CCSCNE '23

The Twenty-Seventh Annual
Consortium for Computing Sciences in Colleges
Northeastern Conference

April 14 – 15, 2023

Hosted by Ithaca College
Ithaca, New York

Undergraduate Student Poster Abstracts

CCSC
CONSORTIUM FOR
COMPUTING SCIENCES
IN COLLEGES
ccsc.org

Contents

Poster 5001

A Novel Smart Electronic Liaison to Support International Students

Ahmed Muntasir Hossain, Killian Meehan, Emmanuel Amoh
Electrical & Computer Engineering and Computer Science Department
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

As the number of international students attending domestic universities increases, it is imperative that administrative programs make changes to accommodate them. International students face critical challenges before and during their time at the university causing the experience to be unnecessarily stressful. The primary channel of communication with service offices is via email. Communication via email is commonly manual, limited to office hours, and poses challenging issues ranging from inconsistent information to long response times, cost for the university, and limited scalability. To overcome this problem, Natural Language Processing (NLP) techniques particularly Conversational Agents (chatbots) are potential candidate solutions. Chatbots are automated, available 24/7, consistent, precise, scalable, and cheap. Therefore, we propose to design, develop, integrate, and evaluate a conversational agent prototype at the University *of* New Haven that will serve as a smart electronic liaison for international students and service offices. The proposed smart electronic liaison will include (1) admission and immigration module, (2) tutoring module, and (3) graduation module for international students. The admission and immigration module intends to address questions related to employment authorizations and program admission requirements. The tutoring module aims to assist international students by suggesting annotated pre-recorded media from tutors in response to questions. Lastly, the graduation module intends to plan and optimize the graduation timeline of international students with respect to offered courses and associated prerequisites. To verify and validate our smart electronic liaison, we are working closely with the University Immigration Services and Center for Learning Resources at the University *of* New Haven to analyze requirements, design the software, develop the software, collect data, and train the chatbot. For future work, the smart electronic liaison will be expanded to target a larger student community and incorporate new services that support housing, finances, dietary needs, and more.

Poster 5002

Hierarchical Ensemble Machine Learning for Fake News Detection

Ryan Himes
Computer Science and Engineering
The Pennsylvania State University, University Park, PA

Faculty Advisor
Truong Tran
The Pennsylvania State University, Harrisburg, PA

Misinformation, or fake news has become a significant problem in the modern day. This can lead to changing peoples' world views that revolve around information that is not true. Hence, the importance of automatically detecting it has become necessary with the amount of news freely flowing throughout the internet. Companies and researchers in the past have found ways to reliably detect whether or not news contains misinformation in it through binary classification methods. Sometimes this classification cannot tell the overall story of how truthful certain news is when parts of the news are truthful while the other parts are not. This problem can be fixed by applying a more precise label than true or false to a story, such as mostly true or half true. To do this more accurately, a hierarchical ensemble method is proposed, which follows a divide and conquer approach that splits the data into smaller sub-sets to label news more accurately. After testing on three different news data sets, the hierarchical ensemble method outperforms all other known methods on two out of the three data sets. These results show that the hierarchical model can be a viable ensemble method for other data sets.

Poster 5003

HellSmoke

Ryan Gibbons, Lester Barrios
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

Pleiadean Systems Corporation created an immersive, virtual reality firefighter training simulation platform, called HellSmoke. This simulation platform incorporates gaming techniques and provide rich simulation scenarii. Hellsmoke offers realistic training exercises, preventing fire fighter injuries, and giving trainers access to assessment metrics to evaluate their performances. However, HellSmoke suffers from critical technological limitations. First, HellSmoke fails to provide realistic fire and smoke simulations because of poor task optimization, limited memory management, and missing parallelism processing. Moreover, the current software architecture is a conventional desktop solution with traditional client-server approach. Because of this architecture style, HellSmoke requires expensive hardware with advanced specifications involving a sophisticated laptop and a set of VR headsets. Such owned hardware incurs operational and maintenance costs and limits scalability, portability, and accessibility of Hellsmoke. To address these challenges, we propose to conduct a performance study which considers a smoke and fire simulation model that enables parallelizes processing leveraging the combination of Central Processing Unit (CPU) and Graphic Processing Unit (GPU). Additionally, we propose to take advantage of the capacity of cloud computing to revise HellSmoke's software architecture to offer anywhere accessibility, flexible resources, and economies of scale. With our contributions, fire science students and professional fire figthers can run HellSmoke without powerful hardware, access a training session anytime and from anywhere without compromising the realistic, interactive, and immersive quality of the learning experience.

Poster 5004

Sentiment Analysis and Recurrent Neural Networks Utilizing the Sentiment140 Dataset

Tye Leckinger

Department of Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisor

Zakariya Qawaqneh
SUNY Brockport, Brockport, NY

As a student at SUNY Brockport, majoring in Computer Science, I will be presenting my experience working with the Student Undergraduate Research Program at Brockport. The problem I worked involved the marketing world, where being able to understand sentiment is a difficult but important task. Customer feedback on an existing product and seeing if people would be interested in buying a new product are both important when making a business decision. However, when analyzing feedback from thousands of people from different mediums, having a human analyze feedback is not feasible. For this reason, the idea of a computer being able to do the same task, but do it much faster, is something of interest to marketing companies. With the Sentiment140 Dataset developed by Alex Go, Richa Bhayani, and Lei Huang the authors attempted to tackle this problem with a Maximum Entropy approach. While their model performed well with an accuracy of 83%, we also hope to attempt this problem as well with a Deep Learning Recurrent Neural Network (RNN) approach instead, and present these results in our poster.

Poster 5005

Sentiment Analysis of Social Media (SASM)

Ahmed Muntasir Hossain

Electrical & Computer Engineering and Computer Science Department
University of New Haven, West Haven, CT

Faculty Advisor

Mehdi Mekni
University of New Haven, West Haven, CT

Digital reputation management systems have become increasingly popular as individuals and businesses seek to maintain and improve their online reputation. However, existing systems, including Online Social Network Interactions (OSNI), are suffering from critical limitations including limited effectiveness, inaccuracy, high costs, and limited scope. Sentiment analysis is a natural language processing technique used to identify and extract opinions, emotions, and attitudes expressed in textual data. Sentiment analysis can be useful in digital reputation management. In this study, we propose to create an open-source, multi-channel, multi-engine sentiment analysis software for social media and digital reputation management purposes. We call this system *Sentiment Analysis of Social Media* (SASM). SASM collects data/posts from three different social media channels, Twitter, Reddit, and Tumblr. It then filters, aggregates, and analyzes trends in the sentiment of content posted on social media while leveraging different sentiment analysis engines including *Microsoft Text Analytics*, *IBM Watson Natural Language Understanding*, and *Google Cloud Natural Language*. To verify and validate our system, we consider a case study focusing on three major information technology companies: Google, Amazon, and Microsoft. The outcomes of this case study aim to explore how social media content about major information technology companies vary depending on various factors including geo-political, socio-economic, and environmental awareness. SASM is original because it generates rich and reliable sentiment analysis results, leverages well-established sentiment analysis engines, and collects content from various social media channels. SASM will allow companies to manage their digital reputation effectively and affordably on social media and evaluate customer loyalty, competition, and demand trends accurately.

Poster 5006

Experiences in Enhancing Functionalities in Apps Using a Customer-Focused Agile-Oriented Approach

Ethan L. Baker, Matthew E. Morgan
Department of Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisor
Sandeep Mitra
SUNY Brockport, Brockport, NY

We describe our experience in developing a student assessment data capture and management app for a campus-based committee. The primary goal was to develop the app to be customized to the specific needs of the customer. The code from a similar app developed for another campus customer during the previous year was available to us. While this app had some features that were also replicated in our project, and that code could be reused as is, much of the functionality had to be adapted to the new customer's requirements, and a number of features were completely new to this project. To implement this application, we took an Agile-oriented approach in which we stayed heavily engaged with the customer throughout, seeking to meet the customer's goal of a user-friendly and intuitive GUI. Throughout the project, we encountered various obstacles. First, in a small university with no graduate program, there were few qualified students available for the project team, so the team was relatively small compared to the size of the project. In addition, we had a hard time limit dictated by the academic calendar and graduation dates. The ability to reuse code from the previous project helped, and the time spent in understanding this code was reduced considerably by the previous project's adherence to appropriate design patterns and that our project used the same technology tools. To succeed in this environment, we also observe that it is necessary to work with the customer to identify the project scope, take a Scrum-based approach to track progress, and conduct code reviews to keep everyone "on the same page". It is also critical to ensure that we get a team with the right skills, and ensure adherence to coding standards, for which we note that faculty member involvement in identifying the right students (rising juniors, for example), and *especially* in ensuring coding standards are met, is necessary.

Poster 5007

Performance Evaluation of MQTT & CoAP for IoT Applications

Diego Segovia Ayala

Computer Science
Eastern Connecticut State University, Willimantic, CT

Faculty Advisor

Garrett Dancik
Eastern Connecticut State University, Willimantic, CT

The Internet of Things (IoT) industry is growing at a record-breaking rate with billions and billions of devices being connected to the Internet to transmit important information and complete critical tasks. In general, IoT devices lack the computational resources of traditional computers making it crucial to use their resources efficiently and to implement appropriate networking protocols. Despite the necessity for reliable data transfer, effective resource usage, and interoperability amongst IoT devices, no application layer protocol offers universal features for all IoT applications and devices. In this project I compare the performance of two application layer protocols, Constrained Application Protocol (CoAP) and Message Queuing Telemetry Transport (MQTT), to choose the most appropriate protocol for resource limited IoT applications. For that purpose, clients and servers are configured on a local area network to transmit JSON data to one another. The protocols are tested for 30 seconds using Wireshark in two different payload configurations. Results show that MQTT achieves higher packets per second and total packets sent for both payload types compared to CoAP. This demonstrates that MQTT is more suitable than CoAP for IoT applications that require quick data delivery.

Poster 5008

Enhancing the Mead Art Museum Metadata: Using Google Cloud Vision API for Tagging Artwork

Reihaneh Iranmanesh, Adam Rogers
Computer Science
Amherst College, Amherst, MA

Faculty Advisors
Lee Spector, Mihaela Malita, Miloslava Hruba
Amherst College, Amherst, MA

The cross-depiction problem, recognizing visual objects regardless of their mediums, is a significant yet under-researched problem in computer vision (CV). In this paper, we present a project aimed at tackling this problem. Specifically, we use CV to enhance the accessibility and searchability of Amherst College Mead Art Museum's database by generating tags for approximately 22,000 web images in the collection. For example, users interested in "dogs" could search artworks tagged with this term.

Google Vision, a CV model trained with realistic images, was employed to classify artworks using thousands of labels. The tagging process involved building a common terminology between Mead's predefined keywords and Vision's terms and sending two API requests to Vision (*label detection and object localization*) for each image. Each API response included the ten tag terms having the highest confidence scores. Instead of using a fixed threshold to filter Vision tags, we refined labels based on the highest confidence score (HCS) in each response, demonstrating greater efficacy. To speed up the process, the images were divided into batches and processed concurrently.

