NMAS
NEW MEXICO ACADEMY OF SCIENCE

THE NEW MEXICO JOURNAL OF SCIENCE

Volume 56
December 2022

2022 NMJS Editorial Board

Table of Contents

Editor's Note

The *New Mexico Journal of Science* (*NMJS*) is the annual publication of the New Mexico Academy of Science (NMAS). Each volume of the *NMJS*, which has been published since 1960, contains research papers and review articles deemed of interest to the scientists, educators, and citizens of New Mexico, in addition to some highlights of the academy's activities from the past year. Some volumes have addressed topics of historical, social, or economic interest while others have emphasized scientific areas in which New Mexico is particularly active. While the *NMJS* is published only once a year, papers are accepted for review and publication consideration on a rolling basis. All papers are peer reviewed by volunteers in the relevant scientific discipline. Additionally, opportunities exist for students and young authors to work with mentors to help get their work published. This volume would not have been possible without the hard work of the authors and reviewers from prominent New Mexico institutions and other organizations including Eastern New Mexico University (Portales, NM); Princeton University (Princeton, NJ); and Stanford University (Stanford, CA). The publication of this volume was enabled by huge professional efforts from Associate Editors Ms. Brittney Van Der Werff (NM EPSCoR) and Mr.Kevin Zhang (Princeton University) and Copy Editors Ms. Sara Pichette (NM EPSCoR) and Dr. Selena Connealy (NM EPSCoR), and all the listed and additional anonymous reviewers. The *New Mexico Journal of Science* is available for free download from the NMAS website at www.nmas.org. This enables the NMAS to reach a wide readership. Prior to 2008, the Academy mailed paper volumes of the journal only to its members. Those hard copies are available to the public upon request. Concurrently, we are also undertaking a digitalization initiative to bring all of the past print-only copies of the *NMJS* into online pdfs and establishing an online indexing system to increase the impact and visibility of the journal to more researchers.

Vladislav I. Sevostianov
Editor-in-Chief
New Mexico Journal of Science

Changes in Rainfall Factors in the Context of Hurricane Harvey

Vladislav Sevostianov [a, b*]

[a)] Department of Civil & Environment Engineering, Princeton University, Princeton, NJ, USA

[b)] Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, NJ, USA

ABSTRACT

This paper uses Hurricane Harvey as a case study to explore factors that affect extreme rainfall. Radar reflectivity analysis at three-hour resolution using raw NOAA data is used to explore the fundamental factors that are affecting extreme rainfall and how these factors, and their changing influence and severity, affect the likelihood of future extreme storms. In particular, geographic and weather timing peculiarities of Hurricane Harvey are discussed, in addition to the factors of urbanization, urban heat islands, abundant aerosols, and the Clausius-Clapeyron relationship with contextualization from literature. Lastly, arguments are made regarding the unbounded nature of extreme rainfall.

KEYWORDS: Hurricane Harvey, Extreme Storms, Rain, Radar, Urban, Hydrometeor Nucleation, Rainfall Bounds.

1 INTRODUCTION

Hurricane Harvey was the second costliest storm in U.S. history (behind Hurricane Katrina), largely due to catastrophic flooding from historic rainfall (Blake and Zelinsky, 2018; Brauer et al., 2020). Emanuel (2017) found the rainfall to be a 2000-year maximum (even using a lower value of 500 mm rather than the 840 mm that was recorded for the storm's accumulation), clearly showing the severity of the precipitation. Global changes, however, will make the likelihood of an event like Harvey from a once in 2000 years to a once in 100 years event by the end of the century, with climate change making the event 3.5 times more likely (Zhang et al., 2018; Emanuel, 2017). Hurricane Harvey started out as a "typical weak August tropical storm" that, initially, actually dissipated over the Caribbean (Blake and Zelinsky, 2018). The storm then reformed and moved northward, rapidly growing into a category 4 hurricane by the time it made landfall near Corpus Christi. After the original tropical storm had dissipated, its remnants remained convectively active. This low-pressure area on around August 23rd exhibited an increase in persistent deep convection, at which time a tropical depression was formed (Blake and Zelinsky, 2018). The storm largely consisted of little shear for wind but with warm water and mid to high layers of moisture.

Making landfall, storms lose energy (Brauer et al., 2020). Upon landfall Hurricane Harvey was weaker; given that most of the storm was often still over the gulf, this fed the entire system. Figure 1 shows the track of the storm, with a large concentration of dots all pivoting

around Houston showing the stationary nature of the storm over the city. Harvey generally slowly moved northeastward, though a north and eastward stationary front developed, keeping Harvey planted over the Houston area with heavy precipitation (Brauer et al., 2020; Blake and Zelinsky, 2018). The center of the storm was to the south of Houston (Figure 1), meaning there was little storm weakening from land—it was continuously fed by the gulf. Only by going completely inland over Louisiana and ultimately as far north as Kentucky did Harvey fully dissipate.

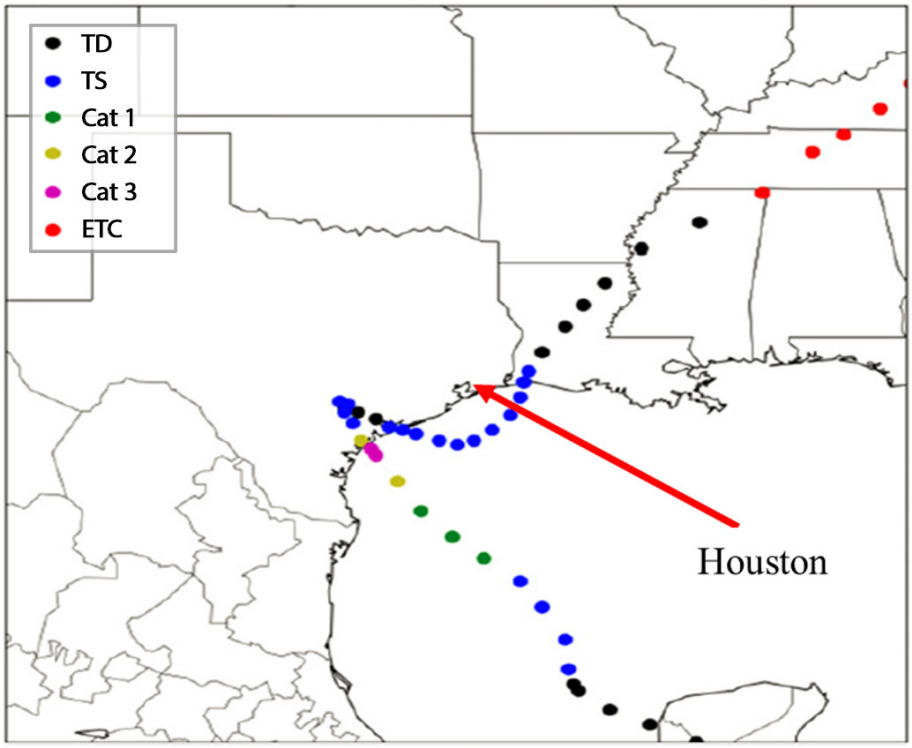

Figure 1. A map based off the HURDAT hurricane database showing the circulation center of Hurricane Harvey in 6 hour increments along with the corresponding storm category characterization. Annotated from Brauer et al., 2020.

2 ANALYSIS

2.1 Three-Hour Resolution Summary of Hurricane Harvey

Using sub-hourly radar analysis of Hurricane Harvey using NOAA data, a three-hour interval summary table (Table 1) was produced to examine the development of the storm and identify possible features of interest at critical points. Overall, this analysis suggests that the largest factors influencing the severity of Hurricane Harvey were its extended stationarity due to larger scale weather and that reflectivity and correspondingly, precipitation, increases over urbanized areas. This helps answer the primary questions addressed here: what fundamental factors are affecting extreme rainfall, how are they changing, and what does this mean for the future?

Date and Time	Description of the Storm	Important Points
August 26, 3 UTC	Landfall as category 4 hurricane near Corpus Christi. Houston is already hit with the outer rain-band.	Time of maximum winds. Surface roughness begins increasing.
August 26, 6 UTC	The eye is completely over land, the outermost band is deformed. Storm winds weaken to category 3.	Weaker winds as the storm moves inland. Based on reflectivity, the strongest rain is southeast of Houston.
August 26, 9 UTC	The storm is mostly overland and has weakened to a category 2 hurricane. Overall, the storm is moving north.	The outermost rain band is largely disrupted; convective rain is hitting north Houston with high reflectivity values around 50 dBZ.
August 26, 12 UTC	The eyewall is beginning to collapse. Slowing winds bring the storm down to a category 1 hurricane.	The outermost two rain bands have merged. Some hail is likely falling though mainly heavy rain. The storm in now moving towards gulf again.
August 26, 15 UTC	The storm has a wind speed around 35 knots, and thus is becoming a tropical storm or weaker for its remaining lifecycle.	It nearly appears as if the storm is splitting into two, with an outer storm of heavy rain and hail and the inner primary core of lighter rain, which is disconnected, meaning the storm is widening.
August 26, 18 UTC	New moister is constantly being brought in from the gulf; high reflectivity values are observed even over water.	Houston itself appears to have a brief respite with heavy rain east of the city. Differential reflectivity values here are low (~1) though very noisy.
August 26, 21 UTC	While the entire region is generally swamped with precipitation (~30 dBZ), small and isolated ultra-high reflectivity hotspots (~50 dBZ) dot the Houston metropolitan area.	Some localized area effects are likely contributing to high rainfall over Houston even as the storm appears to slowly move.
August 27, 0 UTC	The storm has been stationary with little change in the last few hours.	Individual rain bands appear to widen as they approach Houston from the southwest. The intensity appears constant (consisting of heavy rain with hail) but spreading its affected area.
August 27, 3 UTC	The storm moved only slightly northward, but the last highly reflective rain band moved enough to cover eastern Houston.	This may be the last of the major rain bands; the remaining ones closer to the former hurricane eye tend to clump but exhibit lower reflectivity values, suggesting more moderate rain in the future.

August 27, 6 UTC	Very gradual bulk movement takes place eastward towards Houston. Individual rain band reflectivity values tend to increase.	Interestingly, rather than high reflectivity values solely arriving from the southern gulf, values increase from the inland west simply by arriving at the Houston metropolitan area.
August 27, 9 UTC	Individual rain bands are no longer distinguishable. The counterclockwise bulk movement can be discerned through radial velocity, but overall, the storm is stationary.	Reflectivity values at all elevation angles are increased around Houston but lower just a couple miles outside the most developed area. Highest values are directly south. Very high differential reflectivity values around 3 dBZ are over Houston, suggesting a high concentration of smaller rain drops.
August 27, 12 UTC	Significant reflectivity values now form over the gulf (~50 dBZ +), which tend to drift and strength northward.	No particular topographic change appears to play a role, simply by being stationary significant moisture is being funneled from the gulf onto the nearest land: the Houston area.
August 27, 15 UTC	The storm's cyclonic nature is somewhat disrupted from radial velocity measurements. Highest reflectivity values are northeast of Houston.	Given the geography of Texas, the position of the storm center south of Houston necessitates that precipitation falling closer to Beaumont originally spends more time over the gulf (thus strengthening more).
August 27, 18 UTC	The highest reflectivity values are in eastern Houston, arising from the bulk counterclockwise rotational motion of the storm despite its otherwise stationarity.	So long as the storm center does not move, it appears its behavior will be consistent otherwise.
August 27, 21 UTC	Little change has occurred over the past few hours. "More of the same" seems to be a tendency.	The stretch from Galveston and Houston exhibits the highest reflectivity values, which is also the most heavily urbanized area.
August 28, 0 UTC	Differential reflectivity values have grown extremely high to ~6 dBZ over Houston and Galveston, but standard reflectivity values remain more or less constant.	Hurricane Harvey continues to pound the Houston area; its overall severity appears caused by its stationarity though individual area features appear to create local maximum for precipitation.

Table 1. A detailed analysis of the development of Hurricane Harvey from the time of landfall over 2 days using radar analysis. Comprehensive sub-hourly analysis is summarized into 3-hour increments over 2 days. Examined radar data includes wind field measurements, reflectivity, radial velocity, differential reflectivity, and differential phase shift from level II and level III Houston and Corpus Christi NOAA radar stations on August 26 and 27, 2017. The precipitation size and type classifications found are comparable to those found in literature for these days (Wolff et al., 2020).

2.2 Radar Analysis

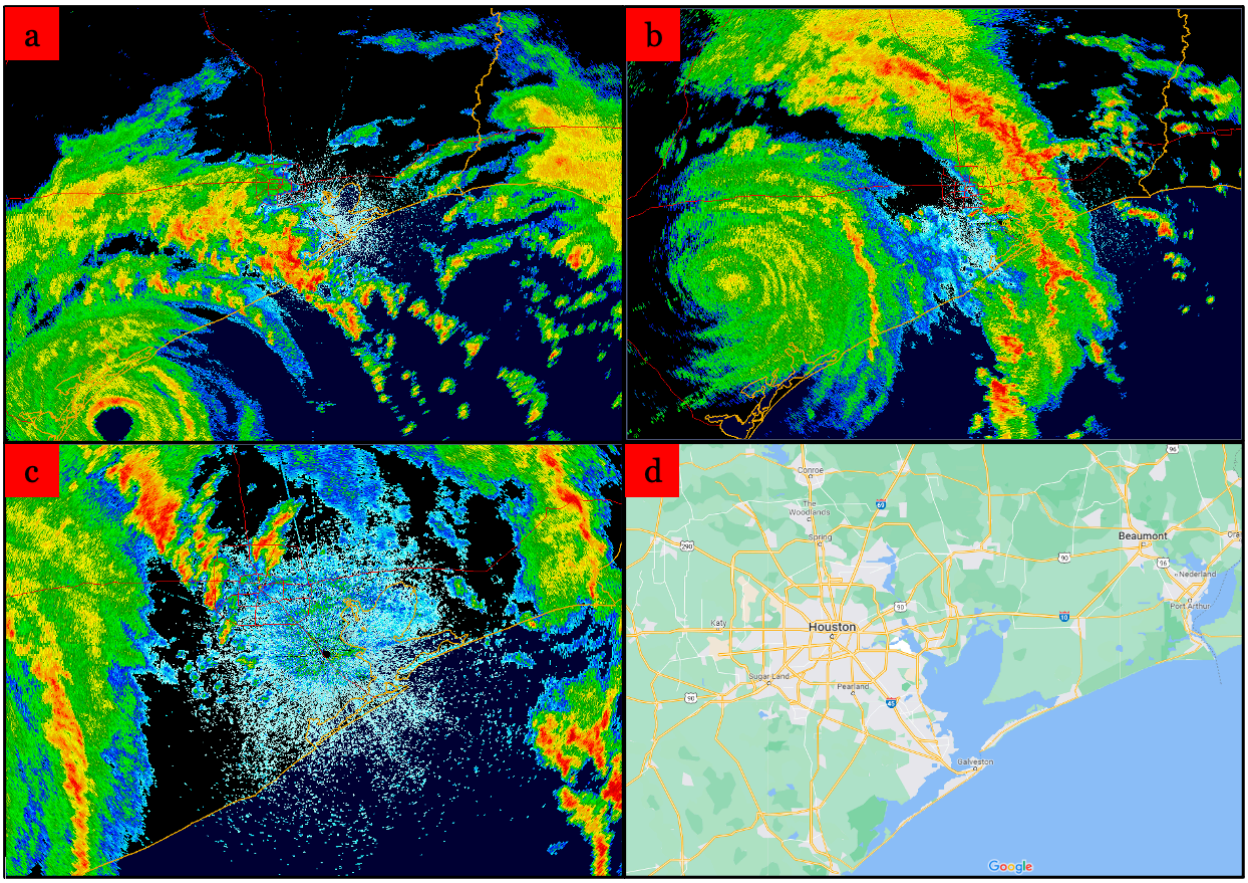

Figure 2. Several images of the Houston area during Hurricane Harvey during critical points all on August 26th, 2017. The same sort of imaging was used to compose Table 1. (a) Reflectivity measurements at 00:05:31 UTC before landfall. The hurricane eye is clearly still over water, but Houston and Corpus Christi are both already experiencing rainfall. (b) Reflectivity measurements at 15:01:26 UTC. This is the point when the winds of the storm have just calmed enough to no longer be classified as a hurricane. The temporary "splitting" nature of bands is visible. (c) Reflectivity measurements 21:01:17 zoomed in to show features near Houston. Individual pockets of high reflectivity like those shown in the northwest tend to pop up around this time as individual "hotspots" near Houston. (d) A Google Maps image showing urban (gray) vs. nonurban (green) areas around Houston. Higher reflectivity values typically match up over the urban areas more for Hurricane Harvey.

