The Development of a Low-Cost Holistic System for the Stratified Screening of Pancreatic Ductal Adenocarcinoma

Rishabh Ranjan, Gopalaniruddh Tadinada

Pancreatic Ductal Adenocarcinoma (PDAC) is a form of cancer with a 5-year survival rate of 8%. PDAC is hardly detected in its premalignant stages (when surgery is viable), due to modern screening method limitations. This study utilizes a panel of miRNA biomarkers that express TFF1, REG1B, and LYVE1, combined with age, gender, and creatinine, for a diagnostic test to screen for PDAC. This panel of biomarkers is an indicator of increased desmoplasia, or the growth of fibrous connective tissue surrounding pancreatic acinar cells. In addition, a low-cost qPCR solution was also developed to amplify and quantify the biomarker levels from patient samples. The device uses an Arduino-controlled PID loop for the thermocycling of the miRNA. Level extraction is done utilizing a fluorescence detection algorithm coupled with a scientific CMOS (sCMOS) camera. Cycle times, amplification curves, and temperature curves are displayed on a web interface for the user. Once the levels are extracted, they can be inputted into the XGBoost algorithm for a PDAC Screen. When compared to the BioRad 384 well system, the qPCR achieved statistically similar quantifications, as well as quick thermal cycling. With an XGBoost approach, PDAC can be detected with 0.9812 AUC and 93.18% accuracy. This solution can be implemented as a standard of care due to the noninvasive, convenient, and low cost nature of urine tests.

I. INTRODUCTION

Pancreatic Ductal Adenocarcinoma (PDAC) comprises more than 90% of Pancreatic cancer cases (Sarantis et al). Early-stage pancreatic cancer, prior to metastasis, can have symptoms such as abdominal pain, loss of appetite, and fatigue, all of which are typically overlooked. Therefore, a majority (up to 80%) of PDAC cases are detected in later stages (III and IV) when treatment is not viable (Sarantis et al). As a result, PDAC has extremely low survival rates, with only 8% of patients surviving after 5 years (Sarantis et al). Furthermore, cancers along the GI tract grow aggressively, due to the abundance of blood vessels, lymphatic vessels, and the biliary tract, which facilitate metastasis and provide nutrients to sustain enhanced growth. For example, more than 50% of pancreatic cancer patients also have a distant tumor in the liver, thus lowering the effectiveness of treatment (Sarantis et al).

A main factor attributed to the abundance of late-stage patients and poor survivability is the lack of effective and reliable screening measures. Various components limit the effectiveness of pancreatic cancer diagnosis. The most common procedures for pancreatic cancer diagnosis are endoscopy ultrasounds, cT scans, and MRIs. Endoscopy ultrasounds involve high-frequency sound waves for imaging one's abdomen, while cT scans use high amounts of X-ray radiation to produce images. These procedures may be challenging for most people, as they can be expensive, invasive/uncomfortable, or inconvenient. Additionally many of these tests are inaccessible in underdeveloped countries due to the cost of deployment and training to conduct such tests. Furthermore, these procedures have low precision and accuracy as well. The sensitivity rate of these methods is approximately 75% (Giannis et al.), signifying several false-positive results being confused for other diseases. For subtypes, such as pancreatic ductal adenocarcinoma, this number drops even lower.

A combination of price and inconvenience yields low screening rates and for the average patient, the seemingly low-risk symptoms cannot justify these challenges of undergoing ultrasounds and cT scans. Consequently, early signs of severe pancreatic cancer get overlooked until malignancy, yielding poor patient prognosis. To address these issues, researchers are investigating options of a non-invasive, low-price, and accurate tool that can be implemented as a standard of care in clinics to refer high-risk patients for further screening and diagnosis.

This research proposes a novel, non-invasive, urine-based detection system. It utilizes a panel of biomarkers that express TFF1, REG1B, and LYVE1. LYVE1 and TFF1 are both homologs of cancer indicators CD44 and PS2, which promote angiogenesis, or the development of new blood vessels to promote metastasis. REG1B is involved with pancreatic lithogenesis and diabetogenesis [3]. All three of these biomarkers contribute to tumorigenesis and metastasis. Additionally, miRNA profiling may help in detecting PDAC. miRNAs are small non-coding RNA molecules that regulate gene expression by binding to complementary sequences in mRNA and inhibiting their translation into protein. They are involved in a wide range of biological processes, including cell proliferation, differentiation, and apoptosis, and have been shown to be aberrantly expressed in many types of cancer. miRNA profiling has been used to identify miRNA signatures that are specific to certain types of cancer, such as breast, lung, and bladder cancer, and could potentially be used to distinguish these cancers from normal tissue or other types of cancer. As a result, miRNAs have emerged as promising biomarkers for the diagnosis, prognosis, and prediction of therapeutic response in cancer.

