A growth curve was constructed based on the data and the percent growth in 5 µM and 50 µM AFB1 was calculated.

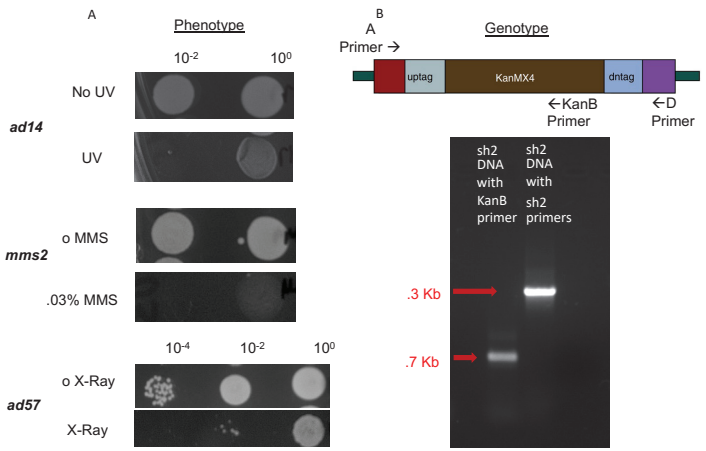
$$\text{Percent Growth} = \frac{\textit{Area Under the Curve of 5 µM AFB1 or 50 µM AFB1}}{\textit{Area Under the Curve of 0.5\% DMSO}}$$

Goals:

- Create a library of DNA repair mutants and document their phenotypes.
- Introduce the CYP3A4 plasmid into DNA repair mutants with sensitivity to at least one stimuli using the SPA method.
- Test the sensitivity of DNA repair mutants with a CYP1A2 plasmid to AFB1.

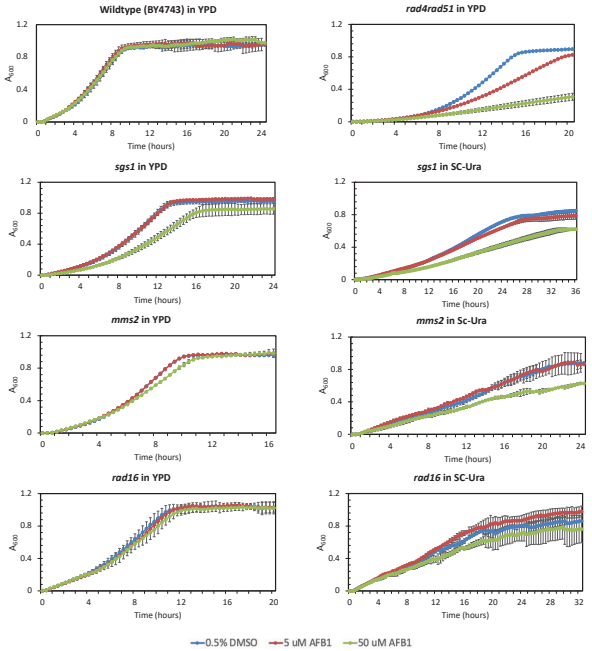
esults:

Confirmation of DNA Repair Mutants



igure 4. A. Spot test results after exposure to UV radiation (60 J/m² for 30 sec), 0.03% Methyl Methanesulfonate (MMS) and X-Ray radiation (46.5 Gy). B. PCR results of *msh2* DNA with an msh2A and KanB primer and an msh2A and msh2D primer.

Testing Sensitivities to Aflatox B 1



igure 5. Growth curve graphs of the exposure of the wildtype strain (BY4743), positive control strain (*rad4rad51*), and experimental strains *sgs1*, *mms2*, and *rad16* to 0.5% DMSO (solvent), 5 µM AFB1, and 50 µM AFB1.

Growth as Measured by Area Under the Curve

Nonselective Medium (YPD)

train	DMSO	5 µM AFB1	5 µM AFB1 % Growth	50 µM AFB1	50 µM AFB1 % Growth
BY4743	13.7	14.0	1.0	14.0	1.0
rad4rad51	8.2	6.1	0.7	2.6	0.3
*sgs1**	10.9	11.1	1.0	8.0	0.7
rad16	13.8	13.6	1.0	13.2	1.0
mag1	13.4	14.1	1.0	14.8	1.1
msh2	13.9	13.2	0.9	13.0	0.9
slx4	14.5	14.5	1.0	14.1	1.0
mms2	13.4	13.6	1.0	13.2	1.0
rad59	13.4	13.4	1.0	12.9	1.0

**sgs1* mutant sensitive to AFB1

elective Medium (SC-Ura)

train	DMSO	5 µM AFB1	5 µM AFB1 % Growth	50 µM AFB1	50 µM AFB1 % Growth
BY4743	10.7	9.4	0.9	9.0	0.8
*sgs1**	.3	4.2	1.0	2.8	0.6
rad16	8.0	8.7	1.1	6.8	0.9
mag1	10.9	10.4	1.0	9.8	0.9
msh2	9.6	9.8	1.0	9.0	0.94
slx4	10.2	9.6	0.9	9.5	0.9
*mms2**	.7	8.1	1.0	5.9	0.8
rad59	8.6	8.8	1.0	8.4	1.0

**sgs1* and *mms2* mutants sensitive to AFB1

Table 1. Area under the curve and percent growth of DNA repair mutants expressing CYP1A2 exposed to 0.5% DMSO, 5 µM AFB1, and 50 µM AFB1 in YPD and SC-Ura media.

Conclusion

- We tested the sensitivities of over 100 DNA repair mutants to UV radiation, X-Ray radiation, MMS, and HU and documented their phenotypes.
- CYP3A4 was introduced to selected DNA repair mutants using the SPA method. Future testing will confirm this.
- *sgs1* is sensitive to 50 µM aflatoxin B_1, which is of interest since SGS1 is involved in decreasing replication stress caused by DNA damage. *mms2* additionally exhibited sensitivity to 50 µM AFB1 in SC-Ura medium.
- Future directions: Test the newly constructed DNA Repair Mutants containing the CYP3A4 plasmid for sensitivities to AFB1 and other DNA damaging agents.

Acknowledgement

I would like to thank Michael Dolan for his help on this project.
Fundraising: NIH R15ES023685

eferences

1. Parke, D V. "The Cytochromes P450 and Mechanisms of Chemical Carcinogenesis." *Environmental H_ th Perspectives*, vol. 102, no. 10, 1994, pp. 852–853., https://doi.org/10.1289/ehp.94102852.

2. Reid, Robert J.D., et al. "Selective Ploidy Ablation, a High-Throughput Plasmid Transfer Protocol, Identifies New Genes Affecting Topoisomerase I–Induced DNA Damage." *Genome Research*, vol. 21, no. 3, 2010, pp. 477–486., https://doi.org/10.1101/gr.109033.110.