The radar analysis demonstrates that the storm's inability to keep moving northward is likely the primary driver of the extreme precipitation, as corroborated by literature. Blake and Zelinsky (2018) noted that Hurricane Harvey's initial northwest motion stopped when the storm became entrained between two light high-pressure steering currents. The bottom one was over the northern Gulf of Mexico, the top one at the mid-tropospheric level from around the four corners region of the U.S. eastward. Nonetheless, besides the stationary nature of the storm due to the larger simultaneously occurring weather patterns, several factors, which have been changing over time, played a role in the severity of precipitation brought by Hurricane Harvey.

3 EXTREME RAINFALL FACTORS

Zhang et. al. (2018) found that increased urbanization not only contributed to more severe flooding, but also directly increased precipitation. Flooding is easier to rationalize: urban areas with asphalt and concrete reduce groundwater infiltration by the rain, increasing the speed and severity of runoff, which increases flooding. However, the direct contribution of urbanization to precipitation is less clear. Through statistically significant modeling analysis, Zhang and colleagues found that the higher surface roughness of urban areas (buildings, as compared to a hypothetical crop field in the same place) increases drag on storm winds, helping create more updraft and cyclonic flow. This brings an increase in low-level convergence and upper-level divergence, which are conditions that help stimulate precipitation (Zhang et al., 2018). These effects are consistent with the radar analysis performed and summarized in Table 1, where ultra-high rainfall reflectivity measurements tend to appear in the most urbanized areas in and around Houston (Figure 2).

Furthermore, local urban surface warming (like the heat island effect) potentially also increases precipitation, as warmer air holds more moisture. However, Zhang and colleagues note that in Houston for Hurricane Harvey, anthropogenic heat played no major role in rainfall. Given that there are physical mechanisms (see following Clausius-Clapeyron equation discussion) that would link the two together though, perhaps in other contexts, urban surface warming plays a larger role. Given the well documented increase in urban heat islands, greater warming would likely only act as positive feedback to increase local precipitation (Sharma et al., 2017). In the case of Hurricane Harvey, a possible explanation for the discrepancy is that the urban area was not a source of moisture and contributed little to transport—the rain was deposited directly over the urban area before having a chance to transport through it. A different geographic configuration, in which the urban heat islands are located in either less rough areas (i.e., very large, but low in height, urbanized areas—like massive parking lots) or areas that otherwise interfere with transport less, would likely lead to a different effect with a large contribution to precipitation. Consequently, as more people increasingly live in urban areas and urban areas grow, precipitation for those people in those areas may increase due to urban surface warming. As a side note, given that convective processes likely give a significant contribution to extreme rainfall, this means that other (nearby, non-urban) areas from which some of the moisture to the urban areas is moving may experience less rainfall simply due to conservation principles. In any case, urban surface warming's effect on rainfall should be investigated further.

Local atmospheric changes also contribute to changes in precipitation. For hydrometeors to form, nucleation seeds are necessary (Beydoun, Polen, and Sullivan, 2016). Urban areas particularly exhibit higher concentrations of aerosols (particulate matter), which are directly emitted from both biogenic and anthropogenic sources. Secondary aerosols are primarily formed through atmospheric reactions between volatile organic compounds (VOCs) and NOx (Ioffe, Isidorov, and Zenkevich, 1977). VOCs are largely produced naturally by vegetation (especially in summer in regions like the southeastern US, which influences Houston) and increasingly, in highly populated areas (also Houston), by consumer products (McDonald et al., 2018). NOx is usually produced through high temperature combustion of fossil fuels (Jacob, 1999). The atmospheric combination of both then leads to the formation of the aerosols, which can act as seeds (with production typically limited by NOx availability in today's

cities) (Carter, 1994; Jacob, 1999). Cloud seeding is known to increase precipitation and has been used around the world, including New Mexico, to stimulate rainfall (Battan, 1977; Downing, 2022). Consequently, as unintentional seeding takes place in urban areas due to increased concentrations of aerosols, total precipitation will increase by increasing precipitation efficiency, provided moisture is available. Though Hurricane Harvey likely had a high concentration of sea salt aerosols as nucleation sites, the idea that air pollution played a role in increasing precipitation (including though stimulating convection) has been corroborated by others (Hu et al., 2020). As air standard restrictions around the world become stricter, the effect of "unintentional seeding" should slowly diminish (Koppmann, 2007; Zeng et al., 2019). Nevertheless, urban and industrialized areas today have significantly more free aerosols that may increase precipitation over preindustrial times (Anenberg et al., 2010; Hu et al., 2020).

Lastly, of the changing factors examined here, temperature is perhaps the most clearly connected factor to extreme rainfall. The Clausius-Clapeyron equation clearly relates temperature to moisture:

$$\frac{1}{e_s}\frac{de_s}{dT} = \frac{L_v}{R_v T^2} \tag{1}$$

Where e_s is the saturation vapor pressure of water, T is temperature (in degree kelvin), L_v is the latent heat of vaporization, and R_v is the gas constant for water vapor (Martinkova and Kysely, 2020). Examining just this equation, with more moisture, all other factors held constant, we expect more precipitation (Ye et al., 2014). This is also part of the mechanism that would suggest urban heat islands contribute to increased precipitation. Per degree Celsius increase in temperature, the atmosphere is expected to hold around 6-7% more moisture (Risser and Wehner, 2017). It should be noted however, that there is a difference between simply the moisture holding capacity of the atmosphere and actual rainfall, and the exact relationship is not fully established, but a moister atmosphere leads to more precipitation (Donat et al., 2016). According to the most recent IPCC, global temperatures have increased around 1.2 degrees Celsius since pre-industrial times and are expected to keep increasing, thus increasing atmospheric moisture as well (Ye et al., 2014).

With a heavily urbanized area and an aerosol abundant atmosphere in a warming climate, Houston can be expected to have a higher precipitation than it would being less urbanized, with fewer aerosols, in a cooler environment. However, the average precipitation may not change significantly—it is the intensity of the precipitation that will. This is best demonstrated in Figure 3, where historical data for both average precipitation (fairly constant) and the number of days with more than 3 inches of rain (an "extreme" event, increasing) from 1900 onwards is shown. Figure 4 shows direct high-resolution rain gauge measurements over the Houston area during Hurricane Harvey with areas east or north of significant topographic changes (sea/land and rural/urban) demonstrating the highest rainfalls, in line with the storm's behavior. Nonetheless, the severity of Hurricane Harvey upon Houston leads to several other important insights.

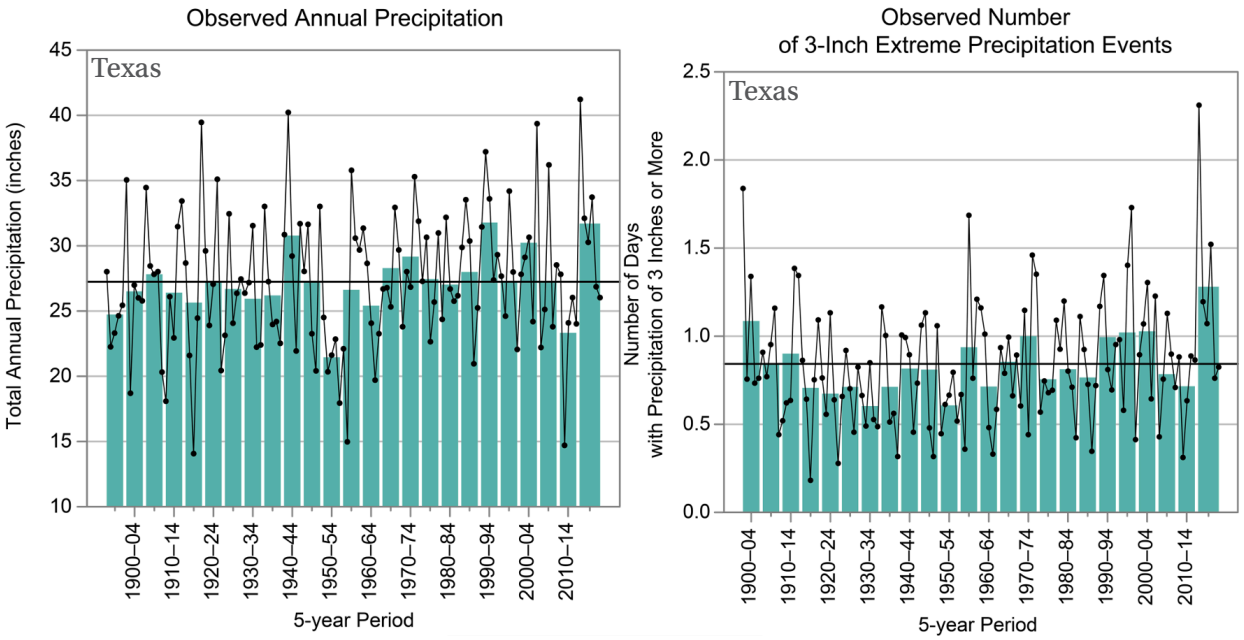

Figure 3. Left: Historical observed annual precipitation in Texas in 5-year increments. Right: Historical observed 3-inch daily rainfall extreme precipitation events in Texas in 5-year increments. During the most recent 2015-2020 period, both the average annual precipitation and the number of extreme events were above average. However, for average annual precipitation, there is no clear increase. Extreme precipitation events generally appear to be increasing, with exceptionary periods. From Runkle et al. (2022).

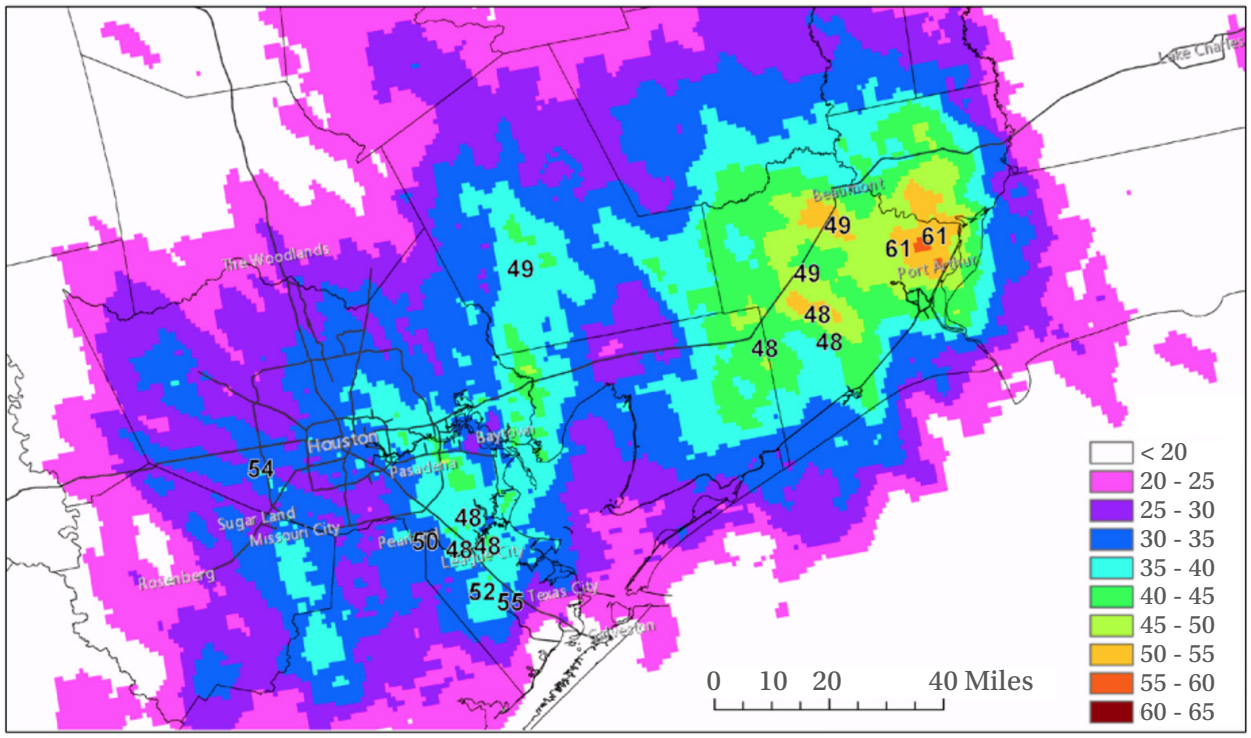

Figure 4. Rain-gauge corrected precipitation estimate for Hurricane Harvey over the Houston area. From Blake and Zelinsky (2018).

Ye and colleagues (2014) define precipitation efficiency to be "the quotient of total precipitation accumulation to the total precipitable water at the surface at the same location within a temporal period" (Brauer et al., 2020). Precipitation efficiencies are typically around 30% for a given column of air (typically due to baroclinic instabilities and deep-layer shear), but if efficiencies are greater than 100% (as was the case for Harvey over Houston from August 26th to 29th), then this provides a possible way to account for dynamic processes and powerful updrafts that lead to extreme precipitation (Brauer et al., 2020). For serious storms, rainfall would then be limited by dynamic processes and the amount of moisture than can be brought in. Brauer and colleagues (2020) found through radar analysis that on August 28th Houston had a precipitation efficiency of around 300%, with a significant concentration of large raindrops being brought in and lofted above the freezing layer of the storm. This would be enabled with an abundant supply of nucleation sites (particulate matter/aerosols), stationarity over the Gulf of Mexico (as opposed to, for example, inland Texas), and a supportive ground roughness (typical urbanization).

4 DISCUSSION

With such high efficiencies, an important follow up question is whether rainfall is bounded; if 100% precipitation efficiency can be easily exceeded then any simple limit is clearly incomplete. Boundness is a critical question for disaster management and design of critical infrastructure, where maximum possible values dictate life-and-death level engineering requirements. Traditionally, strict limits have been set using the previously introduced Clausius-Clapeyron equation—for a given temperature, there is a given amount of water. Hurricane Harvey had over 100% efficiency, meaning that convective processes bringing in moisture are being unaccounted for, in addition to the existing and accounted for precipitation processes not being fully quantified.