This study aims to propose a system that can accurately and noninvasively screen patients for pancreatic cancer, without sacrificing ease of use and cost. The proposed solution attempts to use a panel of miRNA biomarkers associated with PDAC and the optimized machine-learning algorithm for classification. A novel, cost-effective PCR allows for the easy detection of miRNA levels, combined with a client-side web application for customization and usage.

II. METHODS

The methodology behind this project consists of two parts:

(1) Engineering a RT-qPCR device for accurate miRNA biomarker quantification

(2) Developing a machine learning prediction model that can be deployed in clinics

Theory of qPCR

First, a low-cost qPCR device was constructed to accurately quantify the target miRNA levels. An rt-qPCR consists of two functions - thermocycling and an optical system to quantify the DNA/RNA concentration (pg/mL). The objectives of thermal cycling in qPCR are to cycle the temperature between 95℃ to denature the DNA (splitting it into multiple strands), 60℃ for amplification allowing the primers to attach to the various strands, and 72℃ to allow for the polymerase to synthesize a copy of the strand resulting in a 2x increase in DNA concentration every cycle.

1. Preheat at 95℃ for 2-5 mins
2. Denature at 95℃ for 5-15 seconds
3. Anneal at 60℃ for 30-60 seconds
4. Determine the fluorescence intensity of each well/sample
5. Repeat steps 2, 3, and 4, 30-60 times
6. Determine the cycle threshold value (cycle at which fluorescence signal exceeds background intensity)
7. Utilize cycle threshold (cT) and employ a quantification method (ΔΔcT method for example) in order to measure gene expression

Thermal System

In order to accomplish the fast heating times, an aluminum block was used with the sample due to its high and fast thermal conductivity at 235 watts per kelvin per meter. Two 110℃ 12V Positive Temperature Coefficients (PTC) Heaters were used to heat the wells to 95℃, maintained for 5 seconds, in order to denature the sample. To cool the aluminum block, two 12V 7200 RPM fans were installed on both sides of the block. As can be seen in Figure 1, the fans are located accordingly to ensure the even and fast cooling of the block. These were activated, whilst the heaters were turned off to quickly cool the wells to 60℃, maintained for 30 seconds, allowing for the annealing of the primers. Then the heaters and fan are turned on and off sequentially, regulated by a Pulse Width Modulation (PWM) driven Proportional Integral Derivative (PID) controller to cycle between 95℃ (5 seconds) and 60℃ (30 seconds).

Optical System

To quantify the miRNA levels, an amplification curve based on the emitted fluorescence per cycle was charted in real-time on the web page application, and the Cycle Threshold (cT) value was referenced. The optical system works by detecting the fluorescence change after each cycle and then determining the point at which it crosses a certain value. This point is referred to as the cT value. Through a quantification method like the ΔΔcT method, the concentration of the biomarker of interest can be found. In order to quantify the wavelengths and corresponding LEDs, emission filters, excitation filters, and a sCMOS camera needs to be selected. The following optical setup is as follows: a 490 nm blue excitation filter, a 590 nm green emission filter, a 940nm blue LED, and a sCMOS camera fitted with a 140 degree wide lens and custom green IR filter.This allows for the reading of specialized RGB values. The light path is as shown in Figure 1 below. For the purposes of statistical analysis, only the cT values for the conducted trials were analyzed. cT values are indicative of the amount of nucleic acid (the building blocks of RNA). A higher the cT value corresponds to a lower amount of target RNA. Thus, cT values are a reliable metric of measuring the accuracy of the device performance.