The results showed that Vision tends to use general terms like "animals" instead of more specific labeling (e.g. "owls"). Although Vision tagged the artworks rapidly, it is not yet capable of fully recognizing creative renderings of objects across mediums and styles, suggesting that pre-trained CV models like Google Vision can not accurately label art since they were trained with stock image photography. To improve the accuracy, we are building our own model, which is challenging due to the lack of training data and diversity of art objects (paintings, sculptures, textiles, etc). To augment our data, we plan to use synthetic data, GANs, and style transfer techniques. Our project's source code is now publicly accessible, encouraging interdisciplinary exploration of Mead's collection by students and faculty.

Poster 5009

Soldier Activity Recognition: Are they "Rucking", riding, or running?

Andrew Zhang
Department of EECS
United States Military Academy, West Point, NY

Faculty Advisor
Maria R. Ebling
United States Military Academy, West Point, NY

We produced an activity recognition model that can identify a soldier's activity based on sensor acceleration data. The model can recognize nine activities, including three soldier-specific movements and six everyday movements.

Although activity recognition is not novel, our application to soldier activities provides a new utility to advance training for the military. For years, the United States military used the Multiple Integrated Laser Engagement System to allow units to train with more realistic force-on-force contact. However, this system is outdated and inconvenient, making its incorporation into training cycles difficult. A replacement system is under development, but faces challenges in differentiating between soldier activities, specifically differentiating soldiers riding in a vehicle versus running behind it. Thus, we captured accelerometer data with which we created a model, built upon prior research, to detect different soldier activities. In the future, we plan to integrate this model with the broader system to address activity recognition challenges.

We collected data from soldiers performing tasks such as "rucking" (aka backpacking), running, and riding in vehicles (including military). Throughout these activities, participants carried a sensor. We added our data to an existing non-military dataset[1] to ensure a more complete model. Then, we used a Long Short Term Memory (LSTM) network[2] to train the activity recognition model.

Our model was more than 97% accurate in recognizing nine activities, including differentiating soldiers who are riding in a military vehicle from those rucking. Our model offers an activity recognition system compatible with simple modules containing basic triaxial accelerometers. Our work presents the military with an innovative opportunity to improve its decade-old training system.

---

[1] http://storm.cis.fordham.edu/~gweiss/wisdm/

[2] L. Alawneh, B. Mohsen, M. Al-Zinati, A. Shatnawi and M. Al-Ayyoub, "A Comparison of Unidirectional and Bidirectional LSTM Networks for Human Activity Recognition," 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Austin, TX, USA, 2020, pp. 1-6.

Poster 5010

Blood Glucose Prediction for Juvenile Diabetes: Learning RNN-LSTM Models Using Real-World Datasets

Yijia Chen
Department of Computer Science
Wellesley College, Wellesley, MA

Faculty Advisor
Scott Anderson
Wellesley College, Wellesley, MA

I present my senior thesis on applying recurrent neural net (RNN) models to real-world time-series datasets for diabetes case management. The collected 5-year dataset of an individual with type 1 diabetes mainly includes the patient's blood glucose level every five minutes, the insulin boluses, and the carbohydrate intake. After reviewing the literature, I proposed a model for which we would first generate two deciding parameters for the beta curves used to model the absorption processes of insulin injection and carbohydrate intake respectively so that we could split an instantaneous insulin bolus or carb intake into a sequence of values over a certain period after that time point; then I would feed the generated insulin and carb values along with the blood glucose level to our RNN, with Long Short-Term Memory (LSTM) units, to make predictions of the blood glucose level of the specific patient over the short term. During the application process, we encountered the difficulty of a missing value rate of up to 60% in our dataset. We addressed the problem by training on a subset of the data that is relatively free of continuous missing values, which led to an investigation of how prediction accuracy declines with data loss. To better evaluate the degradation of the data loss, we reached out to get access to the OhioT1DM Dataset released in 2018 for the first Blood Glucose Level Prediction (BGLP) Challenge, which is the first publicly available dataset for type 1 diabetes patients to contain continuous glucose monitoring, and each of the 12 patients in the dataset has eight weeks' worth of data included. As a result, I want to present on the efficacy of the LSTM model for predicting BG as well as the effect of missing data in the performance of LSTM.

Poster 5011

Visualizing Music Preferences: Using Physical Features to Construct Preference Models

Henry Zelenak
Computer Science
SUNY Fredonia, Fredonia, NY

Faculty Advisor
Ziya Arnavut
SUNY Fredonia, Fredonia, NY

We present our research on image feature recognition in a consumer-facing product. Music is closely tied to one's personality, as our preferences dictate what we listen to. Our preferences also dictate our style or how we express ourselves. These preferences can be expressed visually, and superficial aspects of those preferences can be discerned in images of oneself. People are often interested in the message conveyed by their appearance, as it reflects their lifestyle preferences. What if there were a website that let you upload an image of yourself daily in exchange for a “song of the day” and a brief horoscope? There is a solution to every problem, so we set out to find it.

To address this, we propose using a machine-learning (ML) object-detection model that identifies physical features in an image of a person, such as the presence of jewelry and hair color. We utilize the TensorFlow framework alongside TensorFlow’s Keras API to train the ML model to accurately identify and categorize these features, providing usable metric data that can be used to construct a preference model for song preferences or other aspects of one’s personality. The ML model is written and trained in Python and implemented through TensorFlow Lite in Swift for iOS development. The app will provide a simple yet seamless user interface, getting the image from the user and displaying it along with the results of the detected features on the screen.

Poster 5012

Internet Store: Time to Go for Shopping

Andriy Martynyshyn

Computer Information Systems
SUNY Buffalo State University, Buffalo, NY

Faculty Advisor

Sarbani Banerjee
SUNY Buffalo State University, Buffalo, NY

Do you know how much of our free time we spend visiting stores? According to statistical data, it is a crazy number, which could be reduced with the help of Internet Stores, so that people will have more time to enjoy or do other things. This research project is focused on designing an Internet Store. It will consist of three parts: database, server, and client. Additionally, the project will build a web server with Node.js, NPM, and React, a JavaScript website client as well as PostgreSQL database. The server will be communicating with the database and the client, where errors and page transitions will be detected. The database will store the necessary information about user and products with its characteristics and images in the Internet store. The user or shopper will need to create an account on the website. The administrator can create different products, types, and brands in this Internet store.

Node.js is selected because it is a cross-platform, open-source server environment that can run on Windows, Linux, Unix, macOS, and more. NPM is a package manager for the JavaScript runtime environment of Node.js. React is a free and open-source front-end JavaScript library for building user interfaces. PostgreSQL, also known as Postgres, is an open-source relational database management system (RDBMS) that will be used in this project, emphasizing extensibility and SQL compliance. The outcome of this project is an active online store where one can create a shopper account, view products with their characteristics and images.

Poster 5013

Walking the Rover

Eric Barton

Computer Information Systems
SUNY Buffalo State University, Buffalo, NY

Faculty Advisors

Gang Hu, Sarbani Banerjee
SUNY Buffalo State University, Buffalo, NY

This research project explores the use of a camera-based laser pattern navigation system for autonomous rover navigation in unfamiliar environments. Traditionally, GPS and inertial navigation have been utilized as driving guidance for rovers, but these methods can be limited by unreliable or unavailable GPS signals and the accumulation of errors in inertial navigation. In contrast, the camera-based laser pattern navigation system is versatile and unaffected by such limitations.

The project is written in Python and uses packages such as OpenCV, NumPy, and Pyfirmata. There will be a remote connection, linking the raspberry pi on the rover and the server that will deal with all the computations. Image processing techniques are then applied to analyze the images and determine the rover's position and orientation relative to the terrain, thus providing directional guidance.

This project specifically investigates a camera-based laser tracking system that is controlled by a user or other controller pointing a laser pointer at the ground/floor. The images captured by the camera mounted on the rover are processed to generate an instructional interface, which guides the rover's movements. However, this system presents challenges such as increased processing power requirements and susceptibility to changes in lighting conditions or shadows. Proper calibration is also necessary to ensure accuracy.

The objective of this project is to develop a real-time camera-based laser pattern navigation system that can serve as a valuable tool for autonomous rover navigation, providing visual directions for precise control of the rover's movement.

Poster 5014

Port is Open!

Farah Al-Karkhi
Computer Information Systems
SUNY Buffalo State University, Buffalo, NY

Faculty Advisor
Sarbani Banerjee
SUNY Buffalo State University, Buffalo, NY

Network security is essential for any community as it helps to protect against cyber threats, safeguard critical infrastructure, maintain trust and credibility, and educate and raise awareness about cybersecurity. The goal of this research project is to create a program that scans a network for open ports, target IP addresses, services, and potential vulnerabilities. Python programming will be used to provide a wide variety of packages for network analysis.

This project will develop the ability to create a single report generated from the scanning of ports and DDOS (Distributed Denial of Service). A DDOS program sends many requests to a target server, while a port scanner tool scans a range of network ports on a remote host to determine which ports are open and accepting connections. There are few tools that will be used to set up this project, such as Nmap, Scapy, IPy, Sys, and socket library, which can be used to create a network scanner that can scan a range of ports on a given IP address, also creating, connecting, and binding sockets for communication.

Once completed, this project is expected to scan a network or a host to detect open ports that may be vulnerable to attacks. By identifying open ports, one can identify potential security weaknesses and take appropriate measures to mitigate them. This will provide practical experience in developing and implementing network security measures, critical thinking as well as problem-solving skills.

Poster 5015

UNICON Server Secure Socket Layer Encryption

Chase Leckinger, Ajay Bhattacharya
Department of Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisor
Zakariya Qawaqneh
SUNY Brockport, Brockport, NY

Hypertext Transfer Protocol Secure, is an extension of HTTP that allows web application communications to be encrypted over a secure socket layer (SSL). This involves the server of the web application storing a certificate, which is sent to the browser and validated on the client side before an encrypted session is established. The server will also store a public and private key, which are used to encrypt and decrypt the data that is being communicated between the server and the client. The goal of this project was to build on the existing HTTP server for the Unicon programming language and use it to build an HTTPS server.

To build a working HTTPS server in Unicon, we first had to modify the code for the language itself. We modified the open function by adding a new option to the mode argument. When this option is read, the open function will look for additional arguments containing file paths to an SSL certificate and a private key, and use them to initialize a new SSL context with functions from the OpenSSL library. We also modified several other functions so that we could check that an SSL context was ready to receive data, access the file descriptor of the SSL context, and close the SSL context properly when the connection is terminated.

Next, we had to modify the HTTP server to open the network connection on an SSL socket, and pass in the certificate and private key. We also created a shell script to automatically generate these files.

The results of this project showed that we were able to successfully establish an HTTPS connection. We were able to run the server on localhost with the new version of the Unicon language and access the contents of the web application from the browser.

Poster 5016

Sink and Source: Understanding the Impact of Urban Heat Island Effect on Campus using IoT and Heat Mapping Methods

Caprice Williams, John Burns, Hephzibah Rajan
School of Computing and Engineering
Quinnipiac University, Hamden, CT

Faculty Advisor
Chetan Jaiswal
Quinnipiac University, Hamden, CT

The impacts of global warming and climate change are being felt around the world, from more frequent and severe weather events to ecosystem disruptions. At Quinnipiac University, our team strives to understand the effects of the Urban Heat Island (UHI), a major factor in climate change, on campus. Our project, *Sink and Source,* consists of three layers: Data Collection, Data Processing, and Data Display.

