

Welcome to QMIND's 2023 Project Look Book. The work presented within this document is the collection of a year-long effort. This year our design teams finished a series of projects spanning innovation, research, disruptive technologies, and client-facing deliverables. Each project is unique in scope and granted our members the creative freedom to imagine solutions to problems yet to exist. Our mission, to empower undergraduate students to unlock the full potential of AI, was delivered through hours of mentoring, problem solving, and fostering collaboration among the undergraduate student body.

This past year, our design teams took their projects to a new level, presenting it in front of their biggest stage yet – at CUCAI with 320+ attendees. These projects have been shown to industry professionals, and fellow enthusiasts just like us. Our excitement for the AI innovations in this past year was expressed on multiple occasions throughout these past months and is something we take pride in. QMIND is excited to show you what it means to lead real change.

Project Index

PureFacts Anomaly Detection	3
Detecting Security Vulnerabilities in Code	4
Branching Quantum Convolutional Neural Networks	5
EEG Brain Computer Interface	6
Lung Cancer Detection Using Deep Learning	7
Towards the Responsible Development of AI	8
DCP Sentiment Analysis	9
Applying Machine Learning to Bipolar Disorder Categorization Using a Motor Activity Dataset	10
Deepfake Lip-Syncing	11
Forecasting Stock Price Movement with Google Trends Data	12
NHL Predictions	13
Analyzing ChIP-seq Profiles using Machine Learning	14

PureFacts Anomaly Detection



DANIAL
Quinn

CONNOR
Rewa

SEBASTIAN
Wakefield

STEPHEN
Li

NICHOLAS
Seegobin

We built an anomaly detection program for financial data, aimed at detecting anomalies in fee calculations in PureFacts Financial Solutions' wealth management solutions. Our work employs unsupervised learning approaches, as the data provided was not classified as anomalous/non-anomalous. We detect anomalies by clustering the data using DBSCAN and then using Davies-Bouldin and Silhouette values to check the optimal epsilon value. The approach successfully detected anomalies in the data, which could save the institution from millions of dollars in fines, regulation costs, and compliance issues.

Detecting Security Vulnerabilities in Code



SYED
Moiz

OLIVIA
Deng

MERCY
Doan

ALICE
Slabosz

KYLE
Levy

While NLP continues to have success with human languages, there is not a lot of research on using it to explore other text-based data, such as code.

Cybersecurity experts define a code vulnerability using the Common Weakness Enumeration (CWE) standards. They can identify vulnerabilities using static code analysis, where they manually pattern match code. In contrast, dynamic code analysis relies on programmers providing their own unit testing. Both methods are tedious and rely on human tenacity. Overall, our question was: can we use AI and NLP techniques to detect vulnerable code that is not immediately visible?

We aimed to use Transformers to classify secure vs. vulnerable code with at least 50% accuracy, and compare our findings to non-Transformer models. Using a dataset of 240k+ labeled PHP files, we were able to compare models using Naive Bayes, neural networks, and Transformers.

Branching Quantum Convolutional Neural Networks



CHRIS
Kingsland

ABDELLAH
Ghassel

JACOB
Ewaniuk

TAYLOR
Balsky

ELAD
Dallal

While NLP continues to have success with human languages, there is not a lot of research on using it to explore other text-based data, such as code. Cybersecurity experts define a code vulnerability using the Common Weakness Enumeration (CWE) standards. They can identify vulnerabilities using static code analysis, where they manually pattern match code. In contrast, dynamic code analysis relies on programmers providing their own unit testing. Both methods are tedious and rely on human tenacity. Overall, our question was: can we use AI and NLP techniques to detect vulnerable code that is not immediately visible? We aimed to use Transformers to classify secure vs. vulnerable code with at least 50% accuracy, and compare our findings to non-Transformer models. Using a dataset of 240k+ labeled PHP files, we were able to compare models using Naive Bayes, neural networks, and Transformers.

EEG Brain Computer Interface



ETHAN
Callanan

JACK
Macaulay

DANIEL
Marlin

KIARASH
Mirkmandari

We built a brain-computer interface, using an EEG and ML classification techniques. The interface works as a binary controller, and we integrated it as the controller to play Flappy Bird. Of the several classifiers we built, the best was a SVM achieving 94.43% accuracy. The interface works in real-time, allowing seamless control of the game.

Lung Cancer Detection Using Deep Learning



JACKSON
Kahoe

JOHN
Zhou

DAVID
Nguyen

SIYUNG
Cho

The study aims to investigate the feasibility and precision of detecting lung cancer solely through x-ray data. Currently, a biopsy and CT scan are required for accurate diagnosis after an initial x-ray screening. By relying solely on x-rays, the process can become more efficient and accessible to individuals in rural areas or those with inadequate healthcare systems. The dataset utilized in this study was obtained from the NIH Clinical Center and consists of multiple angles of x-ray images from a large number of patients. However, some of the images did not display cancer clusters. The study was able to accurately identify cancer-causing images using the x-ray images with an accuracy rate of about 70%. Additionally, the study explored predicting bounding boxes on the x-ray images. However, the available data for bounding boxes was insufficient to produce reasonable results. For future iterations of the project, it is recommended that a model be developed to select the clearest image of a patient using x-rays before predicting cancer. Furthermore, a larger annotated dataset for bounding boxes would improve the accuracy of predicting and locating lung cancer.