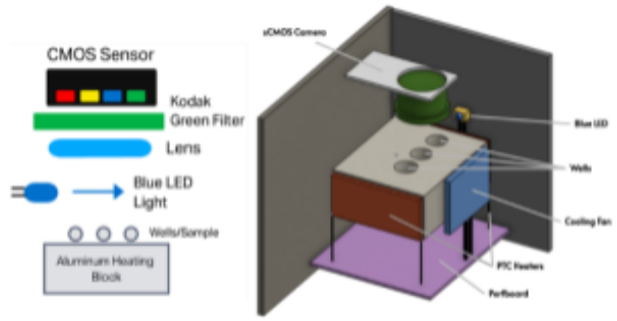

Figure 1: Detailed layered optical system illustration (left); 3D Model of Low-Cost PCR device highlighting each of the major components: thermal system, optical system, wells, circuits (right)

Machine Learning Analysis of Concentrations

To train the machine learning model, a 591-case database was sourced from Kaggle [3]. A total of 6 algorithms were tested, with the training-to-testing ratio 3:1. Prior to passing data into the screening model, the data needed to be converted to a usable and appropriate format. Firstly, a label encoder was called in order to transcribe categorical variables into float values. Secondly, a minmax scaler was fitted in order to convert the miRNA values to values between 0 and 1 for the training of the machine learning and deep learning models. Lastly, Synthetic Minority Oversampling Technique (SMOTE) was applied in order to regularize the frequency of the minority class which is hepatic/liver cancer. Then data was passed into models.

Machine Learning Models can be trained with significantly less data than deep learning models. In order to evaluate the model after training the algorithm, validation data is used to gauge how effective the model is. Tuning, optimization, and regularization options have limited benefits for pattern recognition models, because hypertuning parameters and ensembling generally overfit the models. Accuracy, Area-Under-Curve (AUC), and F1 score were used to determine the effectiveness of each model. False-positive rates were also considered, since they are indicative of

underlying benign pancreatic issues, such as pancreatitis.

	No Cancer	Benign non-Cancerous	Stage 1&2	Stage 3&4
Train	137	156	76	72
Test	46	52	26	25

Table 1: Breakdown of training and testing data for ML algorithms

Testing the Device Against Lab Standards

In order to gauge the effectiveness of the low-cost device, an experimental procedure was devised in which the cT values outputted by our device were compared to the cT values of an actual lab-grade qPCR. Due to availability, a cisplatin-treated A549 Lung Cancer Cell Line was used. Within the A549 Cancer Cell Line, different genes were identified for our device to quantify - B2M, CXCL1, IL6, and CDKN1A. Before testing, A549 plates were treated with the drug Cisplatin at 12 distinct concentrations, clinically relevant to cell proliferation and mRNA levels. Cells were harvested from 0.39-1.56 µG/mL. From these cells, total RNA was extracted utilizing a Qiagen miRNA extraction kit. Afterwards, high capacity cDNA was prepared utilizing Qiagen cDNA synthesis protocol. Lastly, the cDNA was used for rt-qPCR, and the resulting cT values were recorded.

III. RESULTS

Logistic Regression	91.84%	95.83%	0.87
SVM	91.22%	95.55%	0.87
Naïve Bayes	84.46%	91.14%	0.83
Random Forest	91.21%	97.28%	0.99
Light GBM	86.82%	94.87%	0.84
XGBoost	93.18%	98.12%	0.93

Table 2: Machine learning algorithm results with 3 performance metrics displaying how each algorithm performed side by side

A number of metrics were used to evaluate the quality of the screening system. Both quantitative and qualitative results were collected for each sub-system of the device, notably heating, cooling, and sCMOS fluorescence and image capture. The functioning of the device as a whole was also analyzed through lab qPCR of samples. To evaluate the performance of each model, accuracy, AUC score, and F1 score was collected. AUC is the area under a ROC curve, which measures true vs. false positive rate, whereas F1 score is the harmonic mean of the precision-recall.

As can be seen, the ML algorithms performed exceptionally, with XGBoost boasting 93.18% accuracy and 98.12% AUC. It is important to note that a majority of false-positives are patients that experienced some benign growth or gastrointestinal disease that could increase the risk of cancer, for example, pancreatitis. Thus, it is still important for such patients to go in for further tests such as ultrasound, after being classified as high-risk.

