SUMMER INTERNSHIP REPORT

Dan Suissa Ecole Polytechnique Bachelor Program 2nd year

EPFL Laboratory of Quantum and Nano-Optics July-August 2024

Executive Summary

This report details the work I undertook during my summer internship at the EPFL Laboratory of Quantum and Nano-Optics, under the supervision of Dr. Nicolas Villa. The project focused on machine learning applications in antimicrobial susceptibility testing (AST) using integrated photonic crystal optical tweezers. The primary objective was to explore how cutting-edge machine learning techniques could be combined with advanced photonic tools to improve the detection and analysis of antimicrobial resistance in bacteria.

The project was grounded on the findings from the recent doctoral thesis of Dr. Nicolas Villa, which explored the potential of photonic crystal optical tweezers for bacterial manipulation and testing. My role was to extend this research by incorporating machine learning algorithms for analysing the results obtained through optical tweezers. The experience allowed me to gain insight into both the practical and theoretical aspects of machine learning, photonics, and microbiology, while contributing to a rapidly growing field of biomedical engineering.

This report begins with a description of the EPFL Laboratory of Quantum and Nano-Optics, followed by an outline of my tasks, challenges faced, and the key outcomes of the project. The skills gained and their relevance to my academic and professional goals are also discussed.

**Table of Contents**

* Executive Summary, page 1
* Table of Contents, page 2
* Description of the Company/Institution, page 3-4

1. History and Overview
2. Field of Activity and Structure
3. Why I Chose EPFL for this Internship

* Internship Activities and Position, page 5-10

1. Work Conditions and Responsibilities
2. Main Project: Antimicrobial Susceptibility Testing Using Optical Tweezers

* Assessment of the Internship, page 11-12

1. Skills and Qualifications Acquired
2. Impact on Future Career Plans
3. Relation to Classroom Knowledge

* Conclusions, page 13

1. Internship Experience Summary
2. Observations on the Biomedical Engineering Sector

* References, page 14

Description of the Company/Institution

**History and Overview of EPFL and the Laboratory**

The École Polytechnique Fédérale de Lausanne (EPFL) is one of Europe’s leading research institutions, renowned for its cutting-edge research in engineering, technology, and the natural sciences. Founded in 1853 as a technical school, EPFL has since evolved into an internationally recognized hub for scientific and technological innovation. It is one of the two Swiss Federal Institutes of Technology and has established itself as a global leader in various disciplines, particularly in fields like quantum technology, artificial intelligence, nanotechnology, and biomedical engineering.

The **Laboratory of Quantum and Nano-Optics** at EPFL, where I completed my internship, is at the forefront of research in light-matter interactions at the nanoscale. This lab specializes in investigating the unique optical properties of nanostructures and developing technologies that exploit these properties for applications ranging from quantum computing to biomedical imaging. The lab is led by Dr. Romuald Houdre, who’s past student Nicolla Villa, my supervisor, recently completed a significant thesis on the use of integrated photonic crystal optical tweezers for biological applications.

**Field of Activity and Structure**

The EPFL Laboratory of Quantum and Nano-Optics focuses on both fundamental and applied research in quantum and nano-optics. One of its main research areas is the development of photonic crystal optical tweezers, a tool used for manipulating and analyzing microscopic biological particles, such as bacteria, with high precision. This tool has wide-ranging applications in biomedical engineering, particularly in antimicrobial susceptibility testing (AST), which is the focus of my project. The lab is structured with several research groups that work on overlapping areas of nanophotonics, computational optics, and biomedical devices, integrating expertise from physics, biology, and computer science.

The lab also collaborates with other research groups within EPFL and globally, contributing to various large-scale projects funded by European research initiatives and Swiss national programs. With approximately 15 researchers, including post-doctoral fellows, PhD students, and undergraduates, the lab is equipped with state-of-the-art facilities, including clean rooms for nanofabrication and high-performance computing resources for machine learning experiments.