3. St. John, Nick, et al. "Genome Profiling for Aflatoxin B1 Resistance in *Saccharomyces Cerevisiae* Reveals a Role for the CSM2/Shu Complex in Tolerance of Aflatoxin B1-Associated DNA Damage." *G3 Genes|Genomes|G_net ics*, vol. 10, no. 11, 2020, pp. 3929–3947., https://doi.org/10.1534/g3.120.401723.

Circuit Design and Signal Generation for Accessing Neuromorphic Chip

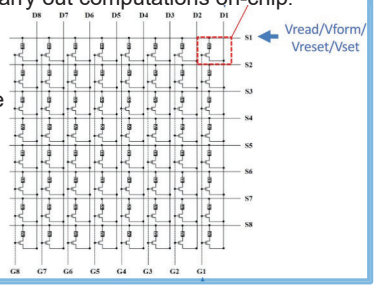
COLLEGE OF ENGINEERING
SUNY POLYTECHNIC INSTITUTE

SURP Student: Alice Lin, SURP Advisor: Yu Zhou
College of Engineering, SUNY Polytechnic Institute, Utica, NY 13502

Introduction

A neuromorphic chip is an IC chip which contains a matrix of memristors. Each memristor is a two-terminal IC-compatible analog device, which can be programmed to different conductance states. The tunable conductance states of memristors allow neuromorphic chips to be used as flash memory and to potentially carry out computations on-chip.

Our objective is to design a programmable pulse generator circuit to generate the necessary voltage pulses for accessing the neuromorphic chip.

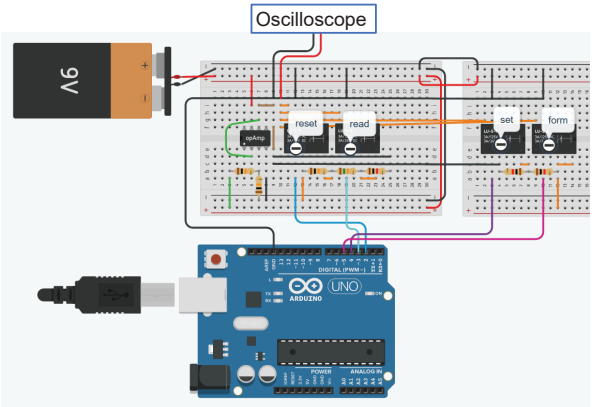


Tasks

To design and construct a circuit to generate 4 different levels of pulse signal required for the chip access

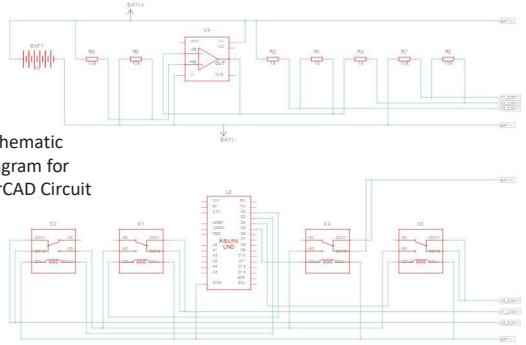
- Forming: Amplitude of at least 3.30V and pulse width is around 1ms
- Setting: Amplitude of 1.0V to 2.5V and pulse width around 1ms
- Resetting: Amplitude of -0.8V to -2V and pulse width around 50ns to 1ms
- Reading: Amplitude of -0.2V to 0.2V and pulse width around 10µs

Circuit Design (TinkerCAD)



Circuit Design

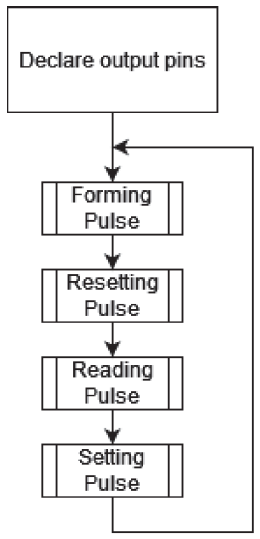
- Uses an Op Amp to convert a 9V input to -4.5V to 4.5V
- The Op Amp outputs 4.5V as the forming voltage
- Generates 2.25V setting voltage by using two 1k resistors
- Generates -2.77V resetting voltage and -0.172V reading voltage by using 10k, 15k, and 1k resistors
- SPDT relays, which act as digital switches, are used to choose among the 4 different voltage levels
- Arduino is programmed to select among different voltage levels and control the pulse width



Schematic Diagram for TinkerCAD Circuit

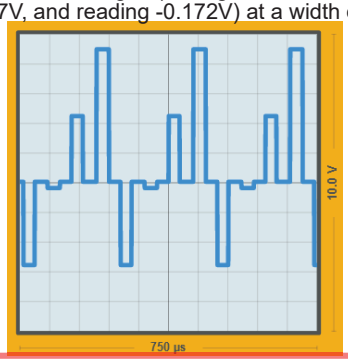
Arduino Program

```
// C++ code
// Alice Lin
void setup()
{
  pinMode(2, OUTPUT); //max 40mA
  pinMode(3, OUTPUT);
  pinMode(4, OUTPUT);
  pinMode(5, OUTPUT);
}
void loop()
{
  form(); //outputs 4.50V
  reset(); //outputs -2.77V
  read(); //outputs -0.173V
  set(); //outputs 2.25V

}

//require -0.8V to -2V
void reset(){
  digitalWrite(2, HIGH);
  delayMicroseconds(10);
  digitalWrite(2, LOW);
  delayMicroseconds(10);
}

//require -0.2V to 0.2V
//used 12k, 1k resister voltage divider
void read(){
  digitalWrite(3, HIGH);
  delayMicroseconds(10);
  digitalWrite(3, LOW);
  delayMicroseconds(10);
}

//require 1.0V to 2.5V
void set(){
  digitalWrite(4, HIGH);
  delayMicroseconds(10);
  digitalWrite(4, LOW);
  delayMicroseconds(10);
}

//require at least 3.30V
//around 1 ms
void form(){
  digitalWrite(5, HIGH);
  delayMicroseconds(10);
  digitalWrite(5, LOW);
  delayMicroseconds(10);
}
```



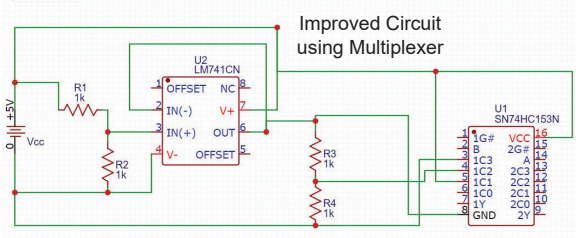
Result

As shown in TinkerCAD simulation, the oscilloscope shows the 4 pulses with different voltages (setting 2.25V, forming 4.5V, resetting -2.77V, and reading -0.172V) at a width of 10µs.