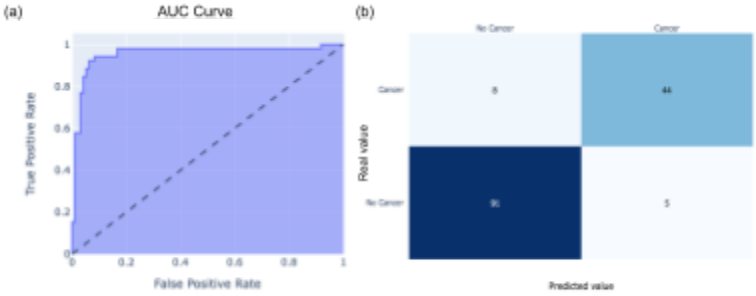

Figure 2: (a) AUC showing false positive vs true positive rate for Logistic Regression (b) Confusion Matrix showing true, false positives & negatives for Logistic Regression

The heating and cooling times for the thermal system were measured over 350 times. The time to heat until 95℃ and cool down to 60℃ was measured in seconds for each cycle and plotted on the graph in Figure 3.

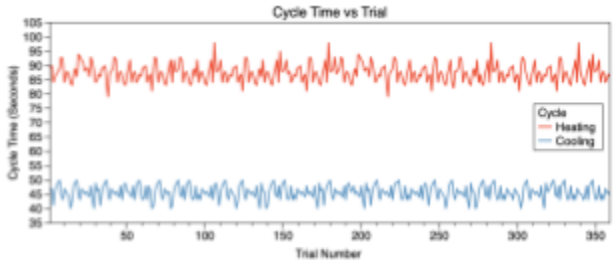

Figure 3: Cycle times for heating and cooling in seconds across 350 trials. The upper red band shows heating times, while the bottom blue band represents cooling times.

As can be seen, the thermal system is extremely consistent across the several trials, for both heating and cooling. The average heating time is 87 seconds, while the average cooling time is 45 seconds. Furthermore, a key component is for the user to be able to keep track of the temperature and cycle times in real time. Hence, our thermal plotter directly receives temperature readings from the device and in real time, plots the cycling curve, shown in Figure 4.

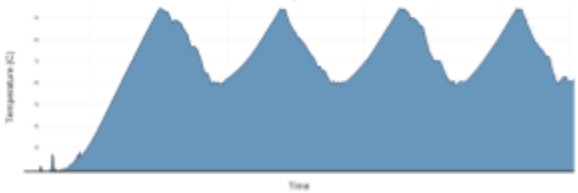

Figure 4: Outputted thermal cycles on the Python based Web application after 4 cycles.

The resulting graphs and cT values that were outputted by the lab qPCR and the low-cost device are displayed in Figure 5.

As can be seen when comparing the graphs for B2M and CXCL1 against the constructed device outputted graphs, the differences appear negligible. Identical behavior can be observed among the graphs and the values when looking at the fluorescence for each cycle. It was determined that the p-value of aggregated cT values was 0.62, indicative of no statistically significant difference in the performance of both tested devices.

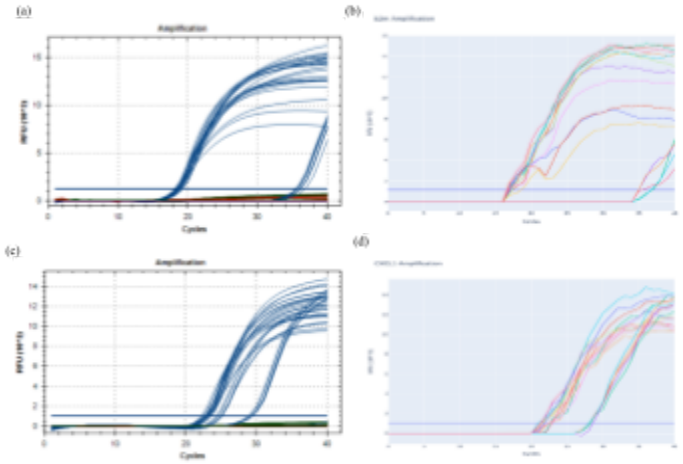

Figure 5: Diagrams (a) & (c), display lab PCR amplification curves for CXCL1 & B2M, respectively. Graphs (b) & (d) display the amplification curve plotted by the constructed qPCR device for B2M and CXCL1, respectively.

