Improving accuracy of fluid detection on the BD Max with machine learning

by Daniel Khoshkhoo
Mentored by Mr. Brett Pearce and Mr. Joshua Sunday

Introduction

Becton Dickinson, a medical device company based in the United States, offers products that range from needles to high throughput machines. One of their products, the BD Max, is a Deoxyribonucleic acid (DNA) amplification machine capable of real-time Polymerase Chain Reaction (PCR) analysis. The BD Max automates the DNA amplification process to detect genetic markers for infectious diseases. Automated machines have the benefit of lessening human error, lessening exposure time to potentially dangerous samples, as well as saving time and money (Almár et al., 2015).

The BD Max begins the process by extracting the DNA. The BD Max then goes through a series of steps amplifying and mixing solutions from a purchasable kit to prepare the elution. The BD Max uses charged magnetic beads separation which isolates the DNA from impurities which interfere with enzymes in amplification (Berensmeier, 2006). The concentration and volume of DNA need to be high enough for the PCR to be effective. PCR is a common method for DNA amplification. In this process, DNA is heated to break apart the DNA strands, allow for primers to bind to the DNA, and allow enzymes to copy the DNA.

The goal of this project was to improve the measurement system on the BD Max by analyzing data from the light refraction sensor as the medium transitions to air (Figures 1 and 2). The data can be noisy and can produce artifacts from several possible sources. Improving the measurement would allow for better confidence in the process of DNA amplification.

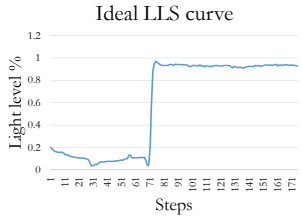

Figure 1 (above): Ideal LLS curve with little to no noise or artifacts.

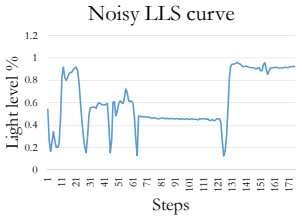

Figure 2 (above): Nonideal LLS curve lots of noise and artifacts reducing accuracy.

Materials and Methods

The BD Max logs were collected from the database at Becton Dickinson. A Python script was written to parse and rebuild the liquid level sensing (LLS) data in the log files. The output was saved into a properly formatted CSV file. PNG files were generated for the LLS graphs to verify the BD Max algorithm and the equation it uses. A Butterworth bandpass filter was applied to reduce the noise and artifacts in the graphs.

Materials and Methods

A graphical user interface (Figure 3) was created for mapping out the machine learning portion of the project to better assess decisions based on human analytical techniques. The Pandas Python library was used to process the data and the TensorFlow Python library was used for training and testing a machine learning model. The data was converted into a TensorFlow dataset and was inputted into a Keras Sequential model with four dense layers. The number of layers was chosen to ensure the dimensionality of the problem was captured.

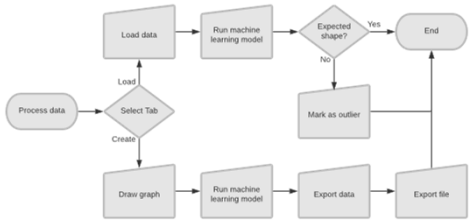

Figure 3: Graphical user interface tool for simulating LLS samples used to train the machine learning model.

To test the new algorithm, 100 standard tubes from the BD Max testing kit were emptied and massed. The code was modified on the BD Max to only run the volume detection subsystem. The BD Max was commanded to aspirate a volume from an interval of 200 µL to 950 µL. The tubes were massed again after the program finished. The volume was calculated using the differences in weights. The output of the machine learning algorithm and the calculated volume were compared.

Results

A Gage R&R study was conducted to compare the measurement systems of the BD Max and the constructed machine learning algorithm. A linearity and bias test and the crossed variant of Gage R&R were run on 90 samples. The study consisted of one operator and parts being sets of four identically commanded tube aspirations.

With a *p*-value of ~0.000, the slope was -0.428 which indicates that the performance is not constant within the 200 µL to 950 µL interval. However, the accuracy tended to increase with larger volumes. This possibly stemmed from an unbalanced dataset.

With a %Study Var of 37.17% (*SD* = 60.399), which is above a targeted 10%, the crossed Gage R&R test suggests that this is not a reliable Gage for volume measurement. The machine learning model had difficulty reproducing the same measurement reducing accuracy. Figure 4 illustrates validation of the model which is believed to be overfitted.

Results

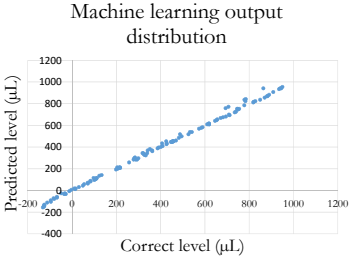

Figure 4: The regression line of the liquid level sensing (r = 0.976) on a test split of the data. The model was able to converge. Despite this, the model was not able to accurately model unfamiliar data.

Conclusions

The goal of this project was to create a machine learning model that could accurately detect the volume of the liquid transferred by the BD Max. The model was not able to maintain consistent readings across the interval with high accuracy. This model tended to be less accurate than the algorithm currently implemented on the BD Max in ideal fluid situations, but machine learning seemed to have better performance when artifacts were present. The machine learning algorithm could be implemented just for those situations.

This project could have been improved using other machine learning algorithms such as OpenCV, which can detect shapes much like a human would when analyzing a graph, creating an algorithm that would be more predictable and easier to manipulate.

In pursuit of state-of-the-art diagnosis in an evolving field of medicine, it is necessary to keep pushing boundaries. With higher accuracy, unnecessary treatments and missed diagnoses can be prevented. There are many ways of measuring volume, some more accurate than others. This project focused on one of the many methods. Overall, this project could serve as a basis for a greater shift to the use of machine learning at Becton Dickinson.

References

Almár, A., Péterfia, B., Wichmann, B., Patai, Á. V., Barták, B. K., Nagy, Z. B., Molnár, B. (2015). Comparison of automated and manual DNA isolation methods for DNA methylation analysis of biopsy, fresh frozen, and formalin-fixed, paraffin-embedded colorectal cancer samples. *Journal of Laboratory Automation, 20*(6), 642–651.

Berensmeier, S. (2006). Magnetic particles for the separation and purification of nucleic acids. *Applied Microbiology and Biotechnology, 73*(3), 495–504.