Conclusions and Discussions

- The design of the circuit has been validated by TinkerCAD simulation. It can generate the desirable pulses for accessing the neuromorphic chip.
- Further improvement: A multiplexer is a more compact choice to replace those relays for selecting different digital pulses.



Acknowledgements

- SURP Program for support the summer project
- Prof. Cady's AFRL award #FA8750-21-1-1019
- Prof. Yu Zhou for project advice
- Miss Jeelka Natwarbhai Solanki (PhD candidate), Prof. Natheniel Cady, and Mr. Steven Wood for providing information and advice on the neuromorphic chip
- Mr. William Denale for advice on circuitry

SUNY POLY

Analysis of Electroencephalography (EEG) Data

Megan G. Lorenz, Student at College of the Holy Cross
Research Mentor: Daniel K. Jones, Ph.D., P.E.
College of Engineering, SUNY Polytechnic Institute, Utica, New York 13502

Introduction:

Electroencephalography, or *EEG*, is a technique in which electrodes are used to measure voltage activity in the brain. Voltage activity is useful in determining patterns in brain responses. During an EEG, electrodes, or ***channels***, are placed on the scalp and activity is measured in **microvolts (μV).**

The analysis of EEG data comes in the form of analyzing ***event-related potential*** (ERP) graphs. These graphs measure the voltage (μV) on the y-axis and the time (*ms*) on the x-axis. While the magnitude of voltage is considered, the ***latency***, or timing, is a key factor in determining response differences across participants.

Objectives:

1) **Compare positive, negative, and neutral responses** between participants.
2) Look for **topoplot differences** in latency among positive, negative, and neutral responses.
3) Look for **ERP differences** between responses.

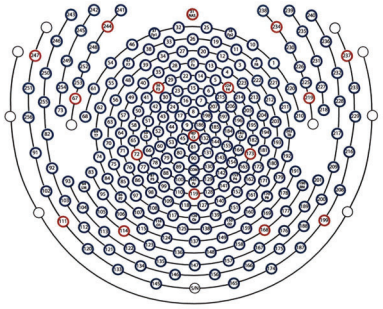


Figure 1: Sensor Net Channel Map

Methods:

Used **MATLAB/ EEGLAB** software to perform graphing and analysis.

1) Creating Topoplot Movies:

- Graphed 2D topoplots from latencies 0ms-800ms, at intervals of 10ms.
- Saved .jpg images at each 10ms interval.
- Used Clideo Video Maker to combine images into a video.
- Created side-by-side video comparisons with Fastreel.

2) Statistical Analysis of EEG Data

- Created a study in EEGLAB with all positive, negative, and neutral datasets for 45 total participants.
- Performed **One-Way ANOVA** tests for **ERP plots**.
- Performed One-Way ANOVA tests for **topoplots**.
- Analyzed results of ANOVA tests to determine significant differences between positive, negative, and neutral ERP data.

Results:

Typically, the **responses to the negative images lasted longer** than the responses to the positive images. With the **negative** images, higher voltage contrast typically occurred from **~150ms to ~550ms**, with a **total duration of 400ms**. Meanwhile, with the **positive** images, higher voltage contrast typically occurred from **~240ms to ~520ms**, with a **total duration of 280ms**. From 400ms to 800ms, the typical channel voltage difference between the positive and negative ERPs was **statistically significant** with a p-value of **0.05**.

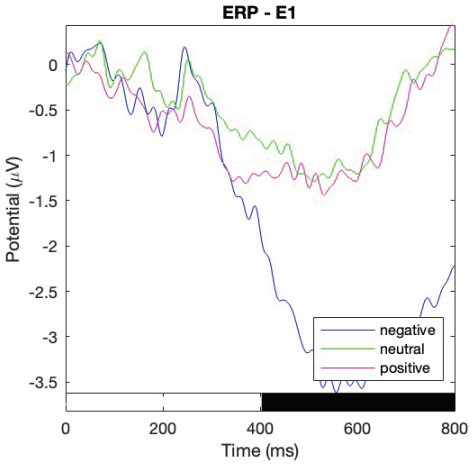


Figure 2: One-Way ANOVA of ERP for 45 Total Participants (statistically significant differences shown by black bar)

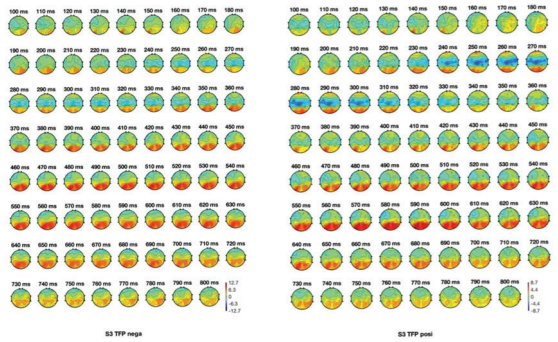


Figure 3: Topoplot Comparison of Negative vs. Positive Images (Participant 3)

Conclusions:

It is reasonable to conclude that negative images leave a stronger, more lasting psychological impression. Specifically, the second half of the response is significantly different after seeing positive versus negative images.

Future Directions:

- Analyze findings in relation to **personality types** of participants.
- Further explore the **400ms to 800ms** time interval for explanations for the significant differences found.

References:

HydroCel Geodesic Sensor Net. Electrical Geodesics, Inc. (EGI). Accessed 29 July 2022.
Luck, Steven J. *An Introduction to the Event-Related Potential Technique*. MIT Press, 2014.

Contacts:

Dr. Daniel Jones: jonesd5@sunypoly.edu
Megan Lorenz: lorenzm@sunypoly.edu

Developing Undergraduate Laboratory Modules on Crystals and Defects

SUNY Polytechnic Institute

Joshua Peck, Kathleen Dunn, Bradley Thiel

WWW.SUNYPOLY.EDU

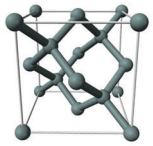
Introduction

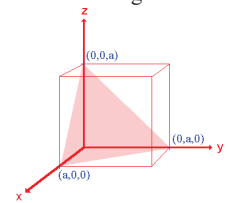
The goal of this internship was develop a series of undergraduate laboratory modules as the basis for a future materials engineering course, focused on the relationship between microstructure and macroscopic properties in technologically relevant materials.