If in fact moisture transport is responsible for setting bounds on rainfall, then rainfall must be effectively unbounded, as this is nearly impossible to quantify (though at an extreme, we can set unhelpful bounds such as by considering all the water on the planet, but the objective here is to investigate possible methodological bounds). While the water may need to move at a higher speed and pressure, fundamentally, so long as there exists movable moisture, rainfall can effectively increase without limit. Much like a wave might bring a little water onto the shore, a big wave—going as far as a tsunami—by moving faster and with greater energy, will bring more water. So long as moisture and energy are available to feed a storm and ample nucleation sites exist, precipitation will increase. Statistical analysis supports this notion as well. By analyzing the generalized extreme value distribution for storms in the Houston area from 1950 to 2016, Risser and Wehner (2017) found that the shape parameter is positive, which supports the notion of an unbounded tail—in other words, an unbounded maximum rainfall. It should be noted that the necessitated small sample size with geographic consideration, and limitations in long-term data records of extreme events, do make such purely statistical analyses fundamentally weak without supporting mechanistic explanations.

Literature tends to be in consensus that the severity of storms like Hurricane Harvey is increasing with time (Emanuel, 2017; Brauer et al., 2020; Risser and Wehner, 2017; Zhu, Quiring, and Emanuel, 2013; Zhang et al., 2018). Overall, Hurricane Harvey does not appear to

have fundamental physical differences from typical storms despite its precipitation severity. Rather, larger scale weather behavior, together with an increase in rainfall inducing factors, all materialized together in space and time to create the devastation of Hurricane Harvey. Sea surface temperature anomalies in the Gulf of Mexico at the time were actually negative, which likely weakened the storm from its "full potential" (Brauer et al., 2020). Given that rainfall is likely unbounded, as the factors of urbanization, particulate matter/nucleation site abundance, and warming temperatures that all stimulate higher rainfall increase, higher rainfall will increase without bound too. Over time, we should expect more extreme storms with more intense rain. Examining other recent extreme events besides Harvey such as Hurricanes Florence, Ida, and Sandy, among others, and the changing rainfall-inducing factors that played a role in them will help further provide a more general understanding of the changes in extreme rainfall currently taking place to better prepare for the future.

AUTHOR INFORMATION

***Corresponding Author**

Vladislav Sevostianov
216 Bowen Hall
Princeton University
Princeton, New Jersey 08540
Email: vs14@princeton.edu

ACKNOWLEDGMENT

The author acknowledges financial support from the Department of Defense and U.S. Army research office sponsored National Defense Science and Engineering Graduate Fellowship and support from Princeton University.

REFERENCES

Anenberg, S. C.; Horowitz, L. W.; Tong, D. Q.; West, J. J. An Estimate of the Global Burden of Anthropogenic Ozone and Fine Particulate Matter on Premature Human Mortality Using Atmospheric Modeling. *Environ. Health Perspect.* **2010**, 118 (9), 1189–1195. DOI: 10.1289/ehp.0901220.

Battan, L. J. Weather Modification in the Soviet Union — 1976. *Bull. Am. Meteorol. Soc.* **1977**, 58 (1), 4–19.

Beydoun, H.; Polen, M.; Sullivan, R. C. Effect of particle surface area on ice active site densities retrieved from droplet freezing spectra. *Atmos. Chem. Phys.* **2016.** 16, 13359–13378. DOI: 10.5194/acp-16-13359-2016.

Blake, E. S.; Zelinsky, D. A. National Hurricane Center Tropical Cyclone Report: Hurricane Harvey. *Natl. Hurric. Cent. Trop. Cyclone Rep.* **2018**, No. May, 1–77.

Brauer, N.; Basara, J.; Homeyer, C.; McFarquhar, G.; Kirstetter, P. Quantifying Precipitation Efficiency and Drivers of Excessive Precipitation in Post-Landfall Hurricane Harvey. *J.*

Hydrometeorolgy **2020**, 21, 433–452. DOI: 10.1175/JHM-D-19-0192.1.

Carter, W. P. L. Development of Ozone Reactivity Scales for Volatile Organic Compounds. *Air Waste* **1994**, 44 (7), 881–899. DOI: 10.1080/1073161X.1994.10467290.

Donat, M. G.; Lowry, A. L.; Alexander, L. V.; O'Gorman P. A.; Maher, N. More extreme precipitation in the world's dry and wet regions. *Nature Climate Change* **2016**, (6), 508–513. DOI: 10.1038/nclimate2941.

Downing, M. New Mexico approves plans for cloud seeding in the southeast portions of the state. https://abc7amarillo.com/newsletter-daily/new-mexico-approves-plans-for-cloud-seeding-in-the-southeast-portions-of-the-state (published May 20, 2022)

Emanuel, K. Assessing the Present and Future Probability of Hurricane Harvey's Rainfall. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 114 (48), 12681–12684. DOI: 10.1073/pnas.1716222114.

Hu, J.; Rosenfeld, D.; Ryzhkov, A.; Zhang, P. Synergetic Use of the WSR-88D Radars, GOES-R Satellites, and Lightning Networks to Study Microphysical Characteristics of Hurricanes. *J. Appl. Meteorol. Climatol.* **2020**, 59 (6), 1051–1068. DOI: 10.1175/JAMC-D-19-0122.1.

Ioffe, B. V; Isidorov, V. A.; Zenkevich, I. G. Gas Chromatographic-Mass Spectrometric Determination of Volatile Organic Compounds in an Urban Atmosphere. *J. Chromatogr.* **1977**, 142, 787–795.

Jacob, D. Introduction to Atmospheric Chemistry, 1st ed.; Princeton University Press: Princeton, 1999.

Koppmann, R. Volatile Organic Compounds in the Atmosphere; Blackwell Publishing, 2007.

Martinkova M.; Kysely, J. Overview of Observed Clausius-Clapeyron Scaling of Extreme Precipitation in Midlatitudes. *Atmosphere.* **2020**, 11(8), 786. DOI: 10.3390/atmos11080786

McDonald, B. C.; Gouw, J. A. de; Gilman, J. B.; Jathar, S. H.; Akherati, A.; Cappa, C. D.; Jimenez, J. L.; Lee-Taylor, J.; Hayes, P. L.; McKeen, S. A.; Cui, Y. Y.; Kim, S.-W.; Gentner, D. R.; Goldstein, G. I.-V. A. H.; Harley, R. A.; Frost, G. J.; Roberts, J. M.; Ryerson, T. B.; Trainer, M. Volatile Chemical Products Emerging as Largest Petrochemical Source of Urban Organic Emissions. *Science*. **2018**, 764 (February), 760–764. DOI: 10.1126/science.aaq0524.

Runkle, J.; Kunkel, K. E.; Nielson-Gammon, J.; Frankson, R.; Champion, S. M.; Stewart, B. C.; Romolo, L.; Sweet, W. Texas State Climate Summary 2022. *NOAA Technical Reports*. **2022**, NESDIS 150-TX. NOAA/NESDIS, Silver Spring, MD, 5 pp.

Risser, M. D.; Wehner, M. F. Attributable Human-Induced Changes in the Likelihood and Magnitude of the Observed Extreme Precipitation during Hurricane Harvey. *Geophys. Res. Lett.* **2017**, 44 (24), 12,457-12,464. DOI: 10.1002/2017GL075888.

Sharma, A.; Fernando, H. J. S.; Hamlet, A. F.; Hellmann, J. J. Urban Meteorological Modeling Using WRF : A Sensitivity Study. *Int. J. Climatol.* **2017**, 37 (4), 1885–1900. DOI: 10.1002/joc.4819.

Wolff, D. B.; Petersen, W. A.; Tokay, A.; Marks, D. A.; Pippitt, J. L. Assessing Dual-Polarization Radar Estimates of Extreme Rainfall during Hurricane Harvey. *J. Atmos. Ocean. Technol.* **2020**, 36 (12), 2501–2520. DOI: 10.1175/JTECH-D-19-0081.1.

Ye, H.; Fetzer, E. J.; Wong, S.; Behrangi, A.; Olsen, E. T.; Cohen, J.; Lambrigtsen, B. H.; Chen, L. Impact of Increased Water Vapor on Precipitation Efficiency over Northern Eurasia. *Geophys. Res. Lett.* **2014**, 41 (8), 2941–2947. DOI: 10.1002/2014GL059830.

Zeng, Y.; Cao, Y.; Qiao, X.; Seyler, B. C.; Tang, Y. Science of the Total Environment Air Pollution Reduction in China : Recent Success but Great Challenge for the Future. *Sci. Total Environ.* **2019**, 663 (24), 329–337. DOI: 10.1016/j.scitotenv.2019.01.262.

Zhang, W.; Villarini, G.; Vecchi, G. A.; Smith, J. A. Urbanization Exacerbated the Rainfall and Flooding Caused by Hurricane Harvey in Houston. *Nature* **2018**, 563, 384–389. DOI: 10.1038/s41586-018-0676-z.

Developing a Control Algorithm and Simulation for Thrust Vector Controlled Rockets

Daniel Kim [a*]

Andres Iturregui [a]

[a)] Los Alamos High School, Los Alamos, NM, USA

ABSTRACT

In the past five years, SpaceX has revolutionized the aerospace industry by introducing a new rocket that can propulsively land. At the heart of these rockets, and virtually all orbital and suborbital spacecraft, is thrust vector control (TVC). The purpose of this project was to develop a robust and versatile control algorithm for TVC rockets and to validate this control system using a simulation and a real-life model of a powered model rocket with TVC capabilities. The simulation was able to recreate a rocket's flight based on physical aspects such as the mass moment of inertia, thrust of the engine, drag coefficient, and the control authority of the thrust vector control mount. The real-life model validated the simulation after cross-referencing the simulation results with data collected from the actual flight. Using these models, a control system and simulation capabilities for use on different rockets were developed.

KEYWORDS: Thrust-vector-control, PID Control, Simulation, Control Systems.

1 INTRODUCTION

Currently, there are over 10,000 companies in space technology development with a combined value of over $4 trillion (Koetsier, 2021) dedicated to advancing technologies in navigation, tourism, national security, communication, and outer space research. This expansive growth prompts the need for new suborbital and orbital-class vehicles to transport these technologies into outer space. Although these vehicles, the majority being rockets, present multiple challenges, one of the biggest obstacles is developing a navigation and control system to guide them.

A majority of these spacecraft use a technique called thrust vector control (TVC) (Figure 1). By angling or vectoring the direction of the thrusting component, the rocket has control over position and orientation, even in a non-atmospheric environment (Hall, 2021). Engineers have designed and implemented gimbals into the rocket engines to allow for this technique across a variety of different rockets, but the software to control these gimbals and the rockets becomes very complicated because of the high speeds, large masses, and precision required from spacecraft. Furthermore, methods to tune these control algorithms also become very complex due to the high cost and the large number of variables involved in a rocket flight. Although a majority of rockets use the same technique as TVC for active control, the control system design and tuning present many challenges.

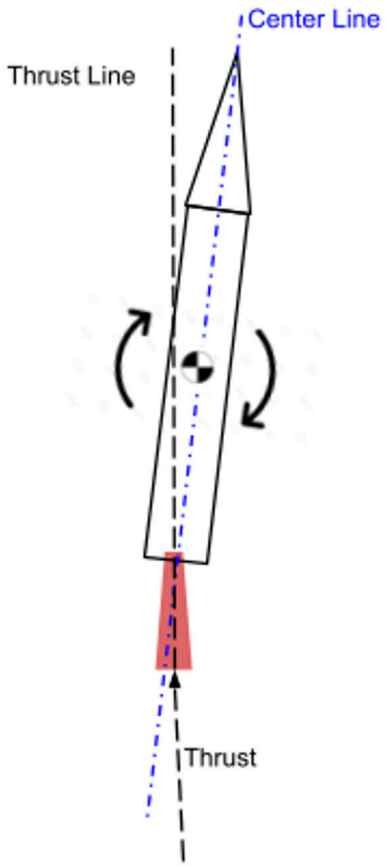

Figure 1. By vectoring the thrust, the rocket's orientation is rotated about the vehicle's center of mass. This method can be used to control the orientation and position of a model rocket.

The purpose of this research and engineering project was to design a versatile control system that is compatible with different dynamics of different rockets and to create a simulation to tune this control system and predict the flight dynamics of different rockets. Implementation of this robust and dynamic control system and simulation offers a simpler way for the growing space industry to solve one of the most challenging problems of guidance, navigation, and control. In this project, a control algorithm, orientation scheme, and simulation were developed and tested in the form of a model rocket with a TVC system and control system implemented.

1.1 Control Theory

The control system that we decided to use and tailor to rocket guidance, navigation, and control, or GNC, was the very widely used Proportional-Integral-Derivative (PID) controller. This is a feedback controller that regulated steam engines during the industrial revolution, improved the yield from windmills and industrial processes worldwide, and lies at the heart of autopilots used in commercial airplanes. This control algorithm allows for a simple integration and tuning process for all rockets of various sizes and purposes by adjusting three values. The PID controller is defined as

$$u(t) = K_p e(t) + K_i \int e(t)\,\Delta t + K_d \frac{de}{dt} \tag{1}$$

Where $u(t)$ is the output, K_p, K_i, and K_d are the three gains, $e(t)$ is the error value and the input to the controller, and Δt is the change in time. There are three terms to this equation: the proportional, integral, and derivative of the error value $e(t)$, which are scaled using the adjustable gains K_p, K_i, and K_d. The proportional, integral, and derivative components allow for control over oscillations, overshoot, and small errors. This simplicity allows for use on lower-powered flight controllers, but the three components work together to produce a very robust, efficient, and versatile control system.

On a rocket, this method can be easily integrated. We can have three PID controllers, one for each axis to control orientation. The error, which is inputted into the controller, is equal to the setpoint subtracted from the current orientation. The setpoint is adjustable based on the desired angle. After performing the proportional, integral, and derivative components, we can then command these angles to the gimbal, or thrust vector control mount (Figure 2).

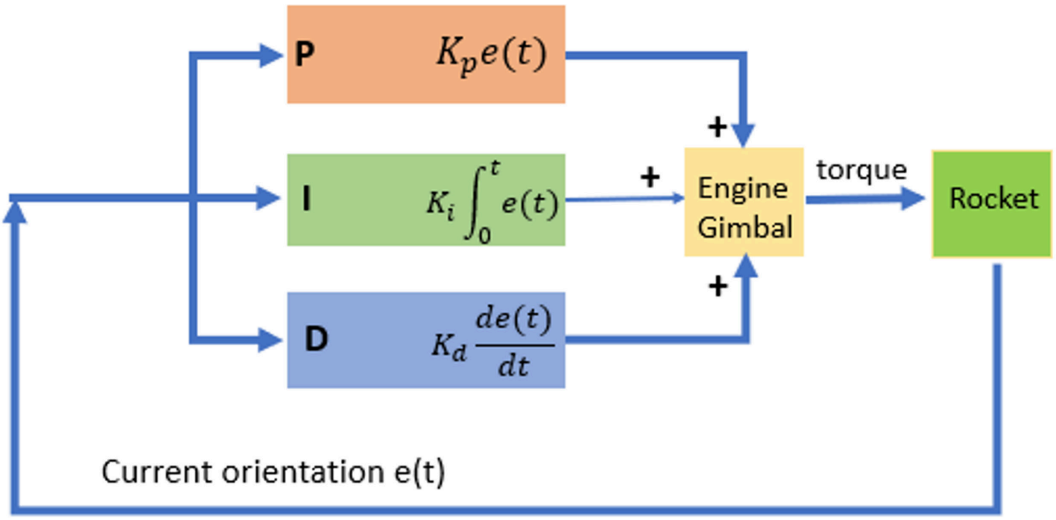

Figure 2. PID Controller for a rocket.

2 METHODS

2.1 Real-World Model

To validate that the control system and simulation worked properly, we designed a model rocket and a TVC gimbal in computer-aided design, CAD, and 3D-printed components to replicate a real-life rocket (Figure 3).