Data collection involves the use of IoT Devices that have been strategically placed in 8 locations around Quinnipiac's Mount Carmel's campus. ESP32 devices are connected to DS18B20 sensors to gather temperature information which is stored in Google's Firebase Realtime Database. The stored data is then classified as a heat sink or heat source using the Getis-Ord Gi* statistic in R Studio. Getis-Ord Gi* is a hotspot analysis tool used in datasets to understand clustering of a parameter, in our case, temperature. The results are displayed as z-values and p-values which help identify the spatial clustering of high and low heat by comparing the temperature of an area to its neighboring area's temperature. Finally, a publicly accessible website will display the data in a user-friendly manner with the assistance of Angular.js in the front-end.

Our team will present the finalized data to Quinnipiac University's sustainability committee in an effort to assist with future infrastructure plans and create a more eco-conscious campus. The utilization of machine learning can expand the dataset and enable weather prediction on campus in the future. As the world becomes more urbanized, the impacts of UHI are becoming increasingly apparent, with implications for human health, energy consumption, and environmental quality. Understanding the effects of UHI is crucial for policymakers, urban planners, and public health officials as they work to mitigate its impacts and we hope to start this conversation at a university-level.

Poster 5017

Evaluating the Efficacy of Worked Solution Videos in a Computer Systems Course

Grace Kim, Dylan Green
Department of Electrical Engineering and Computer Science
United States Military Academy, West Point, NY

Faculty Advisor
Suzanne Matthews
United States Military Academy, West Point, NY

Worked solutions are an educational tool that have been tested in a variety of science based courses, including introductory computer science classes, in which they proved to be an effective learning instrument. The increased reliance on online resources after the COVID-19 pandemic suggests that there is great potential for worked solutions *videos* to help computing students. This study explores how worked solutions, shown through detailed, step-by-step videos, aid student learning in an intermediate-level computer science course, namely computer systems.

In prior semesters of the computer systems course, solutions to in-class exercises were available as PDFs; the authors created 40 worked solution videos of a subset of these in-class exercises that show an in-depth explanation and progression of each solution. To guage the overall helpfulness of the worked solution videos, we gave students currently enrolled in the class a survey that asked them how helpful they found the worked solution videos compared to other course resources. We also compared the quiz performance of students with access to the worked solution videos to those enrolled in semesters that did not, in order to study the impact of worked solution videos on students' grades in the course.

Preliminary results suggest work solution videos are an effective tool for intermediate computer science courses. Over sixty percent of students in the course have consistently watched all or most of the worked solution videos corresponding to each lesson. All of those students shared that the worked solution videos were very helpful and ranked it as the top study aid for succeeding on quizzes. Additionally, students in the section with worked solution videos earned higher average scores (88.7%, 90.8%, 92.7%) on quizzes compared to students without worked solution videos (88.1%, 81.7%, 86.9%). Data collection is ongoing, and we expect this trend to continue.

Poster 5018

The Cyber Domain

Christian Tanguay

Computing & Data Science
Wentworth Institute of Technology, Boston, MA

Faculty Advisors

Frederick Trilling, Hollis Greenberg
Wentworth Institute of Technology, Boston, MA

The United States' current Cybersecurity Defense model of Cyber Deterrence on a practical, realistic level does not work. The current model is designed around altering the risk/reward calculus of hackers so that they are dissuaded from hacking United States persons and entities; this model is explicitly aimed at enforcing and securing the United States' Cyber Operational Domain. The Cyber Domain is defined here as the amalgamation of all things under the United States' control that is directly or adjacent/dependent on the computer networks maintained and secured by the United States. The problem with this model can be summed up in one key question "If Cyber Deterrence is the method used to defend the United States' cyber domain, *why does it still get hacked?*" It would be naïve to assume that we can completely prevent hacking, but the system in place is ineffective for a variety of reasons and thus is ripe for replacement. To understand these issues, research was done in multiple disciplines, including cybersecurity, computer science, and sociology. Through my research of the laws, policies, cases, and people that guide the current model for hacking in and against the United States, we can see how those who hack and those who are hacked can circumvent the standing model, and begin to tell why those who hack are not successfully deterred. I have identified core reasons why hackers are not deterred, by what means they circumvent deterrence, and what is *next* for the Cyber Domain of the United States. The goal of this research is to engage in the conversation on how to find a better way to prevent hacking against the United States, with or without deterrence.

Poster 5019

Unveiling The Ambiguity In Acronym Definitions Using Natural Language Processing

Izhar Ali, Million Haileyesus, Dominick Profico
Computer Science Department
Rowan University, Glassboro, NJ

Faculty Advisor
Vasil Hnatyshin
Rowan University, Glassboro, NJ

In this study, we introduce a new method for identifying and extracting acronym-expansion pairs in digital texts. Abbreviations such as initialisms and acronyms are commonly used in digital texts, which can create difficulties for text mining and natural language processing because of their ambiguity and complexity. Our approach combines custom-made regular expressions and web scraping - implemented in Python - to identify all forms of abbreviations in the text. The regular expressions are carefully crafted to capture different types of acronyms, which are then matched with the text to determine possible definitions. If a definition is not found, the web scraping module is used to obtain the expansion of the acronyms from a pre-determined set of sources. Proposed approach ensures that the algorithm is intuitive and does not require extensive domain-specific knowledge. We compare the algorithm's performance to existing state-of-the-art approaches.

We conducted a study of proposed acronym extraction algorithm using a large set of scientific papers from across different domains. The study identified a number of challenges that readers face when authors use domain-specific acronyms, including the following:

1. Acronyms are not explained in the text
2. Identical acronyms carry different meanings
3. Certain fields of study exhibit an excessive use of acronyms

The insight gained from the study allowed us to develop a practical text mining strategy for identifying acronyms in scientific papers, which is applicable to real-world problems. Proposed solution enhances the efficiency and accuracy of text mining and NLP systems by providing a reliable method for identifying and expanding acronyms.

Poster 5020

Neural Networks Implemented through Bio- Nanorobots for the Early Detection and Elimination of Cancer Cells in the Human Body

Vincenzo D'Aria
Department of Computer Science
Central Connecticut State University, New Britain, CT

Faculty Advisor
Mahdieh Zabihimayvan
Central Connecticut State University, New Britain, CT

I present my experience participating in the CCSU Honors College in conjunction with the Computer Science Department at Central Connecticut State University. A requirement for the honors college is publishing an undergraduate research thesis in a topic of my choice, in which I am researching the early detection and elimination of cancer cells via neural networks and bio-nanorobotics. My thesis looks at both the healthcare and computer science fields by combining cutting-edge neural network and nanorobotics research with the ever-growing demand for early-stage cancer detection and elimination methods. Currently, I am gathering and qualifying data from researchers who have experimented with convolutional neural networks, graph neural networks, and bio-nanorobots in applications such as early-stage cancer cell detection, diabetes detection, age-reversing treatments, and targeted drug delivery. Although results have been preliminary in these experiments due to the research base being relatively new, my thesis explores new ways in which neural networks and bio-nanorobotics can be used to traverse the body and more effectively detect and eliminate cancer cells compared to traditional methods. I have been collecting data from reputable databases, primary sources, journals, textbooks, and conference publishings, such as IEEE, ACM, Pearson and Springer. I review these sources and any data and/or mathematical algorithms with my advisor for my understanding while applying them to my research. My results and findings are currently in-progress, as I am still drafting my thesis publication; however, I expect to present implementing Relation Structure-aware Heterogeneous Graph Neural Networks (RSHN) for more efficient mapping and traversal of the body, convolutional neural network computer vision for more efficient cancer cell detection, and the ethical considerations of these processes.

Poster 5021

PorchfestPro App

Keegan Palmo, Laci Young, Emmett Barry

Computer Science
Ithaca College, Ithaca, NY

Faculty Advisor

John Barr
Ithaca College, Ithaca, NY

In this report, we will discuss our experience in developing the PorchfestPro App, an innovative tool designed to help individuals navigate and enjoy Porchfest events. Our goal in creating this app is to provide a user-friendly interface that makes it easier for people to find and locate bands during Porchfest events, while also promoting community engagement and accessibility. Our aim was to create an app that would appeal to a broad range of individuals and make the Porchfest experience more enjoyable and accessible for everyone. To start, we researched various programming environments that would be suitable for creating a user-friendly app with robust functionality. After careful consideration, we decided to use a combination of React Native and Firebase for our app development. These platforms offered the scalability and flexibility that we needed to ensure that our app could accommodate a large number of users and provide real-time updates on band locations and schedules. Throughout the development process we encountered several challenges, including designing a user-friendly interface, integrating real-time data, and ensuring that the app was compatible with a range of mobile devices. However, with careful planning, collaboration, and iterative testing, we were able to overcome these obstacles and create an app that met our goals and expectations. Our poster presentation will provide a detailed overview of the PorchfestPro App, including its key features, user interface, and real-time data integration. We will also discuss the challenges we encountered during the development process and share our insights and recommendations for other teams interested in creating similar apps. Ultimately, we believe that the PorchfestPro App has the potential to revolutionize the way that people navigate and enjoy Porchfest events, and we are excited to share our work with the wider community. The app will launch in September and will be used in Ithaca's Porchfest event, so we look forward to getting initial user feedback at that time.

Poster 5022

Junkademy: An Interactive Educational Web Application for the Java Programming Language

James Gonsalves, John Theall, MacArthur Mills
Computer Science
Quinnipiac University, Hamden, CT

Faculty Advisor
Chetan Jaiswal
Quinnipiac University, Hamden, CT

With programming becoming an ever-growing field in both scope and demand, the amount of readily available online learning tools is skyrocketing. One fairly popular style of application that is designed to teach users how to program includes the use of code-blocks – predefined code snippets contained in a very high-level description that the user can interact with and assemble in various ways to achieve programmatic results. A crucial drawback of most code-block based applications is that they rarely demonstrate how an assembled arrangement of blocks translates to code belonging to any modern programming language, which drastically limits the applicability of learned knowledge from these applications. Our application prototype, *Junkademy*, aims to bridge this gap by explaining high-level code blocks with relevant examples and user-friendly explanations. Future versions of *Junkademy* will translate user-assembled code blocks into syntactically correct Java code in real-time. This principal feature will peel back a layer of abstraction found in similar educational applications that will benefit the user as they attempt to start writing code because they will have been familiarized with one of the most popular and widespread programming languages worldwide.

This free web-based application is being developed using the Java Spring Boot framework for backend development and object-relational mapping logic, and the React JS framework for frontend structure and development. Furthermore, *Junkademy* has an intuitive user interface that is easy to use which goes hand-in-hand with the high-level code blocks that will allow users to learn using our application regardless of preliminary experience. Our website promotes any user to develop fundamental programming, critical thinking, and problem-solving skills that will provide them with an essential basis for software development.

Poster 5023

Image Compression

George Stark
Computer Science
SUNY Fredonia, Fredonia, NY

Faculty Advisor
Ziya Arnavut
SUNY Fredonia, Fredonia, NY

I present my experience in image compression from my class in Image Processing & Computer Vision in the Computer Science Department at SUNY Fredonia University. This class emphasized the theoretical and practical uses of image computing. In the realm of computer science, compression is an important component of image processing and information as a whole. When storing or transmitting data, the amount of space that can be saved allows for better utilization of system resources. To give an example: if we have a raw set of data that is using the entirety of its storage space, compressing it to 50% of its original size would effectively double the storage space. My goal was to show that a simple lossless algorithm can provide insight into the process of image compression. The technique starts with a decorrelation sequence where the pixel information is translated into a less entropic matrix mirroring the original image; it is then received by an arithmetic encoder that transforms the matrix into a single decimal value that represents the input. Going in reverse we can piece the image back together without losing any data. This process will typically result in around a 50% compression rate. My poster will illustrate this technique to show viewers this image compression visually.