Background

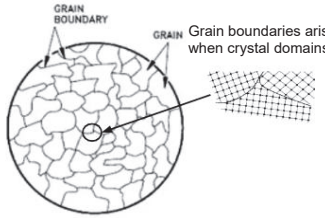
Most undergraduate courses treat crystals as perfect 3-dimensional arrays of atoms. Most materials are not perfect, however, and the defects dictate most observable properties. These modules are designed to introduce undergraduates to the language of crystallography and defects, the measurement of defects and their impact on properties, and the difficulties associated with eliminating defects.

Atoms arrange themselves in regular arrays known as crystals; this is silicon's choice:





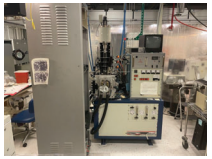
Directions and planes are defined by vector notation called "Miller Indices"



Grain boundaries arise when crystal domains meet

Cz Growth of Silicon

The Czochralski (Cz) method is the standard means for growing defect-free silicon material suitable for the semiconductor industry. This laboratory is intended to introduce students to the equipment, as well as the thermodynamic challenges to producing perfect crystals.



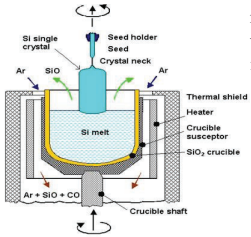
Cz tool in CESTM L233 CNSE Innovation Lab

Laboratory tasks

- Determine status and capabilities of legacy system
- Resurrect attached control system, including hard drive repair
- Work with engineering team on re-installation requirements (water, power)

Curriculum materials created to describe process

- A small seed crystal is lowered to contact a silicon melt, which crystallizes with the same orientation as the seed.
- The crucible with the melt is rotated as the seed is withdrawn, producing an ingot



Silicon ingot

Ideally silicon ingots should be pure & crystalline, but observing the defects in the wafer shows what part of seeding went wrong, offering opportunities for learning.

Future tasks:

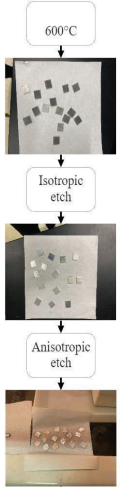
- Obtain or design new control software & missing dongle
- Acquire Si feedstock and seed crystal to test system
- Have site engineering complete re-install checklist

Etching Aluminum Polycrystals

Most defects require specialized microscopes to be seen, but this lab uses a special case of high purity aluminum where the grains can be observed by the naked eye. This lab can thus be used to teach concepts of microstructure and defects without complex equipment.

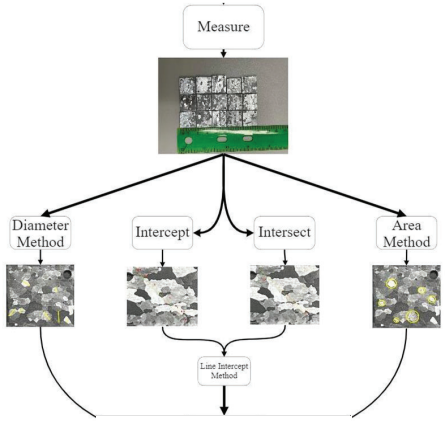
Laboratory tasks

- Anneal aluminum blanks to cause recrystallization and enable grain growth.
- Use chemical etching to remove material and reveal the microstructure of interest
  1) Isotropic etching by NaOH in water was used to remove surface oxide and expose bare Al metal
  2) Anisotropic etching in Poulton's reagent revealed grain structure: etch rate depends on orientation of the grains
- Individual pieces mounted with hardware as zipper pulls



Curriculum materials created to describe analysis and evaluation

- Different methods (diameter, intercept, area) result in slightly different values for grain size
- Discussion of statistics, confidence, and validity



Material Testing

Mechanical properties depend on the ease with which defects can be propagated through a solid. This lab is used to probe two kinds of mechanical response to stress using a commercial tester.

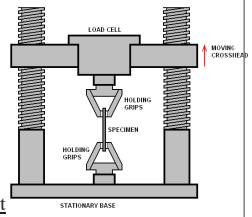
Laboratory tasks:

- Determine status and capabilities of legacy system
- Test everyday materials to calibrate system

Configuration 1: Tensile test

The substance is pulled apart until it fractures.
Tested substances:

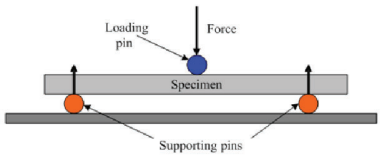
- Electrical wire
- Zip ties



Configuration 2: Flexural test

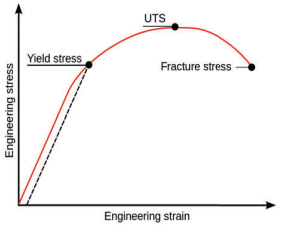
The substance has a load pushed onto it until it until it bends. The yield point is reached. Tested substances:

- Toothpicks
- Silicon substrates
-



Curriculum materials created to describe material response in reacting to the applied force, i.e., the interpretation of a stress vs. strain graph to extract:

- Elasticity(Stress resistance before the yield point).
- Ductility (Stress resistance after the yield point).
- Strength (The total amount of energy absorbed before fracture).



Future Tasks

- Purchase another load cell (existing one is damaged)
- Upgrade computer system for modern data transfer

References

Radavich, Katie. "2010 - The Use of Metallographic Etching as a Decorative Surface Treatment for Jewelry." The Santa Fe Symposium, University of Wisconsin, 2010, https://www.santafesymposium.org/2010-santa-fe-symposium-papers/2010-the-use-of-metallographic-etching-as-a-decorative-surface-treatment-for-jewelry.

Yardley, James T. "Silicon Basics --General Overview. - Columbia University." Columbia.edu, 9 Nov. 2001, https://www1.columbia.edu/sec/itc/ee/test2/pdf%20files/silicon%20basics.pdf.

Schaffer, James P. The Science and Design of Engineering Materials. McGraw-Hill, 2006.

Acknowledgements

The authors would like to thank the following for their invaluable assistance in these projects:

- Brian Taylor
- Kyle Unser
- Ross Pareis
- Eric Wales
- Prof. Ji Ung Lee

SUNY RF

SUNY POLYTECHNIC INSTITUTE

Problem Outline:

Semiconductor wafers are an essential component in modern-day electronics typically used for photovoltaics or ICs. Before these wafers can be used in such components, they must endure several phases of processing. The SiConi™ pre-clean phase is defined by etching and annealing. Etching selectively removes material on the substrate creating a uniformed surface. This removal is necessary because it eliminates any impurities on the semiconductor wafer, allowing for an increase in contact resistance, yield and reliability of middle-of-the-line interconnections.

The current study focuses on carrying out numerical simulations of the etch process to determine conditions necessary for a uniform oxide removal. A 300mm Si wafer with a target etching amount of 4-12 nm is set up under a showerhead that delivers reactants from a remote plasma and the final thickness of the oxide layer is measured.