**Why I Chose EPFL for this Internship**

My decision to undertake an internship at EPFL was driven by its reputation as a world leader in nanotechnology and biomedical engineering. Specifically, the opportunity to work in the Laboratory of Quantum and Nano-Optics was appealing due to its focus on interdisciplinary research, combining quantum optics with machine learning—a field I am deeply interested in. The chance to apply machine learning techniques to real-world problems in antimicrobial resistance offered a unique blend of theoretical and practical knowledge that I believed would enrich my understanding of both fields.

Moreover, the EPFL lab's research on integrated photonic crystal optical tweezers aligns with my interest in exploring advanced optical tools and their applications in biology and medicine. Given the global health threat posed by antimicrobial resistance, I found the lab’s work on AST to be not only academically stimulating but also socially impactful.

Internship Activities and Position

**Work Conditions and Responsibilities**

During my internship at the **EPFL Laboratory of Quantum and Nano-Optics**, I worked closely with my supervisor, **Dr. Nicolas Villa**, from July to August 2024. I was part of a multidisciplinary team focused on the application of **machine learning** in conjunction with **integrated photonic crystal optical tweezers** for **antimicrobial susceptibility testing (AST)**. My responsibilities were aligned with both the theoretical and experimental aspects of the research project.

I worked approximately **40 hours a week**, which included a combination of data analysis, and model development through different experiments. My primary tasks included:

* Conducting a **literature review** to understand the current state-of-the-art machine learning algorithms to perform my analysis.
* Designing and implementing **machine learning algorithms** to process and analyze the data generated by optical tweezers systems.
* Performing **experimental trials** using optical tweezers to trap bacteria and measure their behavior under different antibiotic concentrations (this was more of an observation period where I was more observing the works of my supervisor in order to completely understand the biophysics side of my project).
* Collaborating with fellow researchers to develop my algorithms such as renowned professor Pascal Fua.
* Presenting results in weekly group meetings and writing reports summarizing my progress.

This combination of hands-on laboratory work, coding, and theoretical analysis made the internship an excellent opportunity to deepen my understanding of both **machine learning** and **biophotonics**.

**Main Project: Antimicrobial Susceptibility Testing Using Optical Tweezers**

Initial Task

The primary focus of my internship was to enhance the antimicrobial susceptibility testing (AST) process by combining the precision of **optical tweezers** with the power of **machine learning algorithms**. The overarching goal was to develop a pipeline that could identify bacterial strains and predict their **antibiotic susceptibility** by analyzing the data generated by the optical tweezers in real-time. This innovative approach aimed to provide a faster, more accurate alternative to traditional AST methods, which can be slow and labor-intensive.

The **optical tweezers** system, as described in **Dr. Nicolas Villa’s 2023 thesis**, uses a highly focused laser beam to trap and manipulate bacterial cells. This allows the system to measure various **optical signals** and **force interactions** between the bacteria and the trapping laser. These measurements are critical, as changes in bacterial movement or morphology under antibiotic stress can indicate their susceptibility to the treatment. However, manually interpreting the data from these optical systems is both complex and time-consuming.

My task was to develop **machine learning models** capable of analyzing this data and making real-time predictions regarding bacterial susceptibility to antibiotics. The project consisted of two primary components:

1. **Feature Engineering**: I had to extract meaningful features from the **light scattering patterns**, **force measurements**, and **transmission signals** produced by the optical tweezers.
2. **Machine Learning Classification**: Using these features, I built models to classify bacterial strains and predict their susceptibility to different antibiotic doses.

Solution

The solution to this problem involved building several **machine learning models** and integrating them with the experimental data. To begin with, I worked on preprocessing the raw optical tweezers data. This included removing noise, normalizing the data, and identifying key features. Some of the key features used in my models included:

* **Transmission signal variance**: A crucial indicator of bacterial activity under different conditions.
* **Average and standard deviation of optical trapping forces**: These features provided insight into how bacterial cells responded to antibiotics at the molecular level.
* **Fast Fourier Transform (FFT)**: This technique was applied to capture the **frequency-domain characteristics** of the optical tweezers data, which allowed the detection of recurring patterns in bacterial movement.