Poster 5024

PostgreSQL vs. MongoDB in the Music Industry

Amrita Acharya, Ashley Jagai, Olivia Sorensen
Department of Computer Science
Smith College, Northampton, MA

Faculty Advisors
Razieh Fathi
Smith College, Northampton, MA

We present our project for our Intro to Database Structures course in which we compare the use of two database systems, PostgreSQL and MongoDB, in storing and analyzing data on musical tracks and artists. With the rise of digital music distribution, data on songs, artists, and streaming statistics have never been more accessible. Streaming services like Spotify provide listening data that, when analyzed, can evaluate an artist's success or reveal trends in the current industry. Choosing the best database system for these analyses is an important question and one we sought to answer. We chose two representative database systems, PostgreSQL and MongoDB, to compare; notably, we wanted to compare a relational database system (PostgreSQL) with a non-relational database system (MongoDB) due to their vastly different ways of storing data. Our process started by researching the two databases. Using Spotify's Web API, we developed Python scripts to collect data on a subset of Spotify's artists and tracks. After designing a schema for our database, we imported our data and compared the performance of equivalent queries that we believed would return useful information for the music industry. We conclude from our research and results that a quick, flexible database system like MongoDB is more suited for storing and analyzing the large amounts of complex data one might use in the music industry.

Poster 5025

Database Efficiency in Dynamic Generation of Map Graphs

Olivia Sorensen
Department of Computer Science
Smith College, Northampton, MA

Faculty Advisor
Razieh Fathi
Smith College, Northampton, MA

I present a comparison of two database systems (PostgreSQL and Neo4j) in their ability to store, modify, and retrieve highway data. The Travel Mapping project allows users to mark the highways they have traveled and compare them with other travelers. The geographical data representing travel statistics for each route and hierarchical relationships of regions and highway systems are stored on a frequently-updated open-source GitHub repository. This data is both loaded into the MySQL database that runs the website and used to create text-based graph files for use in the Map-based Educational Tools for Algorithm Learning (METAL) project. Currently, creating these graph files takes up a significant amount of time in the daily site update process; therefore, the Travel Mapping project is looking to transition to on-demand generation of graphs from the database, rather than pre-processing every graph file. A new database system may prove essential for this transition. I compare the efficiency of PostgreSQL and Neo4j for this process, explaining how the inherent differences in data structure affect the performance of the task. Prototype databases are provided for both database systems. In addition, I consider how a transition to either database system would impact the performance of both the Travel Mapping and METAL sites, as well as any challenges integrating the new system may pose.

Poster 5026

Improved Automated Analysis of Mass Spectra for Proteomics Experiments

Dominick Profico, Samuel Costantino
Computer Science Department
Rowan University, Glassboro, NJ

Faculty Advisors
Vasil Hnatyshin
Rowan University, Glassboro, NJ

We present improvements to the algorithm for automated annotation of mass spectra for protein coverage experiments. Liquid chromatography - mass spectrometry (LC-MS) experiments allow scientists to get a glimpse into the structure of a protein, but analyzing the collected spectra requires a large amount of combinatorial computations and robust signal processing. Protein coverge analysis includes modeling the protein digestion (spliting protein sequence into peptides), signal processing, and then searching through thousands of collected spectra to match mass-charge ratio lists with the list of sequences obtained via in-silico digestion. The experimantal data typically consists of thousands of spectra per run with several thousands of data points in each spectrum. That is why the algorithm for comparison of experimental data with the list of hundreds of peptides is computationally expensive. We designed and implemented our improved algorithm for automated analysis of LC-MS data in Python. Leveraging the ability of NumPy library to efficiently perform calculations and Python's native multiprocessing features, we were able to significant improve the effiency and speed of computations. Our implementation performs spectra annotations and applies a scoring function to determine the most likely spectrum that matches to the chosen peptide sequence. We developed a custom plot to visualize the spectrum annotation results which allows the user to confirm and compare the matches between peptide sequences and different spectra. We benchmarked our implementation of automated analysis against other similar proprietary mass spectrometry software and open-source projects. Our product offers a transparent solution which improves computational time and offers visualization of results, including spectra annotation.

Poster 5027

Improving Graph Neural Network-based Fake News Detection with Data-centric AI

Austin Barr
Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisors
Adita Kulkarni
SUNY Brockport, Brockport, NY

Jerome Dinal Herath
Obsidian Security, CA, USA

Social media has gained immense popularity in recent years as it enables users to share their opinions with millions of people. It also enables the wide dissemination of fake news which can have significant effects on real-world events affecting individuals and society. Thus, it is important for the social media platforms to timely detect fake news and prevent its propagation.

In this work, I aim to use the publicly available dataset that includes fake and real news propagation networks on Twitter built from fact-checking websites PolitiFact and GossipCop using 20 million tweets. I plan to use news content and historical posts of users to build a tree-structured propagation graph for each news article based on its sharing patterns on social media. The news post represents the root node and the users who share this news represent the child nodes. Using the vector representations of news and users as node features, I aim to apply Graph Neural Networks to learn a joint user engagement embedding. This user engagement embedding and the news textual embedding will be used to train classifiers to detect fake news.

Once I have a propagation graph and the graph-based models ready, I plan to use Data-centric AI to further improve the accuracy of detecting fake news. Data-centric AI is a novel discipline representing the transition from modeling to systematically engineering the underlying data used to train models. Since significant progress has been made in improving the models, Data-centric AI focuses on improving the quality of the data that is fed to these models.

Poster 5028

Night Life - A Smart Thermal Car Camera for Enhanced Vehicle Safety

Issah Abakar, Scott Van Alstine, Mackenzie Hiddie, Nick Farrell
Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisor
Eric Owusu
SUNY Brockport, Brockport, NY

Night driving can be dangerous due to poor visibility and lighting conditions [1] (* references removed for space reasons *). To address this issue, a smart thermal car camera called Night Life application has been proposed to help drivers detect road hazards in total darkness. The device uses thermal imaging to detect heat-emitting objects, such as vehicles, people, and animals, and alerts the user accordingly. Additionally, it functions as a regular dash camera, recording high-resolution videos that can be accessed through the Night Life application.

Over 36,000 fatalities in motor vehicle crashes in 2019 Vehicle safety raises a significant concern that needs to be addressed. Night driving poses additional challenges due to reduced visibility, increasing the risk of accidents and inability to detect potential hazards in the dark. Night Life targets commuters who drive long distances, particularly those in rural areas, by improving vehicle safety and providing enhanced visibility regardless of light or weather conditions [2].

The Night Life camera was developed based on the software development life cycle (SDLC) model with a focus on agile development practices. This process includes requirements gathering, design, implementation, testing, and maintenance. The camera's technical requirements and necessary features were identified in the requirements gathering phase, and hardware and software components were selected in the design phase. A user interface was designed at this design phase. In the implementation phase, the Night life application prototype was designed, and user feedback and evaluation were analyzed to modify the design. To ensure accuracy and reliability, the camera was tested in various simulated scenarios during the testing stage, such as different lighting, weather, and types of potential hazards, including pedestrians, animals, and other vehicles.

Using the SDLC model and agile approach of iteration for the conceptualization phase, the Night Life smart thermal car camera has shown significant and promising results. Positive feedback was obtained, suggesting that Night Life could enhance vehicle safety during night-time driving, particularly for long-distance commuters in rural areas. The system's ability to detect potential hazards on the road and alert drivers in total darkness was particularly noteworthy, while the regular dash camera feature was an effective tool for recording high-resolution videos.

Poster 5029

Partitioning and Efficiency Improvements for METAL Algorithm Visualizations

Michael Plekan
Computer Science
Siena College, Loudonville, NY

Faculty Advisor
James D. Teresco
Siena College, Loudonville, NY

I present my enhancements to the Map-based Educational Tools for Algorithm Learning (METAL) Project in the Computer Science Department at Siena College. The goal of the project is to help computer science students learn and understand graph algorithms using map based graphs. Students being able to see what the algorithm is doing can help with learning. During my time working on the project we designed a partitioning algorithm visualization, partitioning statistics support, URL options, support for an updated file type, improved efficiency, and an algorithm to simplify graphs. I started with the algorithm to simplify the graphs. The goal of this was to improve efficiency of large graphs. After that, I worked on adding the new file type which is meant to allow more flexibility to students importing their own data. I then started the Recursive Coordinate Bisection algorithm visualization. The goal for this was to start to build the parallel processing side of the project. Next, I worked on making a system to automatically calculate and present partitioning statistics. This is meant to help show how different partitioning algorithms perform and make adding new partitioning algorithm visualization easier. I then worked on improving efficiency of the graph loading system by removing unnecessary arrays and generating pop-up text on the fly instead of pre-loading it. Lastly, I added support for URL options for automatically picking an algorithm visualization and options pertaining to that visualization. The goal of this is to make it easier for professors to set up examples for students to use. The poster will discuss an overview of what I worked on, and the challenges I encountered during the project. The full description and implementation features of the game will be presented in detail.

Poster 5030

Developing an Interactive Simulation for Non-Inertial Reference Frames

Ted Karanja Mburu
Physics & Astronomy
Ithaca College, Ithaca, NY

Faculty Advisor
Colleen L. Countryman
Ithaca College, Ithaca, NY

We are developing an interactive and freely accessible simulation that aims to improve physics students' understanding of challenging classical mechanics concepts. The simulation is being developed using JavaScript with the p5.js visualization library. The simulation is freely accessible and runs on most devices with a browser.

The simulation is split up into different levels that visualize certain physical situations from different reference frames in order of increasing complexity. After experimenting with built-in camera settings, the motion within each reference frame was more efficiently calculated using classical mechanics concepts. These reference frames are split between different HTML5 canvas elements that run concurrently. Each canvas uses Euler's method to animate objects *in situ* by calculating their next steps using the appropriate physics equations.

The motion of non-inertial reference frames and the forces (Coriolis and centrifugal) involved in understanding them are challenging concepts for introductory and advanced mechanics students. Furthermore, the prediction of trajectories of an object due to these fictitious forces is another challenge for students. We will discuss the design, implementation, and planned assessment of the simulation.

Poster 5031

UX Design – Siena College's Programming Contest

Amani Walker, Jeremy Bult, Jason Macutek
Psychology
Siena College, Loudonville, NY

Faculty Advisor
Jami Cotler
Siena College, Loudonville, NY

This study utilized the UX design process to create a website for Siena College's annual programming contest. The purpose of this study is to develop a platform that enhances the organization and efficiency of the event The website will help streamline communication and provide essential information to all users, resulting in a seamless and successful experience for all.

Previously, CS faculty relied on various methods, such as spreadsheets, to manage information about volunteers, high school participants, and food vendors for the annual programming contest. This approach led to information overload and resulted in numerous inefficiencies. It was challenging to access some of the essential information, such as volunteer roles, contact details, and locations, leading to potential delays and miscommunication.

As a part of our research design, our student-volunteers and CS faculty were interviewed to gain a deep understanding of their experiences and pain points. We analyzed statements made by our users and were able to identify the factors that would lead to higher efficiency while eliminating unnecessary features. Based on our findings, we created a problem statement that puts the primary frustrations into an overarching problem. To explore design solutions, sketches and low-fidelity wireframes were produced utilizing a software called Balsamiq. This step enabled us to experiment with various design elements and layout options before committing to high-fidelity mockups, which closely resemble the final website, in a program called Figma.