Towards the Responsible Development of AI



CYRUS
Fung

SETH
Grief-Albert

AAMIYA
Sidhu

Progress in Artificial Intelligence is accelerating. The introduction of AI applications into the public domain will have profound effects on several institutions. This necessitates reevaluating the ways in which revolutionary technology has been and continues to be developed. It is becoming increasingly clear that the mindset of Moving Fast and Breaking Things has led to unintended consequences and negative impacts on society. This paper calls for a shift in focus from rapid development to responsible development, demonstrating the need for a more thoughtful approach to the creation and deployment of AI.

DCP Sentiment Analysis



HUNTER
Hoogendijk

AMINAH
Rizwan

BRIAN
Minkovich

RAYYAN
Ramesh

Our team partnered with Distributive and the Jomo Kenyatta University of Agriculture and Technology to develop an NLP model to conduct sentiment analysis and determine if tweets were classified as bullying or non-bullying. The project was created using a BERT model and a separate LSTM model which were then deployed in the Distributive network for parallel processing of the data.

Applying Machine Learning to **Bipolar Disorder Categorization** Using a Motor Activity Dataset



NASER
Al-Obeidat

RABAB
Azeem

JINGJING
Mao

ARIA
Maz

JULIEN
Pierre Chanel

Bipolar disorder, especially bipolar II disorder, is known to have a high suicide and self-harm rate and a high misdiagnosis rate. This project is an attempt to identify patients with bipolar II disorder among a group of healthy controls and patients with major depressive disorder. We use the motor activity data collected by the motion-sensitive sensor at Haukeland University Hospital, Bergen, Norway. We discuss the preprocessing methods performed and a potential method of feature extraction. We applied a convolutional neural network and a long short-term memory network and the accuracy was 63.48% and 75.77% respectively. Finally, we discussed future directions and suggested methods of improving the accuracy.

Deepfake Lip-Syncing



KHOA
Nguyen

DUKE
Do

YU XUAN
Li

HUY
Nguyen

Using a video of my face and a piece of audio, the program can recreate it saying a piece of audio. The program is meant to be used with a fake webcam, which allows users to attend video conferences and talk without actually having to be in front of a camera. The final result, though checked a lot of boxes, left a lot to be desired in terms of image quality. In a few seconds, the program can generate that resembles the original subject in skin color, face shape, face orientation and some general features, but proves lacking in details. Though the mouth does move to the audio, the change does not appear significant enough to fool anyone.

Forecasting Stock Price Movement with Google Trends Data



THOMAS
Tesselaar

GARY
Faberov

MADDY
Chapnik

ANNA
Lou

Accurately predicting how a stocks' price will behave is one of the most examined problems in finance. More accurate prediction methods have the potential for large financial gain. At the same time, Google Trends can serve as a window into people's interests and a reflection of broad trends. We hypothesized that some search terms may be useful as predictors. To do this, we looked at three stocks and found search terms that correlated with each one. We built a predictor to predict price movement, which outperformed the market for all three equities. Finally, we simulated trading with our model dictating trading decisions and found we returned higher than the market for all three equities we tested.

NHL Predictions



ADAM
Cormier

JORDAN
Capello

JORDAN
Siwak

MAX
Dizy

ELLIOT
Thoburn

Our team created a machine learning model to predict the outcome of National Hockey League (NHL) games during the 2022-23 regular season. We used data and statistics from previous NHL seasons, and the model predicted the outcome (winner / loser) of a game of interest. Each team member focused on a specific NHL team and used different past statistics as input variables for a machine learning model. This provided our team with insights on the effect of different past statistics on future NHL game outcomes. The project was intended to discover the most profitable NHL game betting odds on popular sportsbooks.

Analyzing ChIP-seq Profiles Using Machine Learning



CIARAN
Byles-Ho

KATE
Rodrigues

DOMINIC
Trinh

JONAH
Barclay

PETER
Danilin

THEO
Lemay

Transcription factors (TF) are essential proteins that control the rate of transcription and assist in controlling cellular functions. Understanding how TFs interact with other proteins and bind to DNA has significant implications on understanding cellular processes, identifying mutations and developing effective drugs. The project analyzed chromatin immunoprecipitation (ChIP-seq) profiles using machine learning and evaluated bioinformatics tools for predicting the profiles. Using the FOXA1, HNF4A, and ONECUT transcription factors with the human genome, it was observed that models with increasing layers in neural networks and larger basepair widths improved performance. Additionally, real mutations were applied to human DNA sequences and yielded minimal permutations to non-mutated nucleotide sequences. The project was run in Collaboration with Dr. Wilson's Lab at the SickKids Research Institute.