After extracting these features, I tested several machine learning models:

* **Random Forest Classifier**: This model was chosen for its ability to handle a wide range of features, robustness to overfitting, and ease of interpretation. The Random Forest model was particularly useful for classifying bacterial strains based on their **Gram type** and optical signatures. It achieved high accuracy in classifying **Gram-positive** and **Gram-negative** bacteria, with an accuracy rate of 100% for Gram type classification in both training and validation datasets​(ML\_project).
* **Convolutional Neural Networks (CNNs)**: Specialized CNNs zrchitectures that iv’ve founbd in research appers were employed to analyze **the time series** generated during optical trapping. The CNNs helped automate the process of **feature extraction** by learning important features from the raw data without manual intervention. Although the CNN-based models showed promising results, they faced challenges in accuracy when dealing with small datasets. However, the model's ability to automatically extract complex patterns proved useful in advancing our understanding of bacterial behavior​(ML\_project).
* **XGBoost**: This powerful algorithm was used for the final classification task, predicting bacterial strains and their response to antibiotic treatment. The XGBoost model was highly effective, achieving an overall classification accuracy of 98% when predicting bacterial strains. Its ability to handle both **linear and non-linear relationships** within the data made it a strong candidate for final deployment in the AST workflow​(ML\_project).

Problems Encountered

One of the major challenges I faced was the **high dimensionality** and **complexity** of the data. The optical tweezers system generates large amounts of **time-series data**, and identifying the most relevant features from this data was non-trivial. In some cases, the bacterial responses to antibiotics were subtle and difficult to distinguish, especially when only small amounts of data were available. **Overfitting** was also a concern, particularly with the more complex models like CNNs, which require large datasets to generalize effectively.

Additionally, processing the data in real-time posed another set of challenges. The machine learning models had to be computationally efficient enough to process data streams in real-time without introducing significant delays in the diagnostic process. This required significant optimization of both the feature extraction and classification stages.

Methods Used to Solve Problems

To overcome these challenges, I employed several techniques:

* **Cross-validation and hyperparameter tuning**: I used these techniques to prevent overfitting, ensuring that the models could generalize well to unseen data. For example, in the **Random Forest Classifier**, I fine-tuned parameters such as the number of trees and depth of the trees to achieve the best possible balance between **accuracy** and **model simplicity**.
* **Feature Selection and Data Augmentation**: I used **Fast Fourier Transform (FFT)** and other statistical techniques like **skewness** and **kurtosis** to identify the most important features for classification. In some cases, I also applied **data augmentation** techniques, such as generating synthetic data points based on existing measurements, to increase the size and diversity of the training dataset, thus improving the model’s robustness​(ML\_project).
* **Parallel Computing and Model Optimization**: To achieve real-time data processing, I optimized the machine learning pipeline by compressing the models, using **efficient algorithms** for inference, and leveraging **parallel computing resources** provided by EPFL’s high-performance computing infrastructure. This allowed us to process large datasets more efficiently while maintaining high accuracy​(ML\_project).

End Result

By the end of the internship, I had successfully developed a pipeline that could predict bacterial susceptibility to various antibiotics with a high degree of accuracy. The **Random Forest** and **XGBoost** models demonstrated **80 and 98% accuracy** respectively in predicting bacterial strains, while also providing useful insights into the dynamics of bacterial movement under antibiotic stress. The integration of **optical tweezers** with **machine learning** showed great promise for revolutionizing the field of **antimicrobial resistance testing**, providing faster and more reliable results compared to traditional methods.