SiCo i™ Etchi g Process:

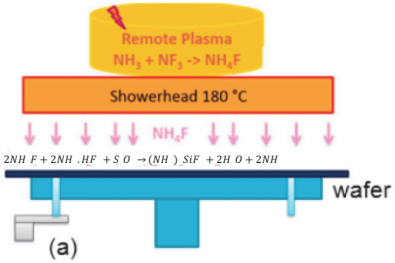


Figure 1. Etching step f the SiC ni™ Process [1]

The SiConi™ etch is one such remote plasma-as isted dry etch process that involves the simultaneous exposure of the wafer substrate to H_2,N F_3, and H_3 pla ma by-products. A a result, fluoro ilicate alt are generated through the reaction of the H4F reactive specie with the Si $_2$ film which i followed by an annealing proce where volatile sale are transformed to SiF_4 and H_3. The combined process allow for a damage-free removal of surface impurities, resulting in increa ed contact re i tance and reliability of middle-of-the-line interconnections. **While the SiConi™ is one of the more popular dry preclea processes, the specific role of various physical parameters throughout the process remains to be determined.**

Virtual Labs & Computational Fluid Dynamics:

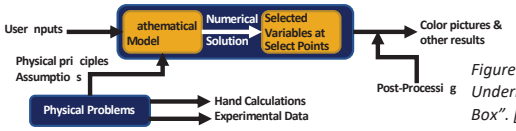


Figure 2. "What Happens Underneath the Black Box". [2]

This graphic illustrates the different components that make up a virtual lab. **Virtual labs are interactive, digital simulations of activities that typically take place in physical laboratory settings**. Virtual labs simulate the tools, equipment, tests, and procedures used in chemistry, biochemistry, physics, biology, and other disciplines.

Finite Volume Method:

The Finite Volume Method divides the domain of the fluid into multiple control volumes known as cells. The cell centers are then evaluated using a system of algebraic equations that are simplifications of the partial differential form of the Naiver-Stokes equations . Each algebraic equation will relate a cell-center value to its adjacent cells.

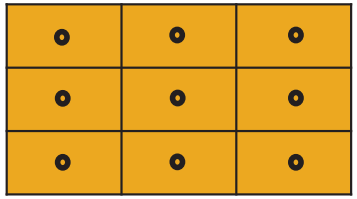


Figure 3. 3x3 cells with cell-centers

Geometry:

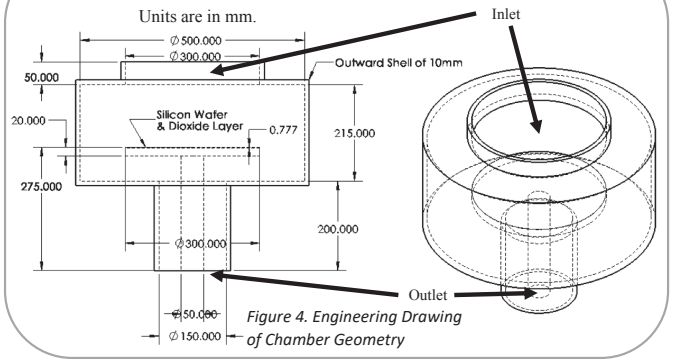


Figure 4. Engineering Drawing of Chamber Geometry

eshing:

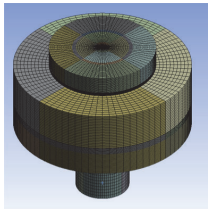
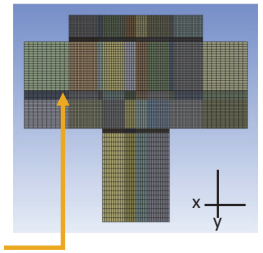


Figure 5a. sometric View of Chamber's Mesh

After the geometry is created, different parameters can be manipulated to create a uniformed mesh.

A series of edge, face, and sweep meshes were used to achieve grid independence.

Figure 5b. Cross-sectional View of Chamber's Mesh



This arrow indicates the thin wafer section, alongside a modeled nanoscopic Silicon-Dioxide layer.

Preliminary results of Simulatio :

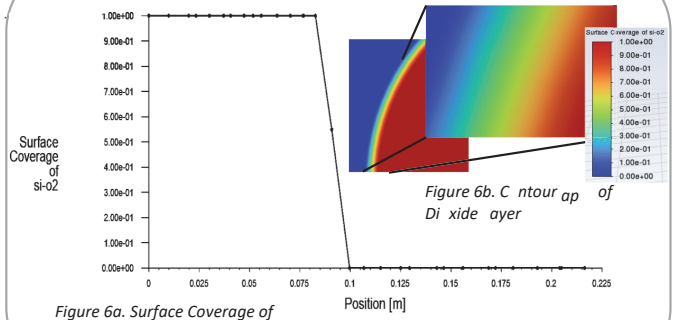


Figure 6a. Surface Coverage of Si- 2 and Position

Figure 6b. C ntour ap of Di xide ayer

Refere ces

[1] Grégoire, M., Horvat, B., Bozon, B. N., Combe, D., Dabertrand, K., & Roy, D. (2019). dditional Siconi™ pr -clean for reliable TiSix contacts in advanc d imager technologies. *icro and ano Engineering, 2*, 104-109.

[2] Bhaskaran, R. B. [Cx Simulations]. (2016, May 10). *CO E G 22016- 006000* [Video].

Faces in TIME

An interactive web resource exploring faces in Time Magazine 1923-2014

Project Overview

This project examines an archive of Time magazine containing 3,389 issues ranging from 1923 to 2014, focusing on images of human faces. We extracted faces from this archive, and tagged them by visual characteristics, including gender, race, age, expression, and context. The research examines how representations of human faces evolved, and reveals relationships between the images and their corresponding socio-political contexts. For example, we found that the percentage of female faces peak during eras when women have been active in public life, and wane in eras of backlash against women's rights. Other findings, among many, include the observation that the percentage of smiling faces increases from 1923 to 1940, which coincides with known shifts in the cultures of advertising and of photography.

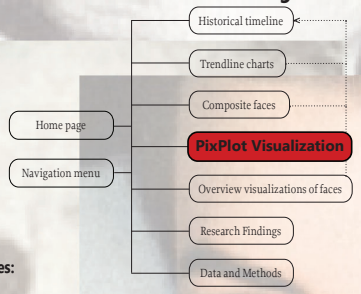
The goal is to develop a web resource that makes our data and research findings accessible to the public through visualizations and interactive narratives.

This summer, I developed one of the pages in the web resource. This page presents a raw visualization of the image data. The images are sorted (using AI) by raw visual features. The user is encouraged to explore relationships between the faces using this tool.