As the outcome of this study, our team will create a website to transform the way Siena College's CS department manages its annual programming event.

Poster 5032

UX Research – Inclusivity Study

Amani Walker
Psychology
Siena College, Loudonville, NY

Faculty Advisor
Jami Cotler
Siena College, Loudonville, NY

This project focuses on the significance of empathy in the development of applications and technology. This study examines the ways in which empathy can be taught and fostered in computer science students, as well as how to ensure that empathetic design principles are integrated into the applications they create.

In the process of creating applications, it is crucial to consider the diverse range of individuals who may utilize the product. Specifically, it is important to ensure that computer science students are constantly thinking of individuals with different disabilities and their specific needs, while designing and developing technology. As a result, our study investigated how to encourage and enable computer science students to prioritize accessibility and inclusivity in their products.

The empathy lab was designed to utilize various techniques to foster an empathetic approach among students. Firstly, the students were introduced to alt tags and their importance in accessibility for individuals with visual impairments. Secondly, the students underwent a simulation of both physical and visual impairments. Thirdly, the lab covered the critical role of color in design, specifically color theory, contrast ratios, and blindness. Lastly, students were tasked with attempting translations to gain a better understanding of the user's language, further promoting inclusivity in their design approach.

The empathy lab provided students with a comprehensive understanding of accessibility and empathy in technology design as well as the ability to keep all kinds of users in mind when developing.

Poster 5033

How Technology is Used to Track Climate Change: Earth Observational Satellites

Sophia Sjogren

Department of Computer Science
Kalamazoo College, Kalamazoo, MI

Faculty Advisors

Pam Cutter, Sandino Vargas-Pérez
Kalamazoo College, Kalamazoo, MI

Understanding the whole picture of the climatological state is how decision-makers and the general public can best respond to climate change. Since 1958, satellites have been launched into space to connect and observe the world. I had the opportunity to participate in NASA's Applied Remote Sensing Education and Training program. These courses taught me how satellite data is collected, processed, and applied to environmental studies and education. Inspired by the training, I created an individual research project focused on how remote sensors track and monitor the growth of harmful algal blooms in Lake Erie. The Great Lakes are a crucial source of fresh water, and pollution from algal blooms affects the ecosystem and surrounding populations. My research involved identifying where phosphorus from commercial farms has runoff access to the lake because this creates optimal living conditions for the algae. As well as the spectral signature of Chlorophyll-a indicates the density and health of the algal blooms. This satellite data, alongside climate models, gives better reasoning and a more coherent account of the recent outbreaks in Lake Erie. Policymakers can use these findings to regulate commercial farming in the area and make educated decisions regarding the lake. Response teams can replicate the data and conditions to predict where future blooms may form. During my research, I learned how satellite data is also applied to monitoring manufacturers' carbon emissions, responding to human and natural disasters, educating the general public with climate models, and many other areas. In recent years, artificial intelligence has been working with remote sensors leading to the exploitation of natural resources. However, if applied ethically, newer satellite technology alongside artificial intelligence can transform our understanding of climate change.

Poster 5034

Hyperspectral Imaging Application

Edward Wynman
School of Computing
Montclair State University, Montclair, NJ

Faculty Advisor
Stefan Robila
Montclair State University, Montclair, NJ

Hyperspectral images (HSI) are collections of grayscale images of the same scene collected over narrow light wavelength intervals. HSI's have been widely used in a variety of fields, be it earth sciences, manufacturing, agriculture, food safety, or defense and homeland security. The growing number of sensing platforms provides a unique opportunity for hyperspectral data to be easily accessible (and usable) not only by scientists and practitioners but also by the broader public in a similar fashion as, say Google Earth, potentially opening unique opportunities for hyperspectral citizen science. Yet, publicly available comprehensive sets of data continue to be limited, and full workflow interfaces are missing. In preliminary work a prototype desktop application for downloading, viewing, and processing hyperspectral data was developed. The application, built using Python, allows an user to browse their hyperspectral image library, examine the spectra of any pixel in the image, assign pixels to material classes, and perform Principal Component Analysis (PCA) and automatic spectra classification using a spectral angle mapping. It also allows the user to create custom material classes of their own and examine the spectra of pixels assigned to each class. The data ingestion provided by an USGS Search module provides an easy search engine to query the United States Geological Survey's vast repository of Hyperion hyperspectral images and seamlessly download them for analysis in the application. We are now re-writing the application to read an image and perform a much more complex set of operations with it.. We have also been able to connect to the USGS repository of Hyperion hyperspectral images to download them and perform the aforementioned analysis. By developing better workflows and interfaces for hyperspectral data, we can help make this technology more accessible and usable for a wider range of scientists, practitioners, and even members of the general public.

Poster 5035

MindBellows: Fan Your Mental Flames!

Michael Tambascio, Loki Simeon
School of Computing and Engineering
Quinnipiac University, Hamden, CT

Faculty Advisor
Chetan Jaiswal
Quinnipiac University, Hamden, CT

MindBellows is an Android app that uses Java, Android Studio, GitHub and SQL to test and improve the mental aptitude of its users. The app tests the user in 5 different categories, those being general IQ questions, Memory, Logical Thinking, Pattern Matching, and Reaction Time. The user can choose to take a test with all of the categories, known as a full-spectrum test, or if they want to particularly hone one skill, they can choose to take a test focused on only one category. Each category is scored out of 20 points, a full-spectrum test is scored out of 100 and impacts the scoring of all five categories. Based on how well the user does, the app gives the user questions on an assigned difficulty scale. The difficulty scale is Novice (0-5 points), Intermediate (6-10 points), Advanced (11-15 points), Expert (16-20 points). A full-spectrum test will scale its difficulty based on the average of the user's five scores in the categories. The app will also try to encourage the user to try questions that are outside their difficulty range, allowing for gradual improvement the more times they use *MindBellows*.

Our target audience is people within the 25-35 and 55-65 age ranges, as these are when brain activity tends to decay - for the former, because people recently graduated from college and put their minds into the workforce, and for the latter, due to the natural progression of the human mind. Studies have also shown that these age groups are most receptive to the effects of brain games, for mostly similar reasons. Studies have also shown that similar apps can decrease the mental effects of aging or degenerative mental diseases, like Alzhiemer's. Our client is the Quinnipiac Psychology Department and the development process is a biweekly scrum sprint using the agile methodology.

Poster 5036

Explainable AI and Data Optimization

Art Marciano

Department of Computer Science
Southern Connecticut State University, New Haven, CT

Faculty Advisor

Imad Antonios
Southern Connecticut State University, New Haven, CT

In machine learning it is generally believed that more data helps models perform more accurately. Larger datasets, however, might potentially present issues with models learning irrelevant patterns or noise that might degrade their performance. Since for many machine learning algorithms the internal logic acts as a black box, it is often impossible to understand what the models might be learning. Explainable AI (XAI) helps to address these problems and allows for management and oversight in the event of undesirable results, such as biased decision making. XAI is typically used to explain a model's prediction and potential biases. XAI examines the dataset and generates explanation of what the model is learning from the dataset. While XAI's primary goal is to shed light on a machine learning model, it can be adapted to produce importance measures of individual data items relative to a model's performance. The aim of this study is twofold: First, we explore the effectiveness of XAI using two techniques, LIME (short for Local Interpretable Model-agnostic Explanation), and ts-keras-vis as methods for visualizing predictions in image data. Second, we demonstrate the joint use of the XAI method SHAP (short for SHapley Additive exPlanation) and the dimensionality reduction technique RFE as a method for reducing data size. Through a process of ranking and pruning features and samples based on their contribution to the model's performance, our results demonstrate that in some cases this approach can maintain, or sometimes improve, model performance using a small fraction of the dataset. In one dataset alone with 506 samples the methods used improved the model's performance by pruning up to 71% of the data. In another dataset for diabetes with 768 samples the model's performance greatly improved by pruning up to 44% of the data.

Poster 5037

Machine Learning in Prosody and Meaning

Shirley Fong, Sofia Hirschmann, Madeleine Guettler
Math, Statistics and Computer Science
Simmons University, Boston, MA

Faculty Advisors
Nanette Veilleux, Alejna Brugos
Simmons University, Boston, MA

This work is part of an ongoing research project funded by the National Science Foundation on mapping prosody to meaning. As students, we were responsible for prosodically annotating the raw speech data, tuning and modeling k-means clusters to find groups of pitch accents. The pitch accent typing serves as a pre-processing step to a random forest model to categorize the meaning of the spoken utterance with respect to mirativity (surprised vs. neutral speech).

Prosody is the aspect of speech that is implemented by alterations in pitch (low/high), duration (fast/slower) and intensity (volume). Prosody is used to mark particular words or chunk running speech into smaller phrases. Furthermore, prosody in spoken language is widely thought to convey meaning but the complexity and variety of prosodic renditions has made modeling a specific prosody-meaning mapping elusive. For this reason, it has been difficult to assert that a particular prosodic rendition is, e.g. “surprised” vs. “neutral”, compared to, e.g. that a spoken word /c-a-t/ means “cat” vs “dog”.

Recent approaches using machine learning techniques show promise in this area. In this work, 238 prosodically-labeled, utterances are used to classify speech as either exclamative or neutral. To do so, a first pass was made to cluster pitch accents types (emphasized words) into three groups. Although the type of pitch accent is thought to be a strong cue to the mirativity category of sentences, pitch accent type is difficult to reliably label, even by experts. In addition, each sentence has one or more pitch accents. In order to capture the importance of a variable number of loosely categorizable elements, clustering is performed and the percentage of accent type is used as an attribute in the final random forest classification step. Current performance in the full model (clustering + random forests) is 83 %.

Poster 5038

Development of Instructional Physics Simulations for Classical Mechanics

Mikolaj Konieczny
Department of Physics and Astronomy
Ithaca College, Ithaca, NY

Faculty Advisor
Colleen Countryman
Ithaca College, Ithaca, NY

We have developed two new physics simulations with learning objectives to address the lack of accessible and satisfactory educational tools for students. Our design process involved researching pre-existing web-based physics simulations and identifying gaps in learning devices for specific physics concepts. Our simulations are coded in JavaScript using the THREE.js library to create a 3D interactive environment with orbit controls. The animated frames are optimized for mainstream desktop internet browsers, with future testing aimed at expanding usability to mobile devices. We are currently revising one of the simulations to include an accessible, simplified, and immersive user interface for new learners. The first simulation explores the conservation of angular momentum and how angular velocity changes with factors that vary the moment of inertia caused by internal forces. The second simulation embodies crucial orbital mechanics concepts, such as Kepler's Second Law, conservation of total angular momentum and mechanical energy, and impulsive orbital maneuvers. Our long-term goal is to determine the impact of the simulations on understanding and engagement with the targeted concepts in undergraduate classrooms.

Mburu, T., & Konieczny, M. Ithaca College PER. Retrieved March 25, 2023, from https://efieldsim.ithaca.edu/

Taylor, J. R. (2005). Two-Body Central-Force Problems. In *Classical Mechanics*. University Science Books.

YouTube. (2021). *Three.js Tutorial For Absolute Beginners*. *YouTube*. Retrieved March 25, 2023, from https://www.youtube.com/watch?v=xJAfLdUgdc4.