The results of this work lay the foundation for further research and development, particularly in the area of **real-time clinical diagnostics**, where rapid identification of bacterial resistance could greatly improve patient outcomes and help combat the growing threat of **antibiotic resistance** globally.

Assessment of the Internship

**Skills and Qualifications Acquired**

Throughout this internship, I gained a range of technical and analytical skills that will significantly contribute to my future academic and professional endeavors. First and foremost, I honed my ability to work with **machine learning models**, particularly in the context of **biomedical data analysis**. I learned how to apply complex algorithms such as **Random Forests**, **XGBoost**, and **Convolutional Neural Networks (CNNs)** to real-world data, enhancing my proficiency in Python and machine learning frameworks such as **scikit-learn** and **Pytorch**.

Moreover, I improved my **data preprocessing** and **feature engineering** skills, especially when dealing with time-series data from the **optical tweezers** system. I also developed a solid understanding of **optical physics** and its application in **biophotonics**, specifically how light-based tools like **optical tweezers** can be used for antimicrobial susceptibility testing (AST). This was complemented by my exposure to **experimental design** and **laboratory techniques**, where I gained practical experience in data acquisition and analysis in a highly specialized research setting.

**Impact on Future Career Plans**

This internship has had a profound impact on my career trajectory, solidifying my interest in the interdisciplinary field of **biomedical engineering** and **machine learning**. The project highlighted the importance of **artificial intelligence (AI)** in advancing medical diagnostics and reinforced my desire to work at the intersection of **technology** and **healthcare**. I now see a clear path forward in terms of pursuing a graduate degree that focuses on **AI-driven medical technologies**, with a particular emphasis on **combatting antibiotic resistance** through rapid diagnostic solutions. The skills and knowledge I gained during this internship have prepared me to contribute to cutting-edge innovations in **medical AI**.

**Relation to Classroom Knowledge**

The activities I engaged in during the internship were deeply related to my academic coursework. Concepts such as **machine learning algorithms**, **data structures**, and **signal processing** were central to the project, allowing me to apply theoretical knowledge from my classes to practical problems. Additionally, my background in **physics** helped me better understand the principles behind **optical tweezers** and their application in biological research, while my studies in **mathematics** were invaluable in performing **statistical analyses** on the collected data. Overall, the internship bridged the gap between academic theory and real-world applications, reinforcing the relevance of my studies.

Conclusions

**Internship Experience Summary**

My internship at the **EPFL Laboratory of Quantum and Nano-Optics** was an enriching and highly valuable experience. I gained hands-on exposure to cutting-edge **optical technologies** and **machine learning models**, contributing to a meaningful project aimed at improving **antimicrobial susceptibility testing**. Through this experience, I developed a deeper understanding of **interdisciplinary research** and the growing role of **artificial intelligence** in healthcare. My contribution to building a machine learning pipeline for bacterial strain classification and antibiotic susceptibility testing represents an important step forward in rapid diagnostic methods.

**Observations on the Biomedical Engineering Sector**

The biomedical engineering sector is undergoing rapid advancements, particularly with the integration of **machine learning** and **AI** into diagnostic and therapeutic tools. My internship highlighted the increasing need for **real-time data analysis** and the development of **automated solutions** for medical diagnostics, such as **antimicrobial resistance testing**. As AI technologies continue to evolve, they will undoubtedly play a pivotal role in addressing global health challenges, including the growing threat of **antibiotic-resistant bacteria**. This sector presents exciting opportunities for future innovation, and I am eager to contribute to its progress.

References:

[ML\_project.pptx](https://1drv.ms/p/s!ArszpYHIXmEF8HDUKAuGdq0mK4Mf?e=LebwCT), PowerPoint presentation updated by myself throughout the whole internship, I used it to give my weekly report to my supervisors Nicolas Villa and Romuald Houdre but also various professors interested after hearing about this project like Pascal Fua.

ChatGPT helped me turn the PowerPoint presentation into a text.