The rocket itself was printed in PLA and supported by four carbon fiber rods. Parachutes that were stored in the upper body were deployed using a small pyrotechnic charge ignited by a load driver aboard the flight computer.

The thrust vector control mount was a two-axis (Figure 4) gimbal that was designed to vector the rocket motor using two servo motors that were controlled by the flight computer. The gear ratio of the motors and the mount was 2:1, which allowed for more accuracy and less error within the mount, and the mount had a range of +/- 5 degrees on both axes.

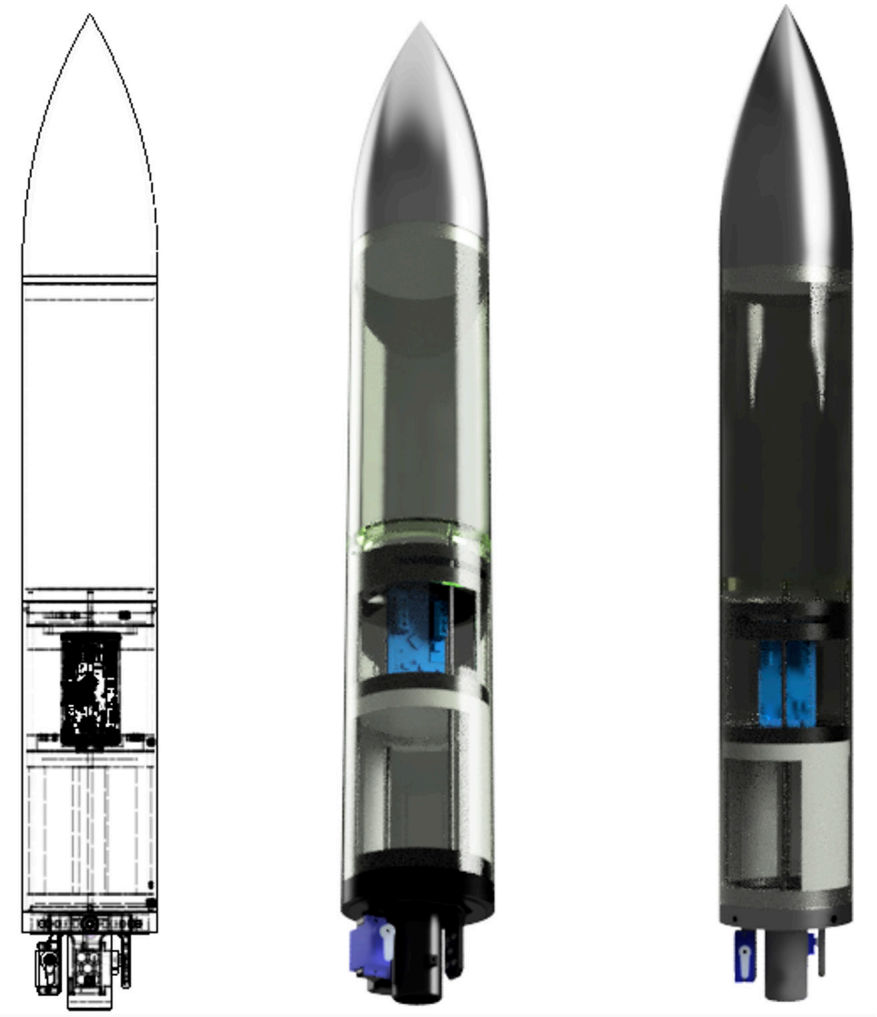

Figure 3. The model rocket design in CAD.

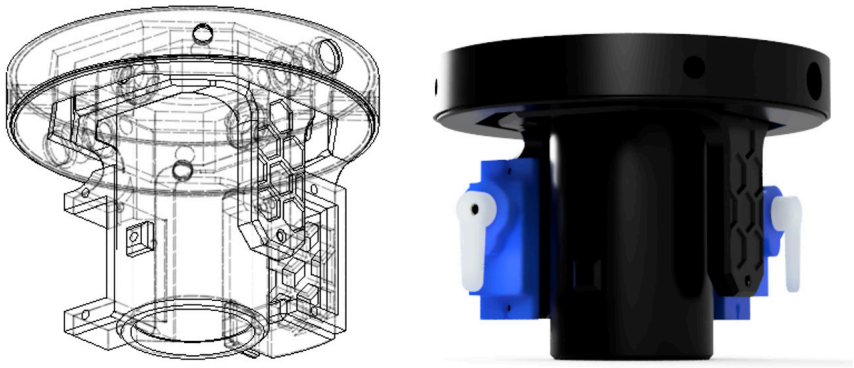

Figure 4. The thrust vector control gimbal.

To control the servo motors, calculate orientation, deploy recovery parachutes, log data, and execute the control algorithm, we designed and fabricated a custom flight computer (Figure 5). By collecting and analyzing sensor data, the flight computer was designed to control all aspects of our model.

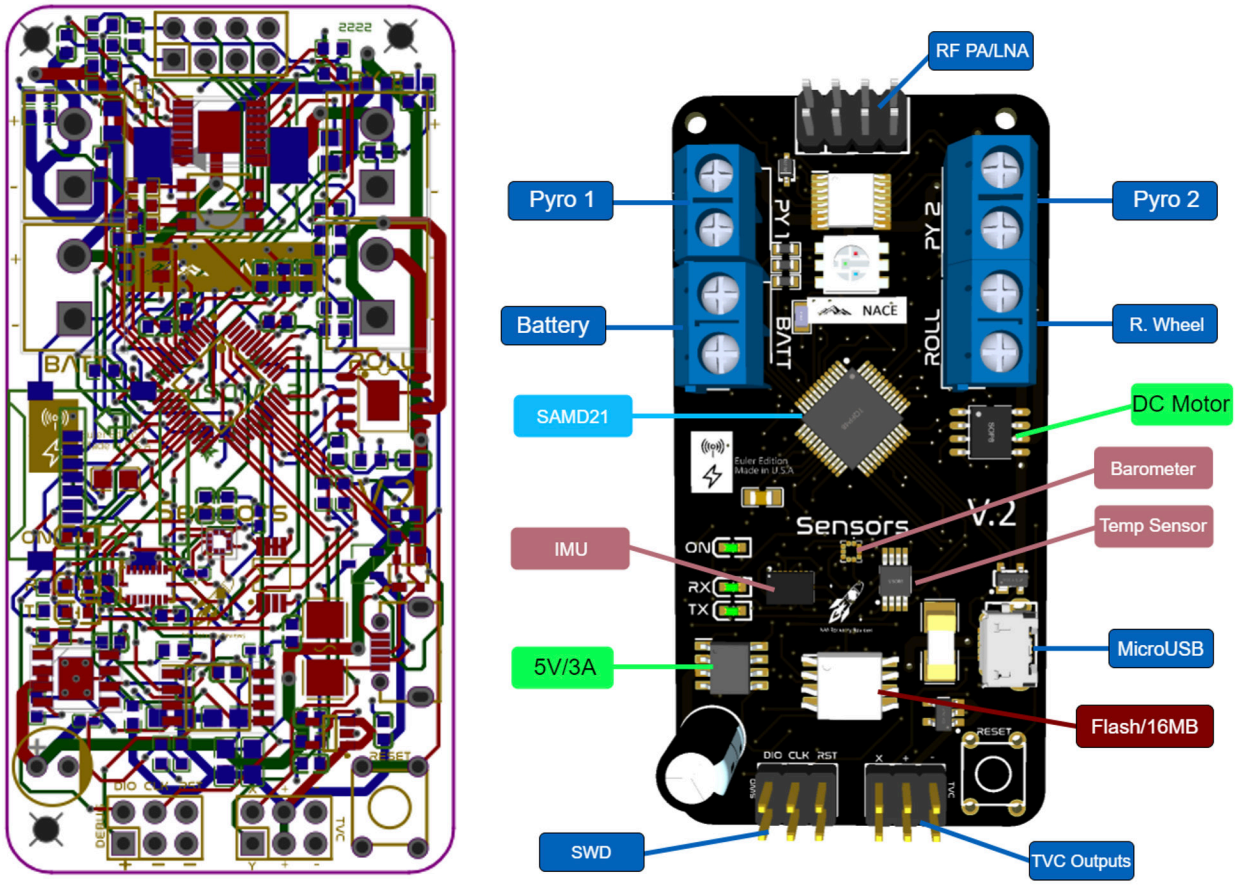

Figure 5. The flight computer in the design phase.

2.2 Flight Software Design

The flight software was written in C++ and controlled all the components of the flight computer and control system. The flight computer ran a finite-state automation function to control parachute deployment and data-logging tasks. To calculate the orientation of the craft using the angular rate measurements from the inertial-measurement-unit, or IMU, the flight computer used quaternions and a linear/1D Kalman Filter to estimate the orientation of the rocket.

2.2.1 Linear Kalman Filter

In our model, a linear Kalman filter is used to filter noise from sensor data (Looney). A Kalman filter calculates an estimate of unknown variables within measurements caused by external factors. Since velocity and position are calculated by integrating and double-integrating acceleration respectively, even a small amount of noise will cause significant drift in the measurements.

We define the Kalman Gain, which is

$$G_n = \frac{p_n}{p_n + r_n} \tag{2}$$

Where:

- G_n is the Kalman Gain where $0 \leq G_n \leq 1$
- p_n is the estimated uncertainty
- r_n is the measurement uncertainty

We then calculate the current estimate X_t by

$$X_t = X_{t-1} + G_n \cdot (Z_n - X_{t-1}) \tag{3}$$

Where:

- X_t is the current estimate at time step t
- X_{t-1} is the previous estimate
- Z_n is the measurement or the sensor data in our case

And the estimation uncertainty can be calculated using the Covariance Update Equation

$$p_n = (1 - G_n) \cdot p_n + |X_{t-1} - X_t| \cdot q \tag{4}$$

Where q is the process variance, which increases the accuracy based on how much the measurement moves. We can then set the previous estimate, X_{t-1}, to the current estimate

$$X_{t-1} = X_t \tag{5}$$

2.2.2 Quaternions and Conversion to Euler Angles

The simple integration of the angular rates provided by the IMU to Euler angles results in gimbal lock, where two axes align and differentiation between the two axes is not possible. Quaternions and quaternion algebra, which were invented by Sir William Rowan Hamilton, were used in this project to solve this issue and to allow for 3D rotations. Quaternions represent orientation in four dimensions in 3D space. Since the IMU outputs the angular rate of *x, y,* and *z* in *rad/s* so we can refer to those as

$$\omega_t = [0, x, y, z] \tag{6}$$

We can also refer to the base world reference frame, and since the rocket points upwards at launch, we can introduce Q_{init}

$$Q_{init} = [1, 0, 0, 0] \tag{7}$$

Next, we can calculate the quaternion derivative that describes the rate of change of orientation relative to the earth.

$$\frac{dQ_t}{dt} = \frac{1}{2} \cdot Q_{init} \otimes \omega_t \tag{8}$$

Where:

- dQ_t/dt is the quaternion derivative at time step t
- Q_t is the orientation at time step t
- $\otimes$ is the Hamilton Product operator (Appendix B)

We can integrate the quaternion derivative to determine the orientation at t with

$$Q = Q_{init} + \Delta t \cdot \frac{dQ_t}{dt} \tag{9}$$

And Δt is the time step. We now normalize the quaternion by first calculating the norm, and then dividing the orientation by the norm

$$Q = \frac{Q}{\sqrt{{Q_1}^2+{Q_2}^2+{Q_3}^2+{Q_4}^2}} \tag{10}$$

Where Q_1, Q_2, Q_3, Q_4 are the individual elements of quaternion Q. This quaternion can be converted back to Euler angles without the risk of gimbal lock by equations 11-13.

$$\alpha = tan^{-1}\left(\frac{2 \cdot Q_3 Q_1 - 2 \cdot Q_2 Q_4}{1 - 2 \cdot {Q_3}^2 - 2 \cdot {Q_4}^2}\right) \tag{11}$$

$$\beta = sin^{-1}(2(Q_2 Q_3 + 2 \cdot Q_4 Q_1)) \tag{12}$$

$$\theta = tan^{-1}\left(\frac{2 \cdot Q_2 Q_1 - 2 \cdot Q_3 Q_4}{1 - 2 \cdot {Q_2}^2 - 2 \cdot {Q_4}^2}\right) \tag{13}$$

Where α is the roll of the craft, β is the yaw of the craft, and θ is the pitch of the craft expressed in radians (Figure 6)

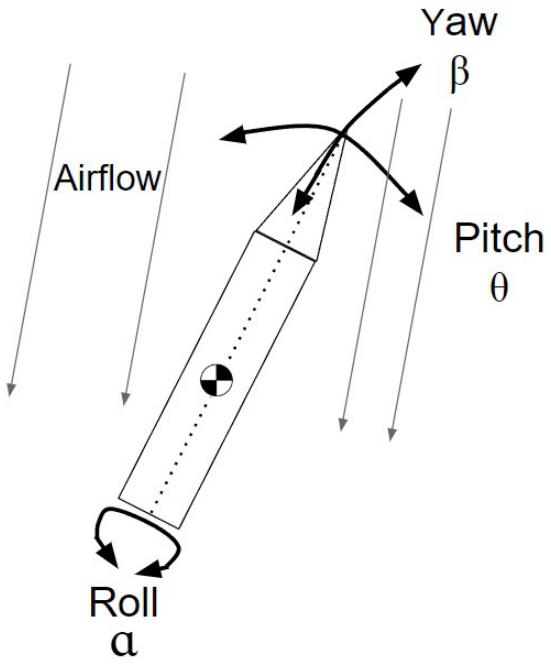

Figure 6. Rocket Rotations

To reduce unnecessary complexity, the real-world model did not include roll control, which is not controllable by the TVC mount. To account for any roll changes during the flight, θ and β must be decoupled from α using

$$\theta = cos(\alpha) \cdot \theta - sin(\alpha) \cdot \beta \tag{14}$$

$$\beta = cos(\alpha) \cdot \beta + sin(\alpha) \cdot \theta \tag{15}$$

θ and β can then be inputted into the two PID controllers for both the pitch and yaw axis.

Quaternions are also extremely useful in rotating different orientations. The accelerometer unit on the IMU provides linear acceleration in the vehicle reference frame. This reference frame is not compatible with velocity and position measurements, as well as accelerometer measurements used to detect launch, burnout, and apogee for the finite state automation function. By finding the Hamilton product of the orientation quaternions and the acceleration readings, we can convert the measurements to a world reference frame. We can refer to the accelerometer readings as

$$a_t = [0, a_x, a_y, a_z] \tag{16}$$

Then, we can find the world reference frame orientation by finding the first Hamilton product of the orientation quaternion and a_t.

$$A_{first} = Q \otimes a_t \tag{17}$$

Next, we find the Hamilton product of A_{first} and $—Q$ to flip the reference frame to the correct orientation

$$A = A_{first} \cdot -Q \tag{18}$$

Where A_2, A_3, and A_4 the 3 of the 4 components of quaternion A, represent the x, y, and z linear accelerations in the world frame respectively.

In conclusion, the flight computer:

1. Reads raw data from sensors.
2. Filters the data using the linear Kalman filter.
3. Converts the raw gyro and time parameters into Euler angles using quaternions.
4. Inputs the Euler angles into the PID controller and commands the output of the PID controller to the servo motors of the TVC mount.
5. Rotates the filtered acceleration readings to the world frame using quaternion operators and uses these values for the state automation function.