Poster 5039

Introduction to AVL Tree through Visualization

Xingyu Dong, Anarsaikhan Tuvshinjargal, Lizbeth Zarate-Hernandez, Ryan Oet
Department of Computer Science
Swarthmore College, Swarthmore, PA

Faculty Advisor
Michael Wehar
Swarthmore College, Swarthmore, PA

We present our experience working on a group programming project in the Software Engineering Course at Swarthmore College. The goal of this project is to teach students about data structures by providing interactive visualizations along with detailed explanations. Specifically, for AVL Trees, we observed that the existing visualization websites either have complex and unengaging user interfaces or lack explanations of each action and how the data structure works in general. To make the learning experience both efficient and effective, we believe that educational websites should explain the materials from different perspectives by providing multiple options and giving students the freedom to choose their preferred learning method. Therefore, we decided to provide a holistic lesson about this particular data structure that includes data visualizations with formal explanations. Our final product has four web pages. The first page intends to introduce AVL Trees with detailed descriptions, including its background, functionalities, relationship to other structures, etc. The second contains a panel for AVL Tree visualizations, where students can create their own AVL tree and manipulate it with different methods. As students make an action, a corresponding explanation will be displayed along with the visualized data operation. Our third page provides students with the actual implementation of AVL Trees in two different programming languages. The final page gives a short quiz, where students can test their understanding of AVL Trees after learning from our website. Ultimately, our goal is to create a one-stop learning platform for data structures that provides interactive visualizations and comprehensive explanations of how data structures function. Our poster will discuss our pedagogical philosophy for studying data structures by displaying our website, introducing each page and section, and demonstrating our AVL Tree visualizations.

Poster 5040

Racial Representation in Television: Building a New Dataset Through Web Scraping and Machine Learning

Jenny Wang
Computer Science
Wellesley College, Wellesley, MA

Faculty Advisors
Eni Mustafaraj, Kyung Park
Wellesley College, Wellesley, MA

My senior thesis takes an interdisciplinary approach, combining computer science and economics, to pursue the question of how racial representation on television has evolved throughout time. First, I combine web scraping techniques with a pseudo-snowball sampling method to construct a novel dataset containing the near universe of television shows and cast and crew members from 1946 to the present. Then, I examine whether augmenting racial classifiers using additional data signals, such as images and biographical texts, can improve classification using ensemble learning methods.

Despite the media's prominence as a transmitter of public attitudes and norms, contemporary data is unavailable to understand and address racial representation and segregation in television. Although existing studies have generated in-depth reports detailing the challenges and impact of increasing diversity in film and TV, we still lack an understanding of why representation on television continues to lag behind population demographics. Scarcity of racial data affects domains beyond television. Social scientists often circumvent this issue by using name-based racial classifiers as a proxy for a person's true race. Through my analysis, I find that name-based racial classifiers may be unlikely to provide reliable predictions about racial representation due to measurement error. Thus, I use a subset of my television cast and crew members dataset to explore alternative methods for racial classification.

These results are relevant given the growing and influential literature that uses name-based classification to study racial cleavages in America, particularly in the social sciences. I test a variety of machine learning classifiers, including Naive Bayes, Logistic Regression, and neural models, on unstructured biographical data available through IMDb. In addition, I use stacked generalization ensemble methods to combine the predictions of my best-performing model with existing name and image racial classifiers.

Poster 5041

Combating Misinformation and Disinformation in the Digital Age: The Importance of Computer Scientists

Aleksandr Molchagin
Computer Science Department
Kalamazoo College, Kalamazoo, MI

Faculty Advisor
Sandino Vargas-Pérez
Kalamazoo College, Kalamazoo, MI

Misinformation and disinformation are two distinct terms with definitions that share a similar concept: spreading false information. While the problem of misinformation and disinformation is not new, the use of computers and the Internet has made it easier to spread false information and more challenging to combat it. In this poster, I review the problem of misinformation and disinformation in the digital era and discuss why people tend to believe in false information. I also propose how computer scientists can help to reduce the spread of misinformation and disinformation through technology and education. Finally, I present a prototype of a possible solution; an AI-powered search engine to highlight the significant role that computer scientists can play in combating misinformation and disinformation.

Poster 5042

Food Justice Map Software

Antonio Craveiro

Computer and Data Sciences Department
Merrimack College, North Andover, MA

Faculty Advisor

Christopher Stuetzle
Merrimack College, North Andover, MA

Food insecurity is an endemic issue that perpetuates societal injustices. The goal of this work is to construct an application that calculates and displays the food security of various local geographical regions in order to identify areas of food security concern. The project uses OpenRouteService's API to access the map database, which then relays information to our website, which visualizes individual census tracts in a given color, thus creating a heatmap. The heatmap shows the food security value of the tract. The security value is a composite of three functions. The first takes the average time to arrive at a grocery store by walking. The second value is the difference in the average price of normal goods relative to all grocery stores. Finally, the third value is the median income of the citizens of the census tract relative to all census tracts. This visualization drastically increases the usability of the software, as the heatmap makes quick work of conveying trends and areas of low food security to the user. This software is planned to be deployed tothe Merrimack Valley Transport Authority in order to identify areas of concern in food insecure areas, improving the standard of living for lower income communities throughout the Merrimack River Valley. This poster will discuss the development of the Food Justice Software, the challenges encountered during the project, future improvements to the prototype, and the impact of such software on communities. The full description and implementation features of the software will be presented in detail.

Poster 5043

Examining the Efficacy of Social Media Data Extraction on Multiple Platforms

Marios Petrov, Shawn Nash
Computer Science
New College of Florida, Sarasota, FL

Faculty Advisor
Fahmida Hamid
New College of Florida, Sarasota, FL

Social media sites are valuable data resources because they offer a vast and diverse array of information that can be used to gain insights into consumer behavior, market trends, and public opinion. During a recent independent study project (ISP), we set out to examine the efficacy of data-mining (and scraping) techniques on three different social media platforms. Our findings revealed that the most data-rich platform, Facebook, is the most complex and convoluted platform for researchers to scrape and gather data from. On the other hand, Twitter and Reddit were substantially more "scrapper-friendly" and enabled us to utilize the plethora of meta-data available. This polarity between platform policies on gathering meta-data can potentially have long-lasting repercussions on how society assesses and studies the impact of social media platforms. We suggest a standardized framework that Facebook and such platforms (current and future) should provide for researchers to access metadata without breaching user privacy and security to bridge this existing polarity effectively and thus let the independent and non-profit researchers use the enormous user-generated data used for research purposes.

Poster 5044

Warrior Shuttle

Owen Enestvedt, Derek Costello, Alex Elguezabal
Computer and Data Sciences Department
Merrimack College, North Andover, MA

Faculty Advisor
Christopher Stuetzle
Merrimack College, North Andover, MA

We present our experience on an interdisciplinary project developing a mobile shuttle tracking application, at Merrimack College. The app tracks the location of the Merrimack shuttle and provides schedule, routing, and destination information for users. The college provides the shuttle service for students and faculty to access on and off-campus locations of interest. The previous shuttle system left much to be desired; it was rarely on time, and the location and schedule were a mystery. Often, those interested in the shuttle system struggled to make use of it due to the lack of information. This meant the college was wasting resources on an underutilized service, and students who did take the shuttle were frequently left stranded in various locations around campus, and the surrounding communities with no support. To begin, our goal was to provide a live shuttle location alongside a schedule, ETA information, and a shuttle hotline. Most importantly, this information had to be quickly and easily accessible to users. We investigated different development platforms, location services, and programming languages. Knowing the app must be available on both iOS and Android platforms, we began writing a Flutter application which relies on the Dart language, to begin formulating a cross-platform UI. We decided to use Java Spring to support a backend model which communicated to devices and a custom location API. Our goal was to design a familiar and fluid mobile application, applying material we had learned in our object-oriented design and web development courses, which would enable students and faculty to capitalize on the shuttle service. Our poster will discuss our planning ideology, the progress of the app, and the challenges we encountered during the project. The full description and implementation features of the app will be presented in detail.

Poster 5045

STEM Data Dashboard

Jared Rathbun, Kathryn O'Neil
Computer and Data Sciences Department
Merrimack College, North Andover, MA

Faculty Advisor
Christopher Stuetzle
Merrimack College, North Andover, MA

We present our experience designing and implementing the STEM Data Dashboard, a web dashboard designed to allow easy analysis of institutional student data, such as analyzing student demographics, GPAs, DWF rates, and course success. Using Python and SQL, contemporary web design techniques such as RESTful APIs, responsive design, and object-orientation are demonstrated. Using data provided by Merrimack's Provost's office, the data is parsed into a relational database where it can be queried by the web dashboard to generate visualizations such as graphs, charts, and subsets of the data. To allow for simplicity, the dashboard features commonly used data and visuals, along with quick access to more general visualizations and a representation of the data in the database. Due to the sensitivity of the data, only users that have the necessary permissions are able to perform data administration tasks, such as adding data to the dashboard. Administrators also have permissions to change user's roles and permissions. Future work, being done during the Spring 2023 semester as an independent study will add features such as, making the database dynamic to new data, developing an API for creation of custom analyses, and adapting the dashboard to help answer some high-level research questions across the institution. Some of these questions include: what is the impact on retention rates, student performance, and class sizes because of COVID, is there a "linchpin" course within each major, around which success in the major is defined, does a particular change made to a course change the distribution of grades in the course, and what are the significant predictors of long term success in each STEM field?

Poster 5046

The Meeseeks Project: Real-Time 3D Motion Capture Data Visualization

Thomas Eckert, Yadel Negash
Computer Science
Quinnipiac University, Hamden, CT

Faculty Advisor
Chetan Jaiswal
Quinnipiac University, Hamden, CT

The sports training industry is valued at 11 billion dollars and growing, with part of that capital being exchanged through coaching and personal training. Our product aims to be an economical alternative by allowing people to train on their own with Machine Learning models. Utilizing just one camera, our product is able to accurately detect the movement of joints in real time. With the help of the machine learning solution Mediapipe, we are able to assist people with many types of exercises and sports. We aim to provide a product that can be used in any space with a UI that is simple to use.

Our product offers features that help users easily identify their movements by overlaying a motion capture ‘skeleton’ on the subject. This drawn skeleton updates with the subject's movements. We use the movements of the subject's joints to plot the coordinates on a graph and produce an animation with matplotlib python library. The skeleton is plotted based on the coordinate value for each joint, which is then written as a JSON file and stored to our MongoDB database. In the backend we used Python and for the frontend, we used HTML, CSS, and Javascript. The IDE we used for development is VScode and GitHub for version control. This animated 3D model can then be compared to other 3D models accessed from our MongoDB database. It is through this comparison that we can provide value to users by giving them examples of how to correctly perform the exercises and sports form (like a golf swing) demonstrated by our models.

Poster 5047

Online Food Ordering App with React.js

Kurt Mueller, Jacob Thibault
Department of Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisor
E. K. Iskrenova-Ekiert
SUNY Brockport, Brockport, NY

At SUNY Brockport, the late-night food joint TRAX is a popular dining spot for many students. A survey we conducted among student athletes showed that 56.2% of the respondents ranked convenience as their #1 motivator for going to TRAX, and 100% said they would order more often if coupons and combo deals were available. We plan to bring online ordering to TRAX by creating a Web app utilizing React.js to increase customer traffic to the dining location. This work was performed as part of a class project for an Independent Study at SUNY Brockport.