IV. CONCLUSION

This project attempts to address the primary issue in current healthcare of pancreatic cancer, poor screening and diagnosis, with a low-cost novel pipeline to detect gastrointestinal cancers with striking accuracy. By combining artificial intelligence, novel spectroscopy methods, and a straightforward user interface for clinical application, this system holds the potential to revolutionize gastrointestinal oncology. The pipeline starts with urine analysis with our PCR prototype. PCR is an extremely common procedure that is used for a plethora of lab applications. However, with increasing dependency on biotechnology for medical use, it is critical that everyone has access to such equipment. Underdeveloped countries still have a severe shortage and this significantly reduces their ability to improve patient prognosis. Unfortunately, people in these areas tend to be more impacted by gastrointestinal cancer due to increased exposure to risk factors, such as poor diet and gut infection. Our PCR device combines an innovative thermal cycling system with a novel spectroscopy optical system to produce reliable analyses of biofluids, such as urine. Thus, the construction of this device is feasible in even underdeveloped countries, due to the low-price and accessible materials. The artificial intelligence model and user interface are also low-resource consumption, only requiring a simple computer and software.

Compared to tests like CT Scans and ultrasounds, urine analysis is convenient, cost-effective, and comfortable. It is a non-invasive approach to diagnosing disease and thereby should be better received by patients. In essence, this means that urine tests in annual checkups can now be used as a preliminary checkpoint for pancreatic cancer testing to refer patients for further measures. Even aside from gastrointestinal cancer, this system easily detects gastrointestinal abnormalities and benign growths, meaning it can associate patients with being at higher risk for obtaining a tumor. This solution not only affects healthcare at a local, regional, or national level, but can be implemented worldwide**.** Since this pancreatic cancer and similar disorders affect patients everywhere, it is urgent to diagnose early.

Specifically, early detection can result in better prognosis trajectory and more desired outcomes for patients. Pancreatic cancer patients can have more treatment options available to them, such as the Whipple procedure, and several other surgical procedures that are fruitful for premalignant cancer. Furthermore, chemotherapy and radiation therapy are far more effective when targeting a single tumor instead of cancer that has metastasized. Although these treatments exist for all patients, they do not yield good results beyond stage II. This urine test system could facilitate cancer detection in regular doctor checkups, making it possible to identify cancerous growths in the pancreas earlier.

Finally, there are many options for using this urine test system for other diseases. Spectroscopy and PCR are methods that can be used for a variety of applications. Although this device is pre-tuned to analyze the miRNA levels corresponding with pancreatic cancers, parameters can be adjusted for other diseases via a connected user interface. For example, this device's use in the detection of water microorganisms and pathogens is currently under investigation. Upon the discovery of other gene marker panels, this device can be customized to diagnose a plethora of diseases, including other cancer types and infectious diseases.

V. REFERENCES

[1] Cole Parmer. (2022, May 31). PCR Process Steps Explained. Cole Parmer. Retrieved June 13, 2022, from https://www.coleparmer.com/tech-article/pcr-process-steps-explained

[2] Debernardi, S., O'Brien, H., Algahmdi, A. S., Malats, N., Stewart, G. D., Plješa-Ercegovac, M., Costello, E., Greenhalf, W., Saad, A., Roberts, R., Ney, A., Pereira, S. P., Kocher, H. M., Duffy, S., Blyuss, O., & Crnogorac-Jurcevic, T. (2020). A combination of urinary biomarker panel and PANCRISK score for earlier detection of pancreatic cancer: A case-control study. PLOS Medicine, 17(12). https://doi.org/10.1371/journal.pmed.1003489

[3] Johns Hopkins Medicine. (2021, August 8). Pancreatic Cancer Prognosis. Johns Hopkins Medicine. Retrieved June 13, 2022, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/pancreatic-cancer/pancreatic-cancer-prognosis

[4] Mendoza-Gallegos, R. A., Rios, A., & Garcia-Cordero, J. L. (2018). An affordable and portable thermocycler for real-time PCR made of 3D-printed parts and off-the-Shelf Electronics. Analytical Chemistry, 90(9), 5563–5568. https://doi.org/10.1021/acs.analchem.7b04843

[5] Pancreatic Cancer UK. (2021, February 10). Types of Pancreatic Cancer. Pancreatic Cancer UK. Retrieved June 13, 2022, from https://www.pancreaticcancer.org.uk/information/just-diagnosed-with-pancreatic-cancer/types-of-pancreatic-cancer/

[6] Yang, J., Xu, R., Wang, C., Qiu, J., Ren, B., & You, L. (2021). Early screening and diagnosis strategies of pancreatic cancer: A comprehensive review. Cancer Communications, 41(12), 1257– 1274. https://doi.org/10.1002/cac2.12204