2.3 Simulation

To tune the PID controller and verify that our real-life system would work properly, we have created a simulation of the rocket in MATLAB Simulink. We worked to account for many real-life variables, including:

- Noise from sensor data
- Latency from flight computer and servo motor
- Wind disturbances

The most fundamental part of the simulation is the 3DOF block, as it utilizes all of the equations in three degrees of freedom motion to simulate an object in 2-dimensional space. The

3DOF block allows us to input x and z direction vector forces and a torque parameter, then outputs the x and z position and velocity, along with an angular position and velocity. To accurately calculate the torque applied by the TVC mount and from wind disturbances, we needed to know the mass moment of inertia (MMOI) and mass of our actual rocket. To find the MMOI, we used the bifilar pendulum method, which involves hanging the rocket by two parallel strings equidistant from the center of mass and measuring the period of the swing rotating around the center of mass. The equation to find the MMOI is:

$$MMOI = \frac{mgp^2r^2}{4\pi^2L} \tag{19}$$

Where *MMOI* is the mass moment of inertia, m is mass, g is the gravitational constant, p is the period for one complete swing of the rocket, r is the distance from the center of mass and where the string is attached, and L is the length of the strings.

2.3.1 Engine Force and Gravity

The two most critical forces being applied to a rocket is the gravity and the engine force. The 3DOF block already has a feature for applying gravity, which makes the implementation extremely simple. As for the engine force, the data for the D12-0 engine we are using is already provided by Estes, so we were able to implement that data into the simulation. However, since we are using a TVC mount, we didn't just want the engine force to be applied in the z-direction, as it needed to be vectored based on the TVC mount's position. To solve this, we split the force into x and z vectors using the following two equations:

$$F_{motor(z)} = F_{motor} \times cos(\sigma + e) \tag{20}$$

$$F_{motor(x)} = F_{motor} \times sin(\sigma + e) \tag{21}$$

In these equations, $F_{motor(x)}$ and $F_{motor(z)}$ are the x and z direction vector forces respectively, F_{motor} is the total engine force, and σ is the angle the TVC mount. The motor mount misalignment, represented by e is to 1 degree to account for any real-life misalignments that may occur. We also used this equation

$$T_{motor} = F_{motor(x)} \times a \tag{22}$$

To calculate the torque produced by the TVC mount, where T_{motor} is the torque produced by the engine and a is the distance from the center of mass to the end of the rocket motor. The way the TVC mount position is calculated will be explained later in the simulation overview.

2.3.2 Noise Addition and PID Calculations

Since the accelerometer used on the real rocket does not yield perfect results, we wanted to simulate some of that noise. A random number generator generating values of +/- 0.1 degrees was added to the actual angular position of the rocket in degrees.

To calculate the PID values, we use the PID equation. To calculate the output, this equation must be the same in both the real-world model and the simulation because we want the K_p, K_i, and K_d, gains found in our simulation to be used in the real-life model with optimal results.

2.3.3 Real-Life Limitations

Next, we need to account for some of the real-life limitations of the rocket, such as flight

computer delay, servo speed limitations, and servo position limitations. The TVC control mount on the real rocket can only move the engine ±5 degrees, so this is accounted for by saturating the output *u(t)* to have a maximum of 5 degrees and a minimum of 5 degrees. In addition, the TVC servos have a limited angular rate. To measure the speed of the servo, we recorded a slow-motion video of the servo moving as fast as it can and calculated the maximum rotational speed to be 258 degrees per second. We used this value as the maximum rate at which the PID output could change. Finally, flight computer latency needs to be accounted for. Although the flight computer latency is estimated to be only 7.5ms, we decided to increase it to 50ms to have the simulation account for more imperfection. After the three of these limitations are applied to *u(t)*, the final resulting angle will be the position of the TVC mount, σ.

2.3.4 Wind Forces

To account for wind that may occur in real flight, we applied torque in the form of a sine wave to the rocket to simulate changes in wind speed. By adjusting the amplitude of the sine wave, we were able to change the maximum amount of torque applied to the rocket body, and by adjusting the frequency of the sine wave, the variance of the wind could be changed.

For the simulation, we used the maximum amount of wind possible–if the rocket could handle very little wind if it could handle lots of wind. After experimenting with different amounts of force, the maximum wind force the rocket could handle was a wave with an amplitude of 0.1N·m.

2.3.5 Aerodynamic Forces

The two most important aerodynamic forces that act on the rocket are the drag and lift force. The drag force acts opposite to the direction the rocket is pointing towards, and the lift force acts perpendicular to the rocket but is applied from the center of pressure. We were able to implement these forces with the help of software called OpenRocket. OpenRocket is a software that allows for the simulation of model rocket launches and lets the user customize the rocket. We used this software and created our real-life rocket, ensuring that parameters such as the mass, center of mass, MMOI, and shape of the rocket were all the same. Then, we exported a large amount of data from OpenRocket for various amounts of wind, including the angle of attack, total velocity, drag force, lift force coefficient, and center of pressure position. The angle of attack is the angle at which the rocket is relative to the air flowing past it as opposed to the ground. All of this data was then implemented into the simulation. However, we only received the lift force coefficient, but not the lift force. The equation to find the lift force is

$$F_l = \frac{1}{2} \times c_l \times d_{air} \times a_{ref} \tag{23}$$

Where F_l is the lift force, c_l is the lift force coefficient, d_{air} is the density of air, and a_{ref} is the reference area of the rocket. The density of air is 1.225 kg/m^3 at sea level, and the reference area is $5.6745 \cdot 10^{-3}$ m^2 according to OpenRocket. To find the torque produced by the lift force, we multiply the lift force by the distance between the center of pressure and the center of gravity, which varies based on the angle of attack of the rocket.

2.3.6 Tuning for Optimal Results

After each component of the simulation was completed, the next step was to tune the PID values K_p, K_i, and K_d. To begin tuning, we set K_i to 0 since its effect was not determined to be significant and began experimenting with K_p and K_d values. We knew that low K_p gains would result in insufficient correction, and high K_p values would result in significant overshooting. K_d gains that were too low would result in overshooting even with low K_p values, and K_d values too high would result in extremely rapid oscillations. When we eventually found a K_p and K_d value combination that resulted in a successful flight, we began fine-tuning these values by moving each of them up and down by small increments individually and seeing if the flight had less or more error. Once we found the K_p and K_d values with the smallest amount of error, we added back the K_i term and increased it gradually until it hindered the results compared to lower values.

3 RESULTS AND DISCUSSION

3.1 Simulation

Once tuning was complete, we got PID values of K_p=0.5, K_i=1.4, and K_d=0.08 inside of the simulation. As seen in Figure 7, the rocket coasts up to an altitude of about 18 meters before falling back down to the ground. Figure 8 shows the angle of the rocket as a function of time,

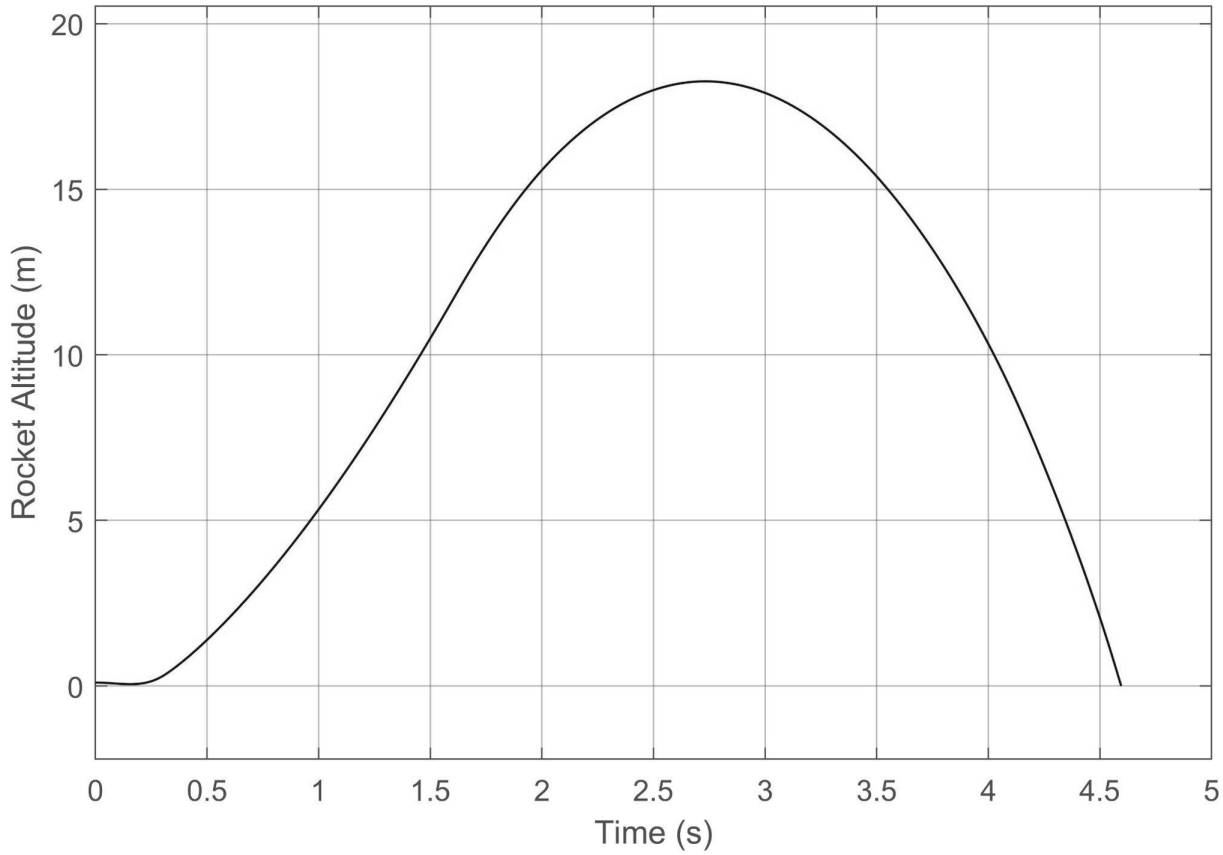

Figure 7. Shows the altitude of the simulated rocket as a function of time.

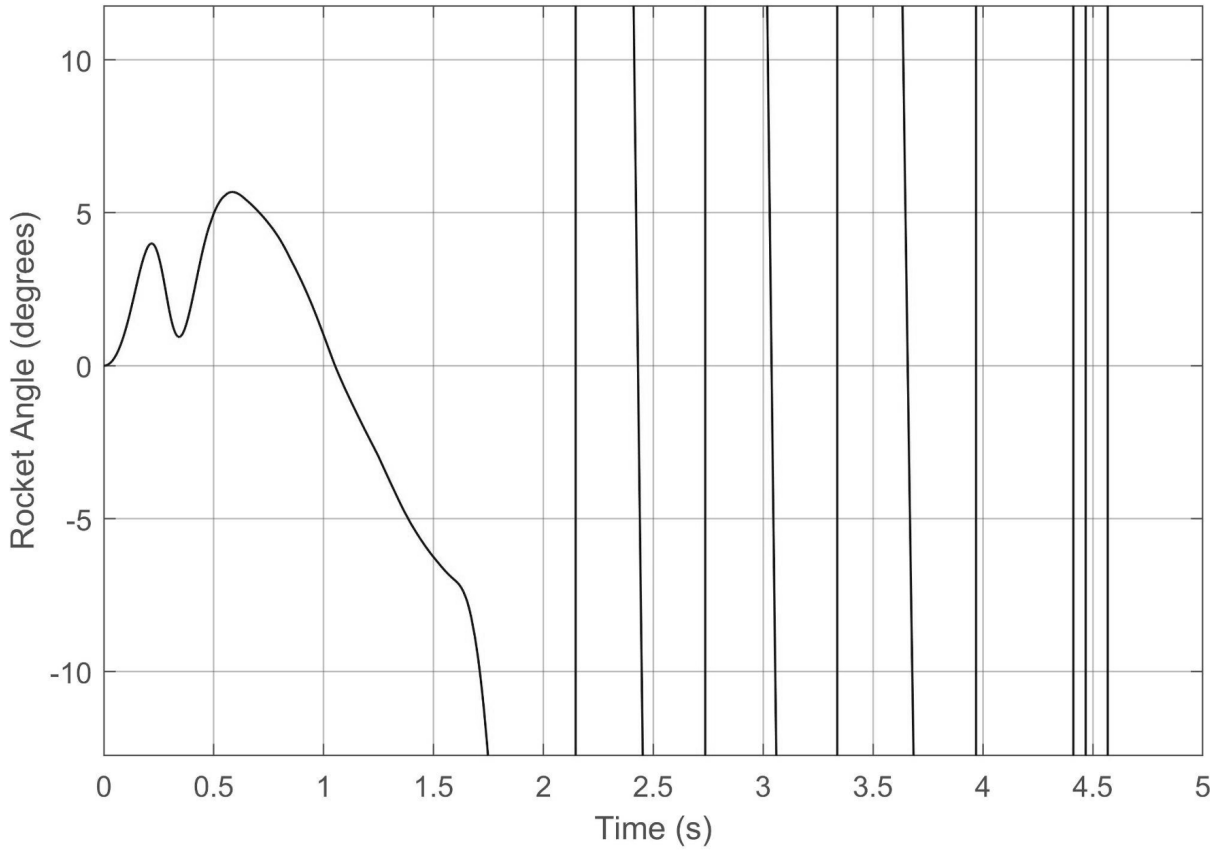

Figure 8. Shows the rocket's angle relative vertical in degrees as a function of time.

which is much more important than the altitude. From 0-1.6 seconds, the TVC mount can handle the wind and prevents the rocket from exceeding an angle of 10 degrees, after this, the rocket motor burns out and is uncontrolled.

At the time we flew the rocket, the simulation did not account for as many drag forces, so the PID gains were instead K_p=0.3, K_i=0.3, and K_d=0.05. These PID gains were then inputted into the real-world model's flight software and flown on the Estes D12-0 rocket motor. For simplicity purposes, the software had a constant setpoint of zero on both axes, meaning that the rocket would try to stay fully upright.

3.2 Real-World Model

The real-world model as discussed above was flown once using the parameters calculated from the simulation. With a height of 412 mm and a diameter of 85 mm, the vehicle was flown on an Estes D12-0. The rocket had a mass of 0.605 kg and a mass-moment-of-inertia of 0.01428 kgm^2.