A Web application built with React.js has short page load times by utilizing dynamic updating of variables and components, such that each time a user interacts with the page a new page isn't required to load. React.js also lessens the load required by the user's device and minimizes communication needed between it and the server. Using React.js in our app provides the capability to easily add components, making the app extendable to accommodate future menu items and features. Additionally, React.js gives the webpage a modern-app-like interface. We are utilizing these features to build an interactive and appealing app for online food ordering from TRAX to improve upon what students already like, convenience, and give students what they're asking for: online-exclusive coupons and combos.

Poster 5048

Campus Navigation Mobile App

Alex Santeramo, Adam Chirico, Jakob DeBroff
Computer Science
Quinnipiac University, Hamden, CT

Faculty Advisor
Chetan Jaiswal
Quinnipiac University, Hamden, CT

As senior computer science students at Quinnipiac University, we wanted to develop an app that will help the incoming students and faculty be able to navigate a university in this case, Quinnipiac's three campuses (Mount Carmel, York Hill, North Haven) with ease. The current mobile application for the university is lacking many features and is quite bare bones in terms of functionality and design so we wanted to create our own app that can be adaptable to any university. Our goal is to add features that both our clients and we feel are needed for a well-developed navigation app.

Some of the main features of our app are: navigation to a building from the user's current location, informing the user on different buildings on any selected campus, and many others. To accomplish all the goals and features that were necessary for the app we took an agile approach consisting of two-week development sprints where we individually focused on implementing the features that we felt were of the highest importance for the app. At the beginning of our development cycle, we focused mainly on setting up the foundation of our app by using Google Maps API and connecting it to our app in Android Studio using Kotlin as our primary coding language. Now, in collaboration with our clients, a Senior Director of User Experience & Engagement as well as a Senior Director of Marketing & Communications, we can cater to specific needs that are needed for our app to be used by students and faculty as well as get their ideas and feedback on the work that we have done so far. We hope to be able to finish this navigation app by the end of March 2023 and have it distributed to the students and faculty at Quinnipiac University.

Poster 5049

Benchmarking Machine Learning Methods on Simulated Bioinformatics Data

Victor I. Robila
Hunter College Campus Schools
City University of New York, New York, NY

Faculty Advisor
Zijun Zhang
Cedar Sinai Medical Center, Los Angeles, CA

We describe the benchmarking of several machine learning methods on synthetically generated DNA data sets. Machine learning (ML) has been extensively used on DNA sequence data including for sequence alignment, classification, clustering, and pattern mining as a way to significantly speedup research [1]. While there are many ML methods currently employed, comparative studies allow researchers to understand their advantages and disadvantages, as and identify the ones most appropriate for a specific problem. This is particularly the case when automated tools like Neural Architecture Search (NAS) are not employed. The data used was formed of synthetically generated DNA sequences using simdna [2]. Such sequences have been shown to constitute a valuable and unlimited source of data for experiments and thus provide an accurate base of evaluation.

This project was completed in two steps. First, a Keras convolutional neural network (CNN) was adapted to provide a baseline for further experimentation. The CNN was then optimized using a custom developed grid search hyperparameter optimization. Second, several popular machine learning methods available in the sklearn library (support vector classifier (SVC), logistic regression classifier, random forest classifier, and gradient boosting classifier) were applied to the same simulated data. This is useful since, as sklearn is often used as a teaching tool, getting a better understanding of what models to apply to what datasets can be valuable for students. A thorough comparison was then conducted between these 4 methods and the original CNN on the basis of accuracy, fit, and time. The CNN outperformed the sklearn methods and the performance comparisons of the sklearn methods heavily depended on training time.

[1] A. Yang, W. Zhang, … and L. Zhang, "Review on the application of machine learning algorithms in the sequence data mining of DNA," *Front. Bioeng. Biotechnol.*, 8,1032, 2020.
[2] "simdna package — SimDNA" https://kundajelab.github.io/simdna/.

Poster 5050

Accessible Learning Labs

Jaden Wedner, Ryan Webb, Heather Moses
Software Engineering Department
Rochester Institute of Technology, Rochester, NY

Faculty Advisor
Samuel Malachowsky
Rochester Institute of Technology, Rochester, NY

We present our experience working for Accessible Learning Labs (ALL), an undergraduate research initiative embarked in the Software Engineering Department at the Rochester Institute of Technology. The goal of the research group is to promote accessible software development practices for individuals with disabilities, and more recently, provide contemporary knowledge about Artificial Intelligence and Machine Learning. The team does this by creating experiential browser-based lab activities which are hosted on the project website, all.rit.edu. To gain knowledge on how our labs influence students from computing and non-computing backgrounds, the research team has conducted several outreach events in which we collect survey and response data. These include class presentations used in several introductory software engineering courses. The team collected data from one control group and two experimental groups. Group A was asked solely to read from general credible outside sources; Group B was asked to complete the interactive lab exercise as well as watch a video from an outside source; Group C was asked to complete the interactive lab exercise and watch an empathy-creating reinforcement video from the lab. As a result, the participants learned about how people with visual impairments need screen readers to access software equitably. The results highlighted that Group B and C were more likely on average than Group A to believe it is easy to create accessible software. These results were determined to be statistically significant after conducting a single factor analysis of variance (ANOVA) test. Additionally, Group B and C were more likely on average than Group A to agree that accessibility should be a top priority for modern software developers. The full description of how the lab is presented and what the data yielded will be presented in detail.

Poster 5051

Web Based Dynamic Graduation Roadmap for Academic Institutions

Peter Sabio, Jordan Mayo
Computer Science
Quinnipiac University, Hamden, CT

Faculty Advisor
Chetan Jaiswal
Quinnipiac University, Hamden, CT

Our project focuses on creating an improved course planning and advising website for use by Quinnipiac students and faculty. Our project looks to improve the current advising and course planning website in use by the school. The current system in place for advising and choosing classes is half on paper and half online. Our goal with this project is to streamline the entire process and make it easy for students to understand the course they need and make educated decisions about their courses. Our project presents to users graduation requirements for different majors, and a list of corresponding classes to fill those requirements. Our project analyzes what courses a user has taken and their graduation path to provide recommendations for what courses to take in the future. In addition to this, users can message with their academic advisor and get approval on their course planning. The framework for this project will be highly adaptable and easy to implement for any university or academic institution.

Our system is a client/server website using AngularJS as the front end of the website. The users can access an interactive GUI that sends information and commands to the back end of the application. Our back end is run on Java using Spring Boot and can handle the requests coming from the front end. The backend uses Hibernate to retrieve information from the database. Our database is a MySQL database running a Google Cloud Platform.
.

Poster 5052

HeartBuddy

Devyn Souza, Joshua Shaw, Matthew Smith, Jake Intravaia, Nicolas Bennett
Computer Science
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

ElectroCardioGram (ECG) is a useful tool for diagnosing and monitoring a variety of heart-related arrhythmia conditions. Existing ECG models provide a limited average accuracy rate of 54% across all training levels. To address this issue, we propose *HeartBuddy*, an open-source software suite that delivers healthcare professionals with ECG data analysis of the electrical activity of the heart and provides valuable information about the heart's function and health. *HeartBuddy* uses Deep Learning (DL) techniques to interpret ECG data and provide confidence values to identify potential abnormalities present in the ECG. *HeartBuddy* offers a user-friendly interface to display identified abnormalities accompanied by a confidence value and relevant extracted ECG values. This can help users spot any potential abnormalities that could be present. The proposed DL model was trained using a verified large scale 12-lead ECG database provided by PhysioNet. This dataset was released in 2022 with ECG samples from 45,152 patients and is comprised of many smaller datasets. Our model's predictions are verified by utilizing an additional validated dataset. Additionally, *HeartBuddy* uses live gathered data and an on-hand cardiologist to validate the correctness of the model. Lastly, the *HeartBuddy*'s ECG analysis provides rationale for possible diagnoses. With the help of our algorithmic-based feature extraction, *HeartBuddy* provides relevant information pertaining to important features within the provided ECG sample. Healthcare professionals can examine the extracted features, such as certain intervals of peaks and valleys within the ECG signals, to better understand the rationale behind the abnormality detection. Moreover, by being open source, *HeartBuddy* boosts transparency by allowing the scientific community to engage in development and improvement of the existing framework to better match its needs.

Poster 5053

Arclight

Robert Debinski, Emiliano Vega, Tyler Nguyen, Tyler Ruocco, Lucien Zheng
Computer Science
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

A Roguelite is a sub-genre of video games that are inspired by the classic 1980 game Rogue. These games usually involve procedurally generated levels, permadeath, and random loot drops. The player must navigate through a series of increasingly difficult levels while managing limited resources such as health, ammo, and food. Existing roguelite games are 2D and room-based which result in repetitive and unbalanced gameplay and limited player engagement, making it difficult to progress or survive. To address this limitation, we propose to design and develop a new 3D roguelite game that is first-person-view, not room-based, and uses procedurally generated levels. The objective of this project is to confirm these characteristics help increase players' engagement. The proposed game, called Arclight, is built using Unity Technology Game engine. Arclight is original because it creates a Player versus Environment (PvE) single player game in the first-person point of view using roguelite elements. The story and setting take place in a SciFi environment where the character is a bounty hunter starting out in a space station city. Game development is constantly growing and open to new ideas, twists, and genres. To verify and validate our results, we propose to qualitatively evaluate gamers' play experience using a Global Environmental Survey. We aim to bring a new face and idea to the roguelite genre which we believe the fan base would like. If we stand by our pillars, Arclight will bring a new twist to the roguelite genre.

Poster 5054

Quantifying the Increase in Climate Instability

Patrick Bierach
Computer Science
Ithaca College, Ithaca, NY

Faculty Advisor
Ali Erkan
Ithaca College, Ithaca, NY

This project is an attempt to quantify the growing instability of climate. We focus on daily minimum/maximum temperature readings from thousands of weather stations distributed throughout the continental US and explore ways to assess the growing variability as a function of time. Our primary approach is based on multi-segment linear regression which is a dynamic programming algorithm to determine the ideal number of straight-line segments to reveal trends in a time-series of points. The algorithm's need to use a growing number of line segments is then interpreted as a proxy for more extreme fluctuations. We then use a simple linear regression per weather station to understand the direction of change as a function of years. An interpolation of these results leads to a raster representation that allows us to see change across the entire country. As a consistency check, we also compare our results with those produced by simpler algorithms designed for the same purpose, such as the number of times the trend in temperature-change itself changes direction (switching from increasing to decreasing and vice versa).

Once we identify areas that are determined to be going through extreme changes in climate, our goal is to seek spatial correlations with consequential weather events such as tornadoes. Another one of our goals is to identify the overlap between agricultural lands; if this overlap is significant, then our work will allow us to associate climate change in financial terms which may resonate with people who may have perceived climate change as inconsequential.

Our data comes from the National Centers for Environmental Information. Because the sampling of the temperature readings is not uniform for all weather stations, we expect there to be upper bounds to the accuracy of our conclusions. However, as we identify better datasets, our results will be more accurate.

Poster 5055

Backpropagation is expensive. Is it necessary?

Nika Gedenidze
Computer Science
Caldwell University, Caldwell, NJ

Faculty Advisor
Vladislav D. Veksler
Caldwell University, Caldwell, NJ

The last decade has seen huge advances in Artificial Intelligence, due in large part to convolutional neural networks (CNN) [1] (* references removed for space reasons *). Traditionally, neural networks only included dense layers (Dense-only) and employed backpropagation for training. CNN frameworks add filter layers in front of dense layers, but continue to employ backpropagation for all layers. Unfortunately, backpropagation is extremely expensive in terms of training times, required hardware, and electricity consumption. This research investigated the necessity of backpropagation for the filter layers in CNNs.