SUNY RF The Research Foundation for The State University of New York

SUNY POLYTECHNIC INSTITUTE

Site Architecture and Navigation



Notes:

Each page contains a navigation menu that allows the user to navigate to any of the other pages.

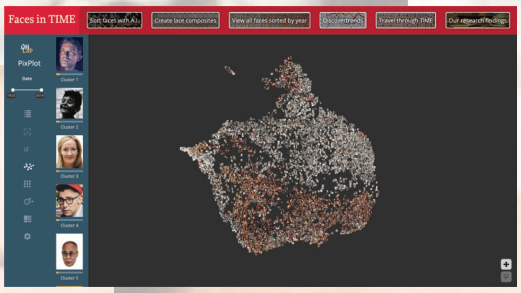
Each of the interactive visualizations includes explicit links directly inviting the user to look at the historical timeline

My Contribution to the Project

The web page I created used software called Pixplot, which was developed by Yale's Digital Humanities Lab. The program receives a collection images as the input, along with their metadata. We used a subset of 8,787 face images from Dr. Jofre's Time magazine collection. The program uses an AI algorithm to output the images in a 3D arrangement organized by visual features.

My contribution was to take the output created by the Pixplot program and redesign it to provide a more user-friendly interface to display our data. To achieve this, I am contributing to the front-end development of this resulting webpage (using HTML, CSS, and Javascript).

Outcomes



Summer Challenges

Since the code was initially programmed by Yale's Digital Team, one challenge was to get into the groove of reading other people's thousands of lines of code and trying to add my changes to the code the best I can.

Additionally there were many times when coding that I ran into problems or roadblocks and had no clues about how to solve those issues. After tons of research and the help of explaining my roadblocks to someone with a different perspective. It would help me to figure out what I was missing or had not tried yet.

Project by: Dr. Ana Jofre and Timothy Tu

Lyme Disease Biosensing & Diagnostic Machine Learning Modeling for Lyme Diseases

Rachel Utterman[1], Benjamin Tubner[2], Frank Doyle[2], Nathaniel Cady[2]
[1]Northeastern University, [2]SUNY Polytechnic Institute
Ciencia, Inc.

Abstract

Supervised machine learning (ML) can be used to support the diagnosis of Lyme disease (LD) using data captured by a grating coupled fluorescent plasmonic (GC-FP) biosensor developed by Ciencia, Inc. The biosensor project aims to create a faster, more sensitive test for Lyme disease than the CDC recommended two-tiered testing (TTT) system. Potential benefits of using supervised ML to process the biosensor data include more accurate diagnoses than our current standard, a receiver operating characteristic (ROC) threshold test, the ability to match or surpass CDC's original results from the standard TTT test, and additional capabilities such as disease stage classification. GC-FP chips were run with 82 human serum samples representing multiple patient groups, and their respective antibody detection ratios from imaging data were used to train various machine learning models. The results of these tests were compared against the results from our standard ROC-based test and the original TTT results to determine each algorithm's efficacy. Multiclass Logistic Regression, the highest performing model thus far, resulted in 95% sensitivity, and 93% specificity for this dataset, outperforming results from both the TTT test and the ROC test, and thus encouraging future usage of supervised ML as a diagnostic tool for Lyme disease.

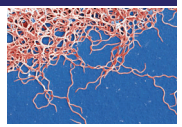
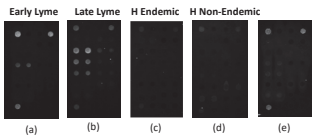


Figure 1. Lyme disease is caused by the infection of spirochetes belonging to the *Borrelia burgdorfe* sensu la o complex. heir high genomic diversity and low degree of amino acid sequence conservation result in grea an igenic diversity, making accura e and reliable LD detection particularly complica ed.

Background

Lyme disease (LD) is the most common tick-borne illness in the Northern hemisphere. The disease is treatable when caught early but can have devastating and prolonged health implications afte just a few weeks of unt eated infection. The standa d TTT, which consists of an enzyme-linked immunosorbent assay (ELISA) followed by a Western blot, takes days to weeks to return results. It also leaves roughly 40% of Lyme cases undetected, the majority of which are in the early critical treatment stage. Howeve , developing a new test for LD is often difficult due to high cross-reactivity, genetic dive sity of the genospecies of *B. burgdorfei*, and ove lapping symptoms with "look-alike" diseases. Howeve , this biosensor has demonstrated the ability to match or exceed the TTT. Previously, the data analysis employed diagnostic thresholds taken from C analysis. It was thought that a machine learning approach could replace the ROC analysis to achieve a higher diagnostic accuracy.

Figure 2. The biosensor, a gold-plated silicon chip etched with nanoscale grating, is printed with chosen target antigens (a), assembled (b), then flowed with human serum, wash buffer, and fluorescent-tagged labeling antibodies (c), and thereafter analyzed with a plasmonic imaging tool developed by our partners at Ciencia, Inc. (d). We were able to use numerical inputs (GC-FP detection ratios) to train ML models to classify the sample as either positive or negative (e) and compare these to diagnoses from our current standard approach.



$$\text{GC-FP Detection Ratio} = \frac{\bar{x}_{\text{target intensity}}}{(\bar{x}_{\text{negative control spot intensity}}) + (3\sigma_{\text{negative control spot intensity}})}$$

Figure 4. Formula for calculating the numerical inputs for ML. Antibody binding is quantified as the light intensity relative to the biochip's background signal (defined by a negative control spot).

Figure 3. Example images of Lyme chips as detected by Ciencia's optical sensing machine from the different patient groups (a and b are positive for Lyme, c, d, and e are negative controls). Ciencia's machine generates target spot intensity values which are then analyzed as GC-FP detection ratios (Figure 4).

Logistic Regression

Logistic Regression (classification)

- Uses sigmoid function to model the logarithmic odds
- During training, optimizes decision boundary
- From inputs, computes probability of class → bins into 0/1 (y)
- Inputs (X) are vectors in *n* dimensions

$$S(x) = \frac{1}{1 + e^{-x}}$$

Figure 9. A variation of the sigmoid function is calculated such that the output falls between 0 and 1 to signify the probability of a sample belonging to either class.

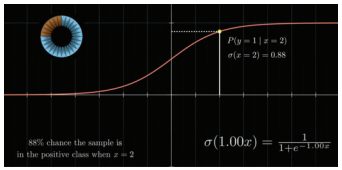


Figure 10. Example of sigmoid function in 2D. X-value input corresponds to a probability on the curve in logistic regression.

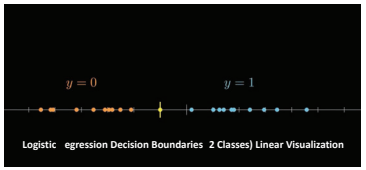


Figure 11. Example of linear decision boundary for binary classification. The purpose of training is optimization of the decision boundary.