Figure 9 shows the estimated orientation of the rocket after liftoff was detected. The rocket almost immediately pitched over after liftoff. At around 0.372 seconds, data logging stopped and the PID loop was terminated because the flight software had an integrated abort system that triggered when the pitch or yaw surpassed +/- 30 degrees. This abort sequence saved all the data into the SD card, detached electrical connections to the servo motors, terminated all

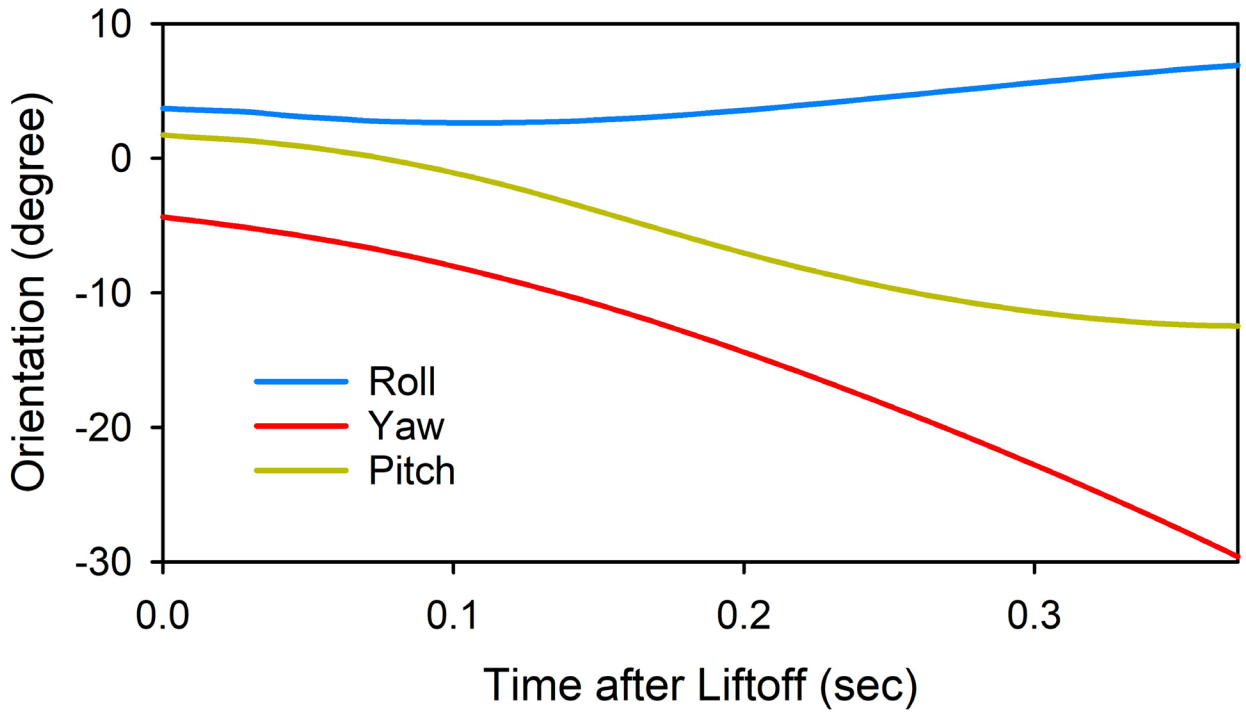

Figure 9. Orientation of the rocket in degrees as a function of time after liftoff was detected by the finite state automation function.

functions, and deployed the parachutes. This sequence ensured that all relevant data would be saved and that damage to the rocket and avionics would be minimized when the rocket's orientation surpassed a point of no return.

The large undesired orientation change was caused by several issues within the physical model. When assembling the thrust vector control gimbal, there were misalignments within the two axes. Therefore, at liftoff, the rocket immediately pitched over, as seen in Figure 9. As the PID controller tried to correct this error, the misalignment caused larger errors to occur. As can be seen in Figure 10, the PID controller tried to output a value much greater than 5 degrees, and subsequently kept outputting values greater than 5 degrees. The TVC mount had an actuator limit of +/- 5 degrees, which means it was not accurately reflecting the command output of the PID loop.

Datalogging occurred when the accelerometers detected a significant upwards acceleration. Figure 9 shows how the calculated orientation was not 0 degrees in all axes upon liftoff, despite being in a stationary position. This drift was most likely caused by small amounts of gyroscope noise being integrated into large errors over time. This large error in orientation readings was then fed into the PID control loop and further caused the large pitch over.

After the abort was called, the rocket flew completely out of control and collided with the ground. As seen in Figure 11, the rocket immediately angled sideways and spun out of control. Although the rocket's outer body frame was shattered and the TVC mount was broken, the flight avionics, data, and servo motors were still intact. Figure 12 depicts other parameters measured during the flight, where X, Y, and Z are the roll, yaw, and pitch axes respectively. All measurements were plotted as a function of time detected after liftoff.

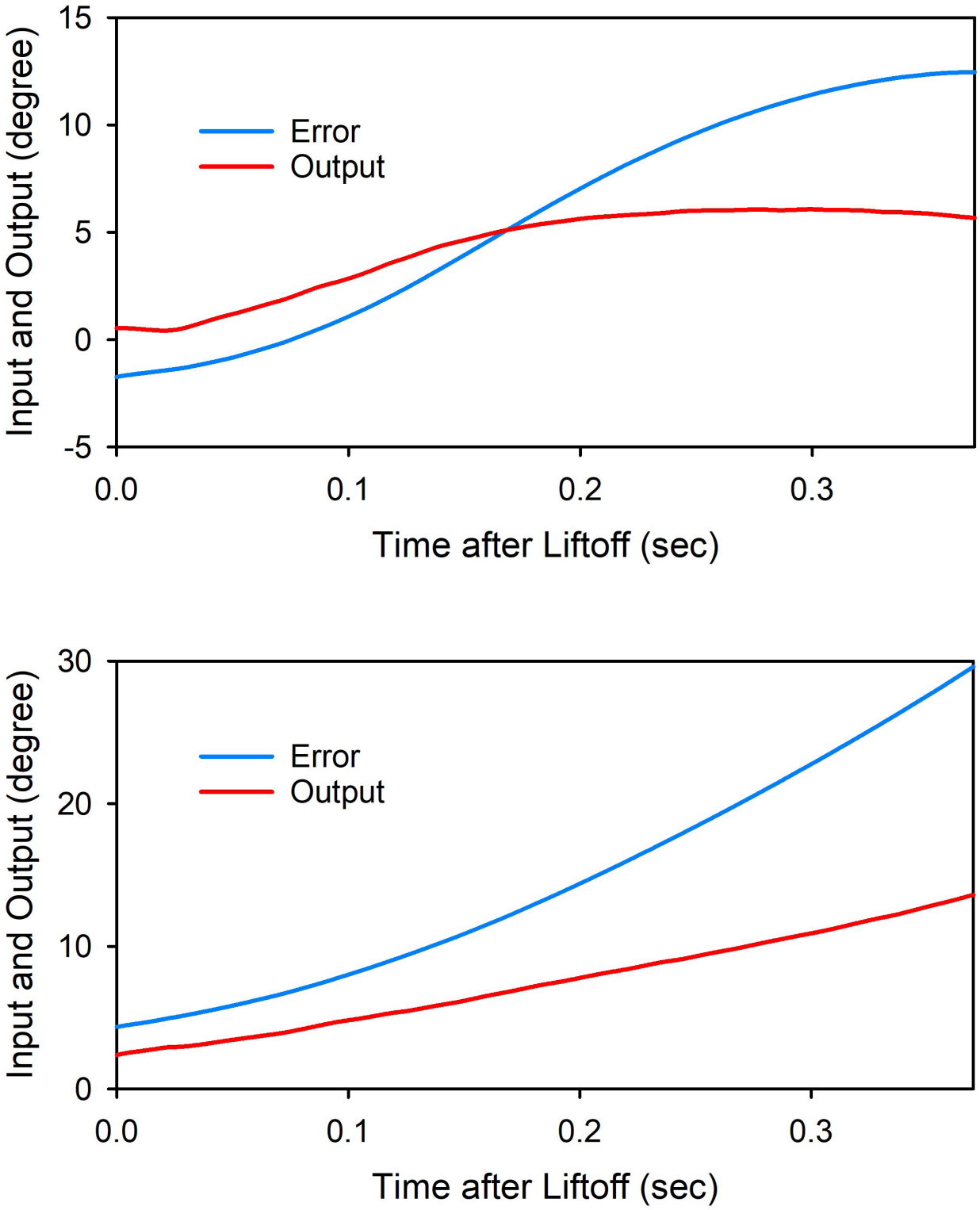

Figure 10. Error inputted to the PID controller versus output as a function of time detected after liftoff.

Figure 11. Different phases of the rocket flight are depicted in one photo. By the second and third freeze frames, the abort sequence had already fired, and the rocket was uncontrolled.

Figure 12 also depicts the raw versus filtered linear acceleration readings in the vehicle reference frame. The raw readings are passed into separate Kalman filters, which smooth the data. However, the Kalman filter over filtered the data, which was most likely caused by measurement uncertainty, r_n, that was too large. Although the filter may have provided skewed data, the accelerometer data was not used for any active control aspects of the flight, therefore, these incorrect readings did not contribute to the lack of control throughout the flight.

3.3 Model and Simulation Verification

Although the flight did not take the desirable flight path due to errors in the physical model, we cross-referenced data from the real-world model's flight to the simulation that was run of that model. Figure 15 depicts the simulated flight orientation versus the actual flights in the Y and X axes (yaw and pitch respectively).

4 CONCLUSIONS

The results of the simulation and model indicate that a PID controller-based navigation system would be a viable option for rockets to use as a control system. Using an accurate simulation that matched real-world data (Figure 15), the three tuning parameters of the PID system could be calculated for an optimal flight. Furthermore, a quaternion-based orientation system was developed to solve gimbal lock and 3D space orientation problems, along with an implementation of a Kalman filter to reduce the effects of noise and increase the robustness of the overall system.

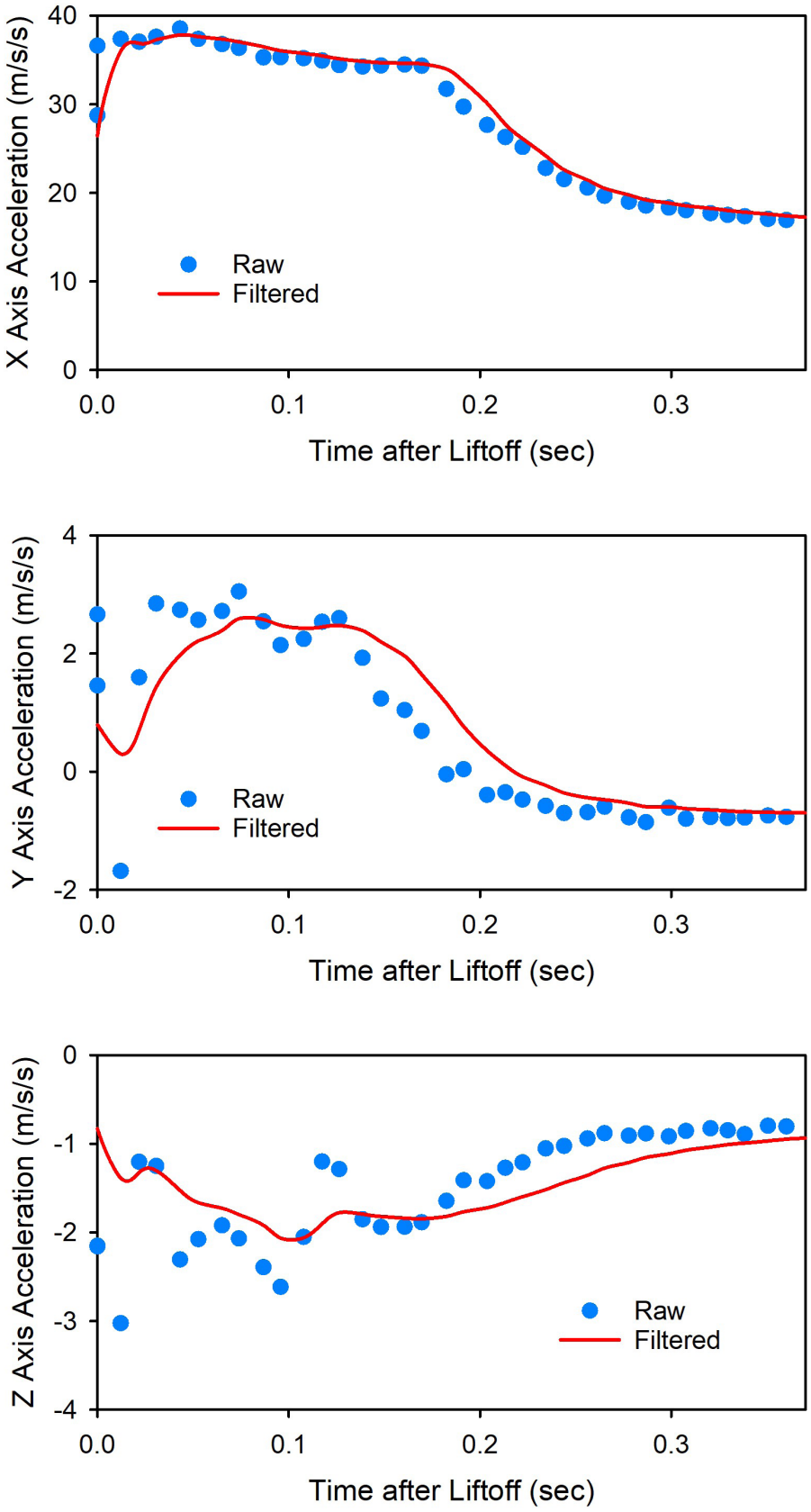

Figure 12. Raw and filtered accelerometer readings.

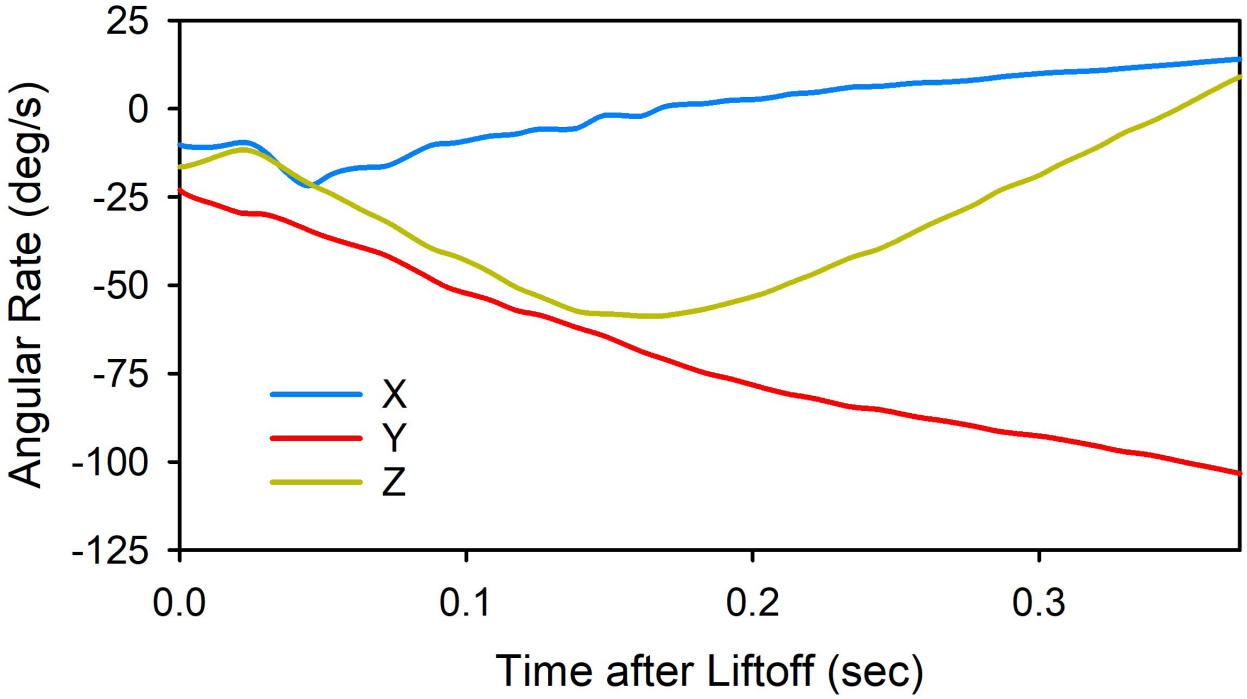

Figure 13. X, Y, Z (roll, yaw, pitch) readings from the gyroscope.

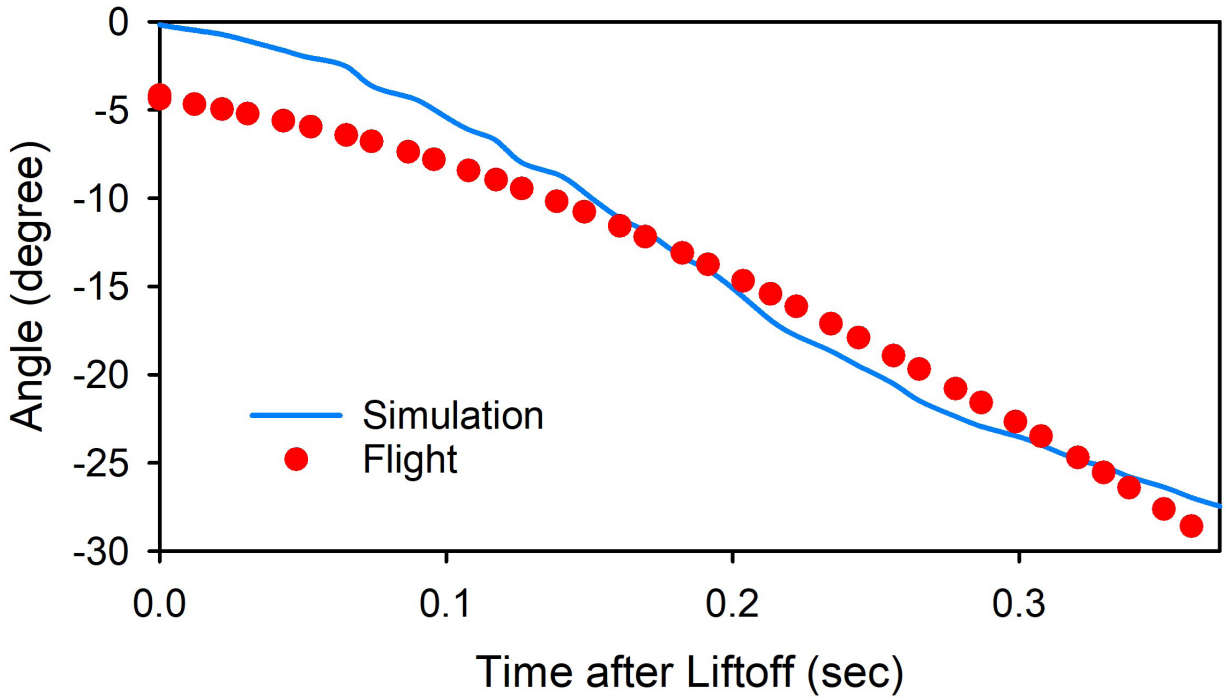

Figure 14. Y-axis (yaw) orientation of the simulated flight of the real-world model and the actual flight.

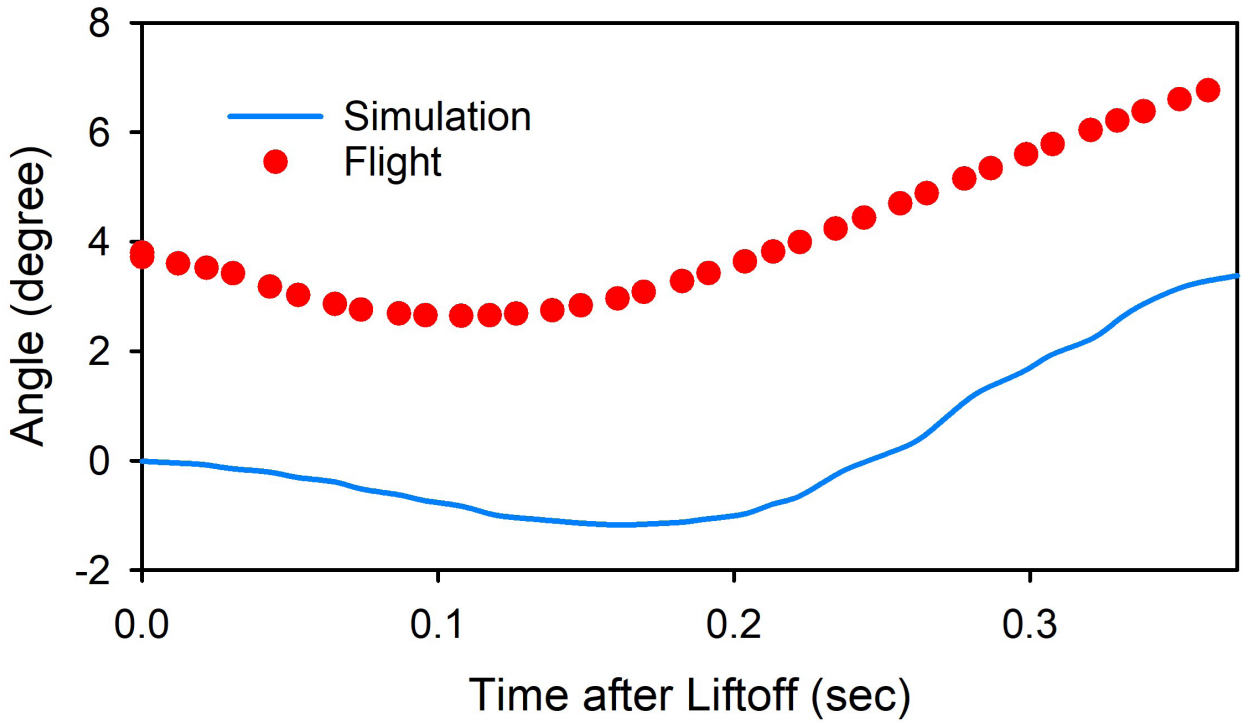

Figure 15. Z-axis (pitch) orientation of the simulation versus the actual flight. The real-world model's orientation values have drifted up about 4 degrees due to the integration of small noise values, causing an offset.

Although the real-world model did not fly as expected, the analysis showed that this was caused by misalignment of the parts and by design errors. We hope to fly another version of the real-world model soon after conducting a few improvements to the model and further improving the simulation for more accuracy.

4.1 Simulation Improvements and Future Implementations

The simulation worked very well at predicting the flight path and properties of the real-world model but also has some future implementations and improvements such as:

- Implementing the actual formulas for the fluid dynamic calculations.
- Adding a visualization of the flight using animations.

4.2 Real-World Model Improvements and Future Implementations

Software

- Faster loop speeds by optimizing the available SRAM onboard the flight computer and creating more efficient functions.
- Add anti-windup to the integral term by clamping the output to the actuator limit of the TVC mount.
- Experiment with different methods of control such as model-predictive control or a

linear quadratic regulator.

- Potentially implementing roll control with a reaction wheel or adding position control.
- A dynamic setpoint that pitches the rocket in a certain direction.

Hardware

- Adding a self-alignment system in the form of a hex driver that inserts into the side.
- Moving avionics upward will lift the center of mass, giving the TVC more torque authority by increasing the lever arm.

Finally, the fact that this project is not the end cannot be overemphasized. The real-world model, control system, simulation, and flight software can have many new improvements, and with more research and development, we hope to see a larger growth within the space industry.

AUTHOR INFORMATION

***Corresponding Author**

Daniel Kim
Los Alamos High School
1300 Diamond Drive
Los Alamos, New Mexico 87544
Email: daniel.kim@studentlaschools.net

ACKNOWLEDGMENT

We would like to express our gratitude to our mentor Dr. Robert Hermes for his invaluable advice and teachings in model rocketry. We thank our parents for their unending support and our sponsor Dr. Ombelli for her patience.

REFERENCES

Becker, A. Online Kalman Filter Tutorial. https://www.kalmanfilter.net/kalman1d.html (accessed Feb 22, 2022).

Center of pressure. https://www.grc.nasa.gov/www/k-12/airplane/cp.html (accessed Feb 22, 2022).

D12-0 engines. https://estesrockets.com/product/001565-d12-0-engines/ (accessed Feb 22, 2022).

Four forces on a model rocket. https://www.grc.nasa.gov/www/k-12/VirtualAero/BottleRocket/airplane/rktfor.html (accessed Feb 22, 2022).

Koetsier, J. Space inc: 10,000 companies, $4T value ... and 52% American. https://www.forbes.com/sites/johnkoetsier/2021/05/22/space-inc-10000-companies-4t-value--and-52-american/?sh=4861da5955ac (accessed Feb 22, 2022).

The lift equation. https://www.grc.nasa.gov/www/k-12/VirtualAero/BottleRocket/airplane/lifteq.html (accessed Feb 22, 2022).

Looney, M. Anticipating and managing critical noise sources in MEMS gyroscopes. https://www.analog.com/en/technical-articles/critical-noise-sources-mems-gyroscopes.html (accessed Feb 22, 2022).

Maths - conversion quaternion to Euler. https://www.euclideanspace.com/maths/geometry/rotations/conversions/quaternionToEuler/ (accessed Feb 22, 2022).

Maths - quaternions. https://www.euclideanspace.com/maths/algebra/realNormedAlgebra/quaternions/ (accessed Feb 22, 2022).

Niskanen, S. OpenRocket technical documentation. http://openrocket.sourceforge.net/techdoc.pdf (accessed Feb 22, 2022).

Vectored thrust. https://www.grc.nasa.gov/www/k-12/airplane/vecthrst.html (accessed Feb 22, 2022).

APPENDIX A

Symbol	Description
TVC	Thrust-vector-control
PID-Controller	Proportional-Integral-Derivative Controller
GNC	Guidance, navigation, and control.
IMU	Inertial-measurement-unit
3DOF	Three degrees-of-freedom
MMOI	Mass moment of inertia
CAD	Computer-Aided Design
PLA	Polylactic Acid
a	Distance from end of engine and center of mass
A	World reference frame linear acceleration
a_t	Linear acceleration
c_l	Coefficient of lift force
a_{ref}	Reference area of the rocket
d_{air}	Density of air
$e(t)$	Error
e	Motor mount misalignment
F_l	Lift force

F_{motor}	The total force from the rocket motor
$F_{motor(x)}$	x-direction vector of the engine force
$F_{motor(z)}$	z-direction vector of the engine force
G_n	Kalman gain
g	Gravitational constant in m/s^2
K_d	Derivative gain
K_p	Proportional gain
K_i	Integral gain
L	Length of string
m	Rocket mass
p	Period of the rocket swing in the two-string pendulum
P_n	Estimate uncertainty
Q	Quaternion orientation
Q_{init}	Base world reference frame quaternion
Q_t	Quaternion orientation at time step t
q	Process Variance
r	Distance from center of mass and string
r_n	Measurement uncertainty
$u(t)$	PID controller output
X_t	Kalman filter estimate
Z_n	Measurement input
α	Roll
β	Yaw
θ	Pitch
σ	The angular position of the TVC gimbal
Δt	Change in time
ω_t	Angular rate

APPENDIX B

$\otimes$ The Hamilton product operator is the noncommutative multiplication of two quaternions. For quaternions a and b, $a \otimes b = i$ is defined as

$$i_1 = a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4$$
$$i_2 = a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3$$
$$i_3 = a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2$$
$$i_4 = a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1$$

CIRCA- CircularRNA for Cancer Active Immunotherapy: A Machine Learning Model to Predict Liver Cancer and Top Genes for Cancer Vaccine

Aditya Kiran Koushik [a*]

[a)] La Cueva High School, Albuquerque, NM, USA

ABSTRACT

Circular RNAs (circRNAs) are long non-coding RNAs with excellent prognostic and diagnostic biomarker properties for many diseases including cancer. By using liver tissues of Hepatocellular Carcinoma (HCC) patient dataset, this study designed and tested a robust machine learning pipeline to predict HCC and circRNA targeted hub gene immunogenicity for immunotherapy. First, a publicly available circRNA microarray dataset was analyzed in Python for the top twelve deregulated circRNAs in tumor tissue compared to healthy tissue. Next, classification models were trained and tested on the circRNA data. microRNA (miRNA) and gene targets (mRNA) of deregulated circRNAs were predicted and top hub genes were found from gene interaction network analysis in Cytoscape. Finally, an immunogenicity predictor in Python was built with a T-cell epitope prediction framework. This study found: 1) hsa_circ_0005284 is strongly upregulated in tumor tissue and hsa_circRNA_089372 is strongly downregulated, 2) the Logistic Regression and Naive Bayes classification models most accurately predicted tumors from circRNA data, 3) the *TMED10* and *RAB1A* hub genes were most immunogenic based on Python predictions. In conclusion, this project identifies hsa_circ_0005284 and hsa_circRNA_089372 as well as their linked immunogenic hub gene peptides as biological candidates for a liver cancer vaccine.

KEYWORDS: Liver cancer, circRNA, miRNA, T-cell epitope, Machine Learning, Artificial Intelligence, Bioinformatics, Immunogenicity, Oncogenes and Cancer Vaccine.

1 INTRODUCTION

1.1 Liver Cancer Pathophysiology

Liver cancer is a deadly disease where abnormal cell division occurs in the liver. Liver cancer remains a global health challenge as it is a leading cause of cancer-related death worldwide. Hepatocellular carcinoma (HCC) is the most common form of liver cancer, and it accounts for over 90% of cancer cases. Viral infection from hepatitis b and c are the most common causes of development, although non-alcoholic steatohepatitis associated with diabetes mellitus is also a frequent risk factor now rising in the west (Llovet et al., 2021). A few known mechanisms of HCC development/progression include genetic predisposition, viral and non-viral risk factor interaction, and cellular microenvironment alteration. Genes that play oncogenic (cancer-causing)/tumor suppressive roles can create changes in the cellular

microenvironment. With high throughput next-generation sequencing and artificial intelligence (AI) analysis, these genes can be identified and targeted (Balogh et al., 2016).

1.2 Screening, Diagnosis, and Prevention

Liver cancer can be diagnosed with a detection of a liver nodule in an abnormal ultrasonography test and with high serum α-fetoprotein levels (>20 ng/ml). Lesions less than 1 cm in diameter can be detected with ultrasonography, but for bigger lesions, a quadruple-phase CT scan or MRI is required. More and more methods to diagnose and screen HCC are being developed, which includes circular RNA (circRNA) biomarkers–which is being further explored in this study. Currently, HCC can be prevented from viral hepatitis vaccines, but there is no cure or vaccine for non-viral HCC (Llovet et al., 2021).

1.3 CircRNAs Functions and Mechanisms

CircRNAs are non-coding RNAs that form closed, continuous loops that and regulate genes in mammals. CircRNAs are generated via the back splicing of exons and introns to form exonic or intronic circRNAs (Greene et al., 2017). CircRNAs lack 5'-3' ends and poly a tails. The image in Figure 1 demonstrates the mechanisms and functions of circRNAs (Conn et al., 2015). Recent evidence implicated circRNA-mediated mechanisms in many cancers, including liver cancer (Liu, Zhang, Yan, & Li, 2020; Shen et al., 2021; Su et al., 2019; Zhang et al., 2019).

The present study is focused on using publicly available circRNAs expression data in HCC tumor tissue versus healthy tissue and identifying optimum machine learning model that correlate with tumor occurrence and likely predict tumor occurrence based on circRNAs expression levels. The study also extends to the test circRNAs -miRNA-mRNA/protein network for circRNAs target hub gene (protein) for potential cancer immunotherapy.

1.4 miRNA Sponging

Micro RNA (miRNAs) are small noncoding RNAs. miRNAs control gene expression by binding to the 3'-unstranslated region (3'-utr) in mRNAs and inhibit/suppress messenger RNA (mRNA) and translational processes. Many studies demonstrate that circRNAs contain miRNA response elements (MRES), which serve as miRNA sponges. CircRNAs can regulate gene expression via releasing miRNA to target genes (mRNAs). These mRNAs that are targeted/regulated by circRNAs via miRNA sponging will be referred to as circRNAs-targeted genes. CircRNAs that have a superior ability to bind to circRNAs are called "super sponges." Figure 2 explains the process in which circRNAs can cause cancer. Since miRNAs in turn regulate their target mRNAs, which can affect the protein expression, some of which may be oncogenes and others can act as tumor-suppressors; therefore, it is important to understand circRNAs -miRNA-mRNA-protein expression relationships for developing cancer therapies at different stages. Because proteins are the ultimate drivers of the disease process, it is also important to test if they can be directly targeted by immunotherapy via T-cell mediated blocking of such oncoproteins.