We ran simulations comparing efficiency and accuracy of Dense-only, CNN, and CNN without backpropagation (CNN-nobp) neural networks using two datasets commonly used to evaluate neural nets: MNIST handwritten digits (MNIST) [2] and MNIST fashion images (Fashion) [3]. We used Python and TensorFlow to implement all models with the same hyperparameters, and Keras to visualize their performance during training. All networks had a single dense layer and were trained for 5 epochs. CNN and CNN-nobp also had 3 filter layers.

(* Figure removed for space reasons *)

Figure 1. Accuracy and Efficiency of Dense-only, CNN, and CNN-nobp.

Results of the simulations are shown in Figure 1. CNN-nobp seems to provide major advantages over CNN in terms of training times, especially on the more complex problem (-57.7% on MNIST, -66.2% on Fashion), while losing only several accuracy points (-1.4% on MNIST, -4.9% on Fashion). Moreover, CNN-nobp is more accurate than Dense-only (+4.9% on MNIST, +3.0% on Fashion), suggesting that convolutional filter layers provide accuracy benefits regardless of whether they are trained via backpropagation.

Our future work aims to evaluate CNN-nobp with hardcoded (rather than random) filters, and with filters being automatically generated based on the data. Although automatically generating filters still requires calculation, it is exponentially cheaper than backpropagation.

References: (* removed *)

Poster 5056

Predictive Modeling of Critical Business Decision-Making Amid Russia-Ukraine War

Lily Tobin, Chi-Ni Lin, Anthony Gatti
Department of Computer Science
Wheaton College, Norton, MA

Faculty Advisor
Zhou Tong
Wheaton College, Norton, MA

Since the invasion of Ukraine by the Russian forces, companies around the globe have announced plans to withdraw or suspend their business operation in Russia. However, making such a decision has a substantial economic impact on the company and is far-reaching on the global economy. To better understand how the decision was made and what factors may have influenced such critical decision-making, we analyzed a wide range of factors that could affect a company's decision-making and reported our findings.

In this study, we collected data from various sources and created business profiles of 1,329 world-leading companies with a total market capitalization of 48.44 trillion, which accounted for approximately 39.7% of the global market. Each company profile includes information such as industrial classification, business operation in Russia, ESG risk rating, and financial performance. Among them, 760 companies (57.2%) announced their suspension or withdrawal and 569 companies (42.8%) remained in Russia. We analyzed and identified the top three factors that influenced a company's decision, including the headquarters country, industrial sector, and business exposure in Russia. We found that the remaining companies on average tend to have higher revenues and larger workforces than the companies that are exiting Russia. Companies in the EU that are directly threatened by Russia's military power are more likely to withdraw their business compared to companies in the U.S., Japan, and other NATO countries. Furthermore, we built a logistic regression model to predict whether a company will cease its business operation in Russia based on select factors. The preliminary results indicate that our model can predict such decisions with high accuracy. This study can also help companies understand and make informed critical business decisions in future conflicts.

Poster 5057

Multi-Armed Bandit EvaLuator (MABEL)

Alex Hetrick, Kaiy Muhammad
Department of Software Engineering
Rochester Institute of Technology, Rochester, NY

Faculty Advisor
Daniel Krutz
Rochester Institute of Technology, Rochester, NY

We present our Multi-Armed Bandit EvaLuator ((MABEL) that evaluates several Multi-Armed Bandit (MAB) models for real-world efficacy. This is the first known tool of its kind that can evaluate several MAB models, evaluating their ability to deal with corrupt and missing values. We also present a robust evaluation of several Multi-Armed Bandit (MAB) and Contextual Multi-Armed Bandit (MAB) models.

The MABEL tool uses a real-world scenario of autonomous connected vehicles (ACVs). These ACVs rely on sensors and inter-vehicle communication to operate safely and effectively as a fleet. This interconnectedness exposes the ACV system to cyberattacks in which an attacker injecting or manipulating sensor readings for even one vehicle could disrupt the entire ACV system. One approach to filtering out compromised sensor reading is with the use of MAB and CMAB machine learning models to assign penalties from a Gaussian distribution to the ACVs for crashes which often result from trusting data outliers. We researched algorithms that could be applicable and began with Epsilon-greedy, Linear Thompson Sampling, and LinUCB (Upper Confidence Bound). To fairly compare the models while handling the different input requirements, we created the ACV robust simulation MABEL using a MAPE-K adaptation control loop to run 100s of trials. We present our findings applying various MAB/CMAB models to MABEL to serve as a benchmark determining which model would be best suited for sensor outlier identification. The MAB/CMAB algorithm implementations and MABEL will both be presented as well as the findings from the evaluation itself.

The findings suggest that while not all MAB algorithms are suited to address the scenario we presented, some models such as LinUCB demonstrate notable success in filtering out compromised sensor readings in ACV systems.

Complete source code is available at: https://icmab.github.io/MABEL/.

Poster 5058

Snail Language Support

Charles Reinhardt
Mathematics, Computer Science and Statistics
St. Lawrence University, Canton, NY

Faculty Advisor
Kevin Angstadt
St. Lawrence University, Canton, NY

A high quality programming environment (often an Integrated development environment or IDE) can be vital to enhancing developer productivity. Visual Studio Code (VS Code) is a popular, open-source text editor maintained by Microsoft [2]. VS Code delivers language-specific features through freely downloadable, community-built extensions on an online marketplace. Many of these extensions allow developers to take advantage of editing features such as syntax highlighting, code-autocompletion, or debugging support. The Snail Language (Strings Numbers Arrays and Inheritance Language) is a simple, object-oriented programming language meant to be implemented in a one-semester undergraduate course [1]. I present my experience in developing a VS Code extension to provide language support for the Snail Language. The extension implements support for syntax highlighting, rudimentary auto-completion, and dynamic error-checking diagnostics using VS Code's Language Server Protocol (LSP) [3]. The extension also makes use of VS Code's Debug Adapter Protocol (DAP) to implement a debugger that supports breakpoints, start and stop behavior, and variable inspection [4]. First, I will give an overview of the contents of a VS Code extension. Next, I walk through how a VS Code extension runs, particularly highlighting the functions of VS Code's Language Server and Debug Adapter Protocols. Then, I discuss good software development practices including version control and documentation. Finally, I investigate the process of publishing a VS Code extension on the public extension marketplace.

References

[1] Kevin Angstadt. 2022. Strings Numbers Arrays and Inheritance Language (snail). Retrieved from https://snail-language.github.io/
[2] Microsoft. 2022. Visual Studio Code. Retrieved from https://code.visualstudio.com/
[3] Microsoft. 2022. Official page for Language Server Protocol. Retrieved from https://microsoft.github.io/language-server-protocol/
[4] Microsoft. 2022. Official page for Debug Adapter Protocol. Retrieved from https://microsoft.github.io/debug-adapter-protocol/

Poster 5059

A Unified Election Database in Relational as well as Key-Value Form

Emmett Barry
Computer Science
Ithaca College, Ithaca, NY

Faculty Advisor
Ali Erkan
Ithaca College, Ithaca, NY

Election data at the federal level can be found at many granularities from many sources. MIT's Election Lab site and Harvard University's Dataverse are, for example, repositories that provide datasets at the highest resolution (precinct level) for inquiries from anyone of any discipline. However, in all of these datasets, election cycles are presented as a different dataset which means any analysis through time requires the user to merge them; this is an obstacle for researchers that do not have a computing background. The goal of our project is therefore to create a relational database that contains US election data at the precinct level for multiple federal elections over a span of time (currently 2016 to 2020) including US President, US Senate, and US House.

The inclusion of the time dimension widens the set of queries to be ascertained. For example, what are the regions (as aggregations of precincts, or districts, or counties) whose winners are always consistent with the national election results? What are the regions that are steady in terms of their political choices? Our database also allows us to associate election results with other national measures of interest. As a proof of concept, we explore possible correlations between print media distributions (and their associated changes) with federal election results.

To assess the suitability of different database technologies for queries through time, we explore alternate representations as well. In particular, we have a Mongo DB instance of the same data, populated programmatically from the SQL instance. An empirical performance comparison of these competing representations is within our area of interest as well.

Poster 5060

Analyzing Colonialism in Africa through Relational and Graph-Based Databases

Simret Melak, Nardos Mamo
Computer Science
Ithaca College, Ithaca, NY

Faculty Advisor
Ali Erkan
Ithaca College, Ithaca, NY

While history is typically expressed through narratives, the actual information within a narrative is best represented as a graph; different thematic connections one can overlay over the same set of historic facts is in fact a consequence of this representation. As visualization systems have become more prevalent, it is now common to encounter network visualizations of history based on such relational/graph models.

Inspired by a particular visualization that expressed the conflicting political relations of the recent past among different actors within the middle-east, we decided to represent the colonial past of Africa as a relational database. Our work started with a study of authoritative textbooks to define all relevant forms of entities and all consequential forms of relations. We then created a properly normalized SQL database that represented colonized countries (and their economies), colonizing countries, causes of conflict, international organizations and other countries involved in these conflicts, and descriptions of what the colonizers did.

Simple join-based queries confirm that our database can restore the facts we read in our source books. However, examining the lasting impact of colonialism on Africa requires us to explore long chains of events, such as tracing roots of modern conflicts to the colonial era, demonstrating the effects of these conflicts on the continent's economy, resources, and geopolitical relations. When the length of these chains are unknown, SQL falls short, as it requires us to know the number of joins needed to make particular connections. Consequently, our current work is also focused on porting this relational structure into an actual graph based database (Neo4J) to benefit from its declarative query language to search for paths. However, we retain the relational version as the canonical instance and automate the creation of the graph instance programmatically. We also plan to reveal our findings through a GIS system.

Poster 5061

Spatial Analysis of the Iron March Data Dump

Will Regalado Succop, Mateo Hernandez
Computer Science
Ithaca College, Ithaca, NY

Faculty Advisor
Ali Erkan
Ithaca College, Ithaca, NY

PEGIDA (whose English translation is an acronym for "Patriotic Europeans Against the Islamisation of the West") is a pan-European, anti-Islam, far-right political movement that was founded in Dresden in 2014. When the addresses of its members were leaked around 2016, a choropleth map of their distribution (normalized by district populations) revealed a strong correlation to the east-vs-west division within Germany as well as areas of high unemployment. While this did not reduce the threat of their racist agenda, the correlation shed some light on factors that encourage the formation of such groups.

The Iron March Data Dump of 2020 provided a similar window into a particular white supremacy group from the US, but (to the best of our knowledge) no spatial analysis of their activities has been conducted since then. In this project, we first worked on providing this missing analysis through (i) the creation of a relational database of the leaked data, (ii) the drawing of all spatial information from these data, and (iii) a Geographic Information System (GIS) representation of what we discerned. We then proceeded to couple this dataset with census data we obtained from the U.S. Census Bureau. Due to the limited extent of data we have at our disposal, we have not yet been able to provide strong correlations between the distributions of the Iron March members and the other layers within our system. But our work is indicative of the value of (i) spatial representation of data, (ii) GIS, and (iii) the intersection of computing with real-world problems, especially with datasets of societal importance.