Comparing Machine Learning Algorithms

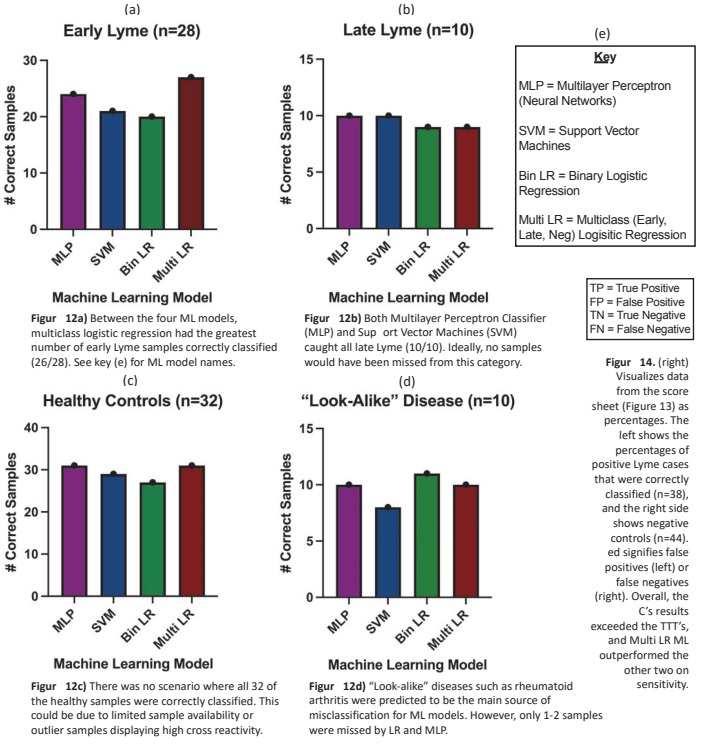


Figure 12a) Between the four ML models, multiclass logistic regression had the greatest number of early Lyme samples correctly classified (26/28). See key (e) for ML model names.

Figure 12b) Both Multilayer Perceptron Classifier (MLP) and Support Vector Machines (SVM) caught all late Lyme (10/10). Ideally, no samples would have been missed from this category.

Figure 12c) There was no scenario where all 32 of the healthy samples were correctly classified. This could be due to limited sample availability or outlier samples displaying high cross reactivity.

Figure 12d) "Look-alike" diseases such as rheumatoid arthritis were predicted to be the main source of misclassification for ML models. However, only 1-2 samples were missed by LR and MLP.

TP = True Positive
FP = False Positive
TN = True Negative
FN = False Negative

Figure 14. (right) Visualizes data from the score sheet (Figure 13) as percentages. The left shows the percentages of positive Lyme cases that were correctly classified (n=38), and the right side shows negative controls (n=44). ed signifies false positives (left) or false negatives (right). Overall, the C's results exceeded the TTT's, and Multi LR ML outperformed the other two on sensitivity.

Comparing Logistic Regression to R C & CDC

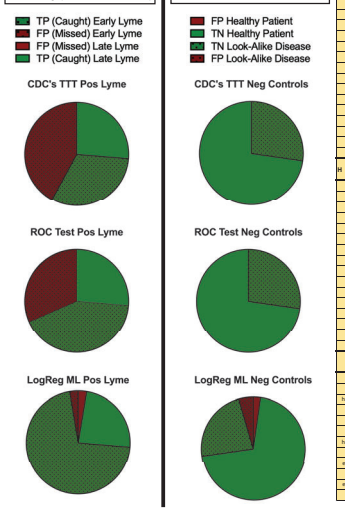
	TTT	C	L Log Reg
Overall ensitivity	58% (22/38)	68% (26/38)	95% (36/38)
Overall pecificity	100% (44/44)	100% (44/44)	93% (41/44)
ensitivity by class:			
Early Lyme	42.9% (12/28)	53.6% (15/28)	96.4% (27/28)
Late Lyme	100% (10/10)	100% (10/10)	90.0% (9/10)
pecificity by class:			
Healthy on-Endemic	100% (16/16)	100% (16/16)	100% (16/16)
Healthy Endemic	100% (16/16)	100% (16/16)	93.8% (15/16)
"Look-Alike" Disease	100% (12/12)	100% (12/12)	83.3% (10/12)
Color Key:	Po i i e	* = if/then left r ig % y thi s val t > adj the c and ed as negat ve	Nega i e

(a) Summary Table

P ei	P i ive Th e hold
O C	> 1.841
D A	> 2.342
D B	> 3.148
P45	> 4.672
E	> 2.435

(b) R C Thresholds

Figure 13. The final product is a score sheet (c) of the 82 samples that will be presented to the CDC using multiclass logistic regression as the chosen ML model. The summary data (a) breaks down sensitivity and specificity by patient group. C threshold values (b) are included, as well as the original TTT results.