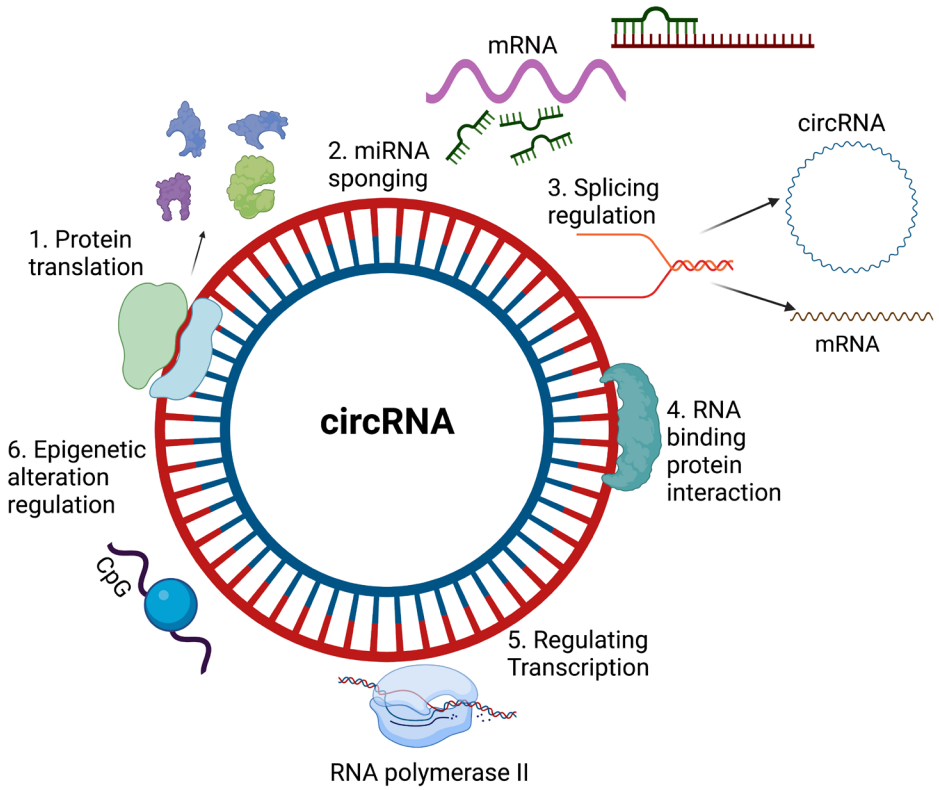

Figure 1. CircRNA interactions. 1. Protein translation–classic translational machinery can occur on circRNA producing proteins; 2. miRNA sponging–circRNAs can act to bind to miRNA and deploy when needed for anti-sense inhibition of complementary mRNAs; 3. Splicing regulation–to lead to production of circRNAs and mRNAs; 4. Interact with RNA binding proteins–circRNAs can interact with RNA binding proteins and can regulate post-transcriptional processes; 5. Regulation of transcription–circRNAs can regulate transcription of mRNAs. 6. Epigenetic alteration regulation–circRNA can also alter epigenetics. Created with BioRender.com.

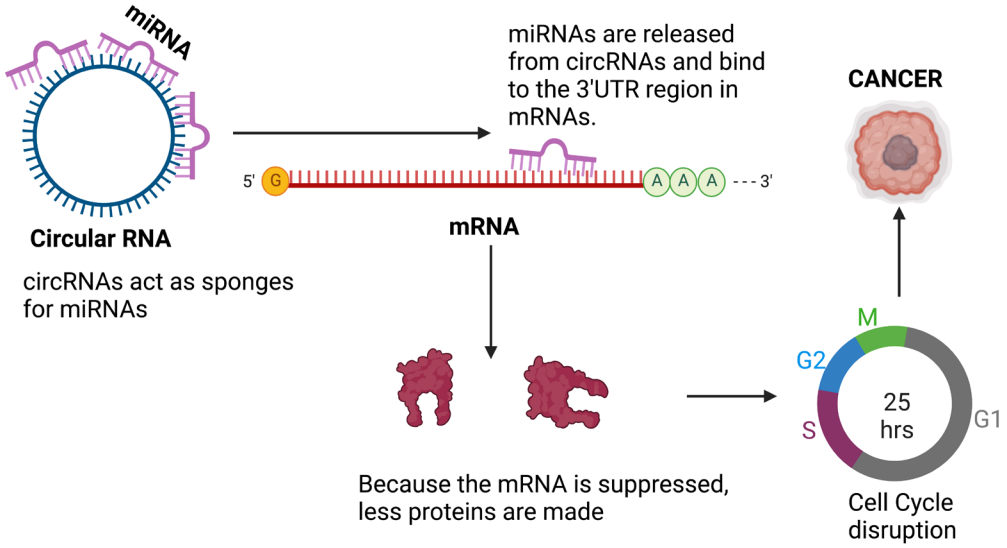

Figure 2. Schematic showing how circRNA may play a role in cancer. Created with BioRender.com.

1.5 T-cell activation and killing

T-cells are a type of leukocytic cell that play an essential role in the immune system (Henderson, 2021). T-cells originate in the bone marrow and are matured in the thymus. T-cell activation can involve antigen specific simulation. The T-cell receptor (TCR) binds to the antigen on the major histocompatibility (MHC) complex present on the surface of the antigen-presenting cell (APC) such as dendritic cells. This results in the activation of a killer T-cell (cytotoxic or CD8+ T-cell) which can kill tumor cells as shown by Figure 3 (Cavanagh, (n.d.)). After antigen simulation, CD8+ cytotoxic T lymphocytes kill tumor cells by secreting granzymes and perforins. Perforins allow granzymes to enter the cell by holes in the cell membrane, and granzymes activate the protease caspases, resulting in apoptosis. CD4+ helper T-cells perform several different functions (De candia, Prattichizzo, Garavelli, & Matarese, 2021). As an example, they produce cytokines which enhance effectiveness of cytotoxic T-cells and respond to MHC class II antigen stimulation and are required to produce antibodies from B-cells (Garnelo et al., 2017). T-cell epitopes are peptide sequences that are presented by the MHC class II receptors on APCs, which result in stimulation of CD4+ helper T-cells and cytotoxic lymphocytes, resulting in immune system activation (Figure 3).

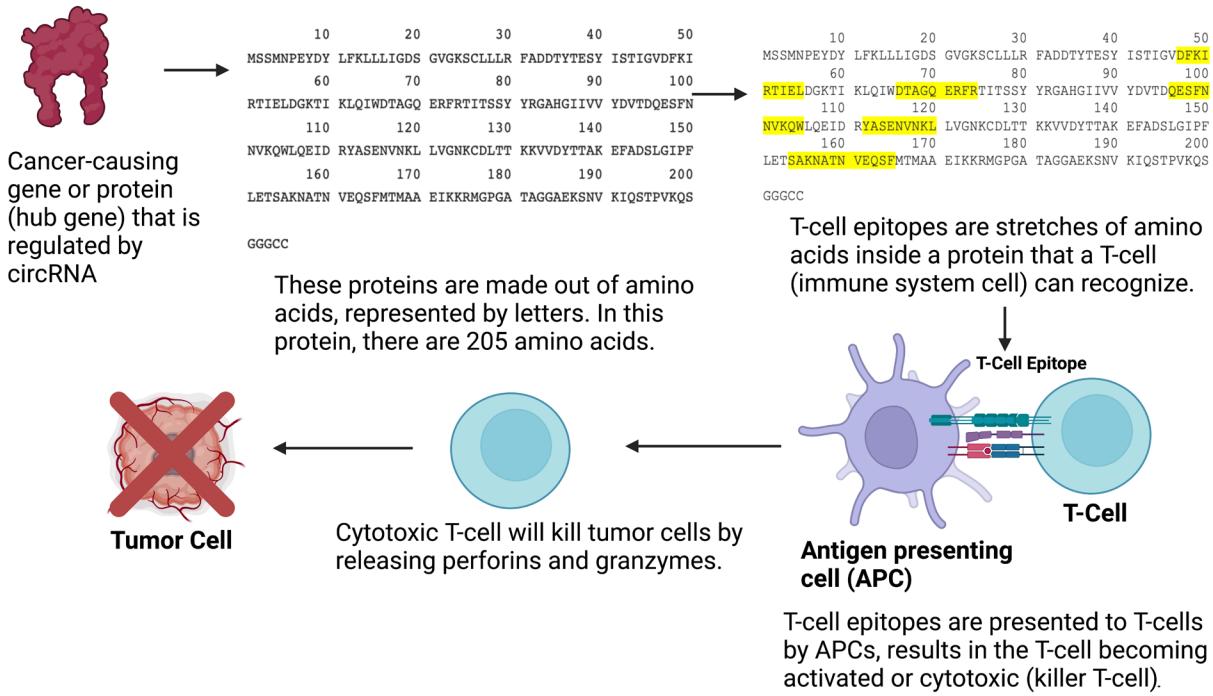

Figure 3. Potential tumor therapy strategy to identify oncogenes/oncoproteins in the hub-genes coded by mRNAs that are in turn regulated by circRNAs. Created with BioRender.com.

1.6 Immunogenicity

Immunogenicity is the ability of a protein to elicit an effective immune response. In any oncogene(s), a high T-cell epitope count can mean a higher immunogenicity and more effective immune response, and thus cytotoxic T-cell killing of tumor cells as shown by Figure 3 (De Groot et al., 2020). Peptides from proteins with high immunogenicity are great candidates

for peptide vaccines, which offer a significant alternative to whole cell antigens because of protective immune response, specificity and recognition to specific antigen, and fewer side effects (Muhammad et al., 2020).

Based on the above background, the present study used publicly available circRNAs expression data and identified optimum machine learning model that could predict, i.e. highly correlate with, the occurrence of hepatocellular carcinoma in patients and to predict the immunogenicity of deregulated circRNAs gene targets (called "hub genes") associated with tumor growth.

2 MATERIALS AND METHODS

2.1 Materials

The data for the present study was obtained from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155949 for data on ~60,000 circRNA expression levels in tumor (n=49) and adjacent healthy tissue (n=49) from n=49 liver cancer (HCC) patients (total n=98 HCC patient samples) (publicly available from GEO – gene expression omnibus). Additional materials/resources used for this study are as follows: Jupyter notebook running Python 3, Cytoscape (Shannon et al., 2003) for circRNA-miRNA-gene network, Protein-protein interaction network, and hub gene analysis, circInteractome to predict circRNA-miRNA pairs from https://circinteractome.nia.nih.gov/mirna_target_sites.html, miRWalk (Sticht, De La Torre, Parveen, & Gretz, 2018) to predict miR-gene pairs from http://mirwalk.umm.uni-heidelberg.de/, fully licensed version of BioRender to prepare model of the concept, Microsoft Excel/GraphPad Prism for data analysis, Human protein atlas for validation (https://www.proteinatlas.org/), GitHub repository for T-cell epitope data https://github.com/pirl-unc/cd8-tcell-epitope-prediction-data, Python modules used include SciPy, Pandas, NumPy, Matplotlib, Sklearn and Seaborn. Graphics included in Figure 1-4 are generated using fully licensed version of BioRender®–an online graphical tool.

2.2 Methods

First, all necessary modules were installed on Python/Anaconda using “pip install”. Next, data publicly available from GEO was downloaded. Using Pearson correlation, the top 12 circRNAs most correlated with tumor occurrence ($|r| > 0.68$ & $p < 0.01$) were identified. In this study, Pearson correlation can be used for categorical variable data. To validate the top 12 circRNAs, the mutual information (information gain) algorithm was used. Next, data on the top 12 circRNAs found previously was loaded into a new Jupyter notebook and the data was shuffled to prevent bias. Next, the data was split into training and testing with 70% – 30% split (70 for training ML models, 30 for testing). Machine learning analysis was performed for the data with seven models (K-nearest neighbor (KNN), Random Forest, Decision Tree, Support Vector Machine (SVM), Naive Bayes, Gradient Boosting, and Logistic Regression (described in methods). The predictive ability of these models was measured with classification accuracy (percent of correctly predicted tumor occurrences in patients), area under ROC curve (True positive vs False positive rate), and confusion matrix. To eliminate biases from the train-test split, Stratified K-fold Cross Validation was performed in Python, and model accuracies were compared with train-test split (normal machine learning). The confusion matrix was

performed on 12 circRNA features. Top two circRNAs were only selected for further analysis involving miRNA and mRNA expression. The confusion matrix was generated with true positive (model predicted positive and the actual label is positive), false positive (model predicted positive and the actual label was negative), true negative (model predicted negative and the actual label is negative) and false negative (model predicted negative and the actual label was positive). The second part of the project is to find circRNA- regulated hub genes (mRNAs) and their immunogenicity. To do this, another circRNA filter was created but made more statistically stringent to only get the top two circRNAs (strongest correlated with tumor occurrence) with criterion ($|r| > 0.7$ & $p < 0.01$). Using the circInteractome (Dudekula et al., 2016) and mirWalk algorithms and Microsoft Excel, datasheet was compiled of circRNA–miRNA-gene (mRNA) interactions for the top two circRNAs. Next, the interactions were graphed in Cytoscape as well as interaction strength/probability. Using the STRING function in Cytoscape, a gene interaction network (PPI) was created to model the circRNA-targeted gene interactions. Using the MCODE algorithm (finds clusters in the gene networks), the top hub genes (strongest correlation) for each circRNA were determined. Next, the Human Protein Atlas website was referenced for validation of the genes (check to see if the predicted hub genes are known oncogenes). To predict the immunogenicity of the hub genes, a T-cell epitope dataset was downloaded from GitHub. Using epitopes found in this dataset, a Python program was written to determine the number of T-cells epitopes as well as clusters of epitopes that are found in each hub gene protein sequence. More T-cell epitopes equates to higher immunogenicity since more T-cell epitopes can elicit more effective immune response by CD8+ and CD4 T-cells. To compare predicted epitopes to a negative control, a random "fake" protein sequence was generated in Python that has the same length as the hub gene, and the number of epitopes found was recorded. Finally, the number of T-cells epitopes found for each hub gene was normalized to the peptide length of each hub gene, and all immunogenicity scores and results were recorded. A brief overview of the methodology is shown in Figure 4.

Additional statistical methods, Python code and machine learning algorithms are described below.

Code snippets: All code can be found at this GitHub repository: https://github.com/adityakoushikk/ml-bio. About 1730 lines of code in total was needed.

Data Analysis and Filtering: The code (in Figure 5) filters the top 12 circRNAs using Pearson correlation. CircRNAs that have correlation coefficient $|r|>0.68$ and p-value (significance) < 0.05 are shortlisted. Pearson correlation is given by the equation (1):

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2(y-\bar{y})^2}} \quad (1)$$

Where r is the correlation coefficient, x is the values of the independent variables (each of the 12 circRNAs), $\bar{x}$ is the mean of the independent variable sample, y is the values of the target variable (tumor occurrence represented in binary 0-No tumor, 1-Tumor), $\bar{y}$ is the mean of the values in the target variable sample. Pearson correlation on categorical data works the same as point biserial correlation.