mple Group	mple ID		OC	ML Log g
tage of disease or type of negati e control	CDC abels	CDC's Diagnosis by ig or IgG	Antigens e ar positi e threshold	robability score and predicted class
Early Lym (acut)				
	C21813		I_B	1 (Ea)
	C21819			1 (Ea)
	C21825			0.52* (Early)
	R49820			0.99 (Ea)
	R49817			0.62 (Ea)
	C21822	IgM	IP 2C, 39kD, 41kD,	1 (Ea)
	C21808		I_B	1 (Ea)
	C21823			0.61 (Ea)
	C21836		c	1 (Ea)
	C21841		I_B	1 (Ea)
	R49814	IgM		0.98 (Ea)
	R49824			0.52* (Early)
	C21809	IgM		0.76 (Ea)
	C21801			1 (No)
Early Lym (conval sc nt)				
	C21824			0.52* (Early)
	C21840	IgM	I_B	1 (Ea)
	C21849	IgG	39kD, 41kD,	0.75 (Ea)
	R49829	IgM	23, C,	1 (Ea)
	C21835	IgM	2C, 23, 28, 39kD,	1 (Lat)
	C21827	IgM		0.86 (Ea)
	R49812	IgM	23, C,	1 (Ea)
	C21817	IgG	39kD, 41kD,	1 (Ea)
	C21846			0.99 (Ea)
	C21830	IgM	23, C	1 (Ea)
	R49827		23, C,	1 (Ea)
	C21821	IgM, IgG	2C, 39kD, 41kD,	1 (Ea)
	C21803	IgM	23, C,	1 (Ea)
	R49811	IgM, IgG		1 (Ea)
Lat (Lym Arthr tis				
	R4983	IgG	39kD, 41kD,	1 (La e)
	C21847	IgM, IgG	39kD,	1 (La e)
	C21806	IgG	39kD,	1 (Early)
	C21850	IgG	39kD, 41kD,	0.53 (No)
	C21811	IgG	39kD, 41kD,	1 (La e)
	C21828	IgG	39kD, 41kD,	1 (La e)
	C21843	IgG	2C, 23, 28, 39	1 (La e)
	R49823	IgG	2C, 28, 39kD,	1 (La e)
Lat (N uro Lym)				
	R49804	IgM, IgG	39kD, 41kD,	1 (La e)
	R49802	IgM, IgG	2C, 28, 39kD,	1 (Early)
Healt y Non-Endem c				
	C21810			1 (No)
	C21818			1 (No)
	C21845			0.97 (No)
	C21838			0.77 (No)
	C21816			1 (No)
	C21807			0.66 (No)
	C21832			1 (No)
	C21815			0.85 (No)
	C21837			1 (No)
	R49807			1 (No)
	R49822			0.82 (No)
	R49830			1 (No)
	C21844			0.99 (No)
	C21826			0.97 (No)
	C21812			1 (No)
	R49826			1 (No)
H alt y Endem c				
	C21839			0.42* (Early)
	C21848			0.76 (No)
	R49818			1 (No)
	C21831			1 (No)
	C21805			0.85 (No)
	C21829			1 (No)
	R49821			1 (No)
	R4986			1 (No)
	C21834			1 (No)
	C21814			1 (No)
	C21804			0.92 (No)
	C21802			1 (No)
	C21820			0.82 (No)
	C21833			1 (No)
	C21842			1 (No)
	R49819			0.64 (No)
Look-Al s				
Fibromyalgia	R49801			1 (No)
ultiple Sclerosis	R49809			1 (No)
heumatoid Arthritis	R49825			0.87 (Ea)
Fibromyalgia	R49832			1 (No)
yphilis	R49831			0.96 (No)
yphilis	R49816			0.99 (No)
heumatoid Arthritis	R49808			0.95 (No)
ononucleosis	R49815			0.97 (Ea)
e ere Periodontitis	R49828			1 (No)
ononucleosis	R49813			1 (No)
e ere Periodontitis	R49805			0.51* (No)
ultiple Sclerosis	R49810			1 (No)

(c)

C and Explor tory Data Analysis

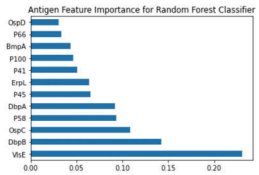


Figure 5. Exploring feature importance illuminates key target proteins. For example, VlsE held the most weight in the Random Forest Classifier, which corroborates findings from the ROC analysis.

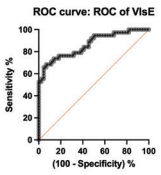


Figure 6. Example OC analysis for VlsE protein. ROC thresholds were chosen based on sensitivity and specificity tradeoffs. Our R C-based test served as a comparison ground for ML.

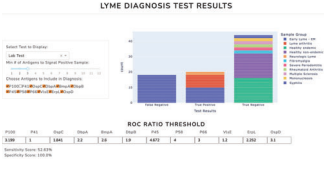


Figure 7. A Plotly interactive Dashboard was built to toggle with various values used in simple R C test. If a certain number of antibody GC-FP detection ratio values exceeded their specified R C threshold, the sample was classified as positive under the R C-based test.

uper ised Machine Learning

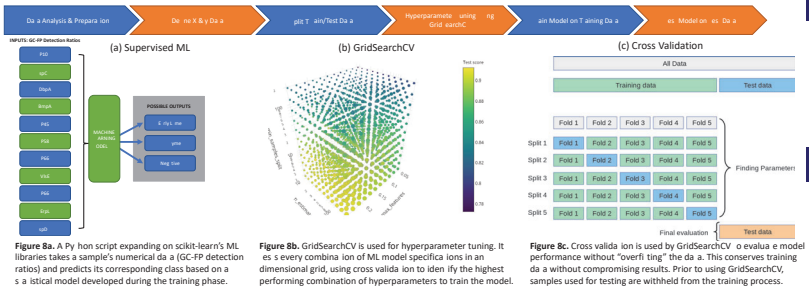


Figure 8a. A Py hon script expanding on scikit-learn's ML libraries takes a sample's numerical da a (GC-FP detection ratios) and predicts its corresponding class based on a s a istical model developed during the training phase.

Figure 8b. GridSearchCV is used for hyperparameter tuning. It es s every combina ion of ML model specifica ions in an dimensional grid, using cross valida ion to iden ify the highest performing combination of hyperparameters to train the model.

Figure 8c. Cross valida ion is used by GridSearchCV o evalua e model performance without "overfi ting" the da a. This conserves training da a without compromising results. Prior to using GridSearchCV, samples used for testing are withheld from the training process.

Conclusions

- Multiclass logistic regression pr vided highest accuracy with additional capacities f r confidence scoring and stage classification.
- was particularly better at detecting early yme but produced 1-3 false positives f r healthy and "lo k-alike" diseases → more data and test standardization should increase diagnosis specificity.
- demonstrates promising results, supporting the GC-FP biosensor as a new test f r yme disease.

Future Directions

- Widen sample database and increase patient diversity t strengthen model.
- ssible incorporation of additional inputs such as relevant patient history, etc.
- As biosensor continues t ev lve, c ntinue t monit r and impr ve machine learning while keeping computational complexity l w.
- Unsupervised t observe data's natural clustering behavior (need larger dataset).
- Biosensor may be applied t ther viral r bacterial infections requiring diagnostic model.

References

1. Chou E., Lasek-Nesselquist E., Taubner B, Pilar A, Guignon E, Page W, et al. (2020). A fluorescent plasmonic biochip assay for multiplex screening of diagnostic serum antibody targets in human Lyme disease. *LoS ONE 15(2): e0228772*. https://doi.org/10.1371/journal.pone.0228772.
2. Grążlewska, W., Ferra, B., Rudzińska, M., & Holec-Gąsior, L. (2021). *Borrelia burgdorferi* BmpA-BBK32 and BmpA-BBA64: New Recombinant Chimeric Proteins with Potential Diagnostic Value. *thogens (Basel, witzerland), 10*(6), 767. https://doi.org/10.3390/pathogens10060767.
3. Pedregosa *et al*. (2011). Scikit-learn: Machine Learning in Python, *JMLR 12*, pp. 2825-2830.

Acknowledgements

We would like t thank Ciencia, Inc. for instrumentation and technical assistance, the CDC f r pr viding the human serum samples and Dr. Frank Doyle f r his c de contributions. Finally, we w uld like t thank the SURP Program at CNSE f r this summer research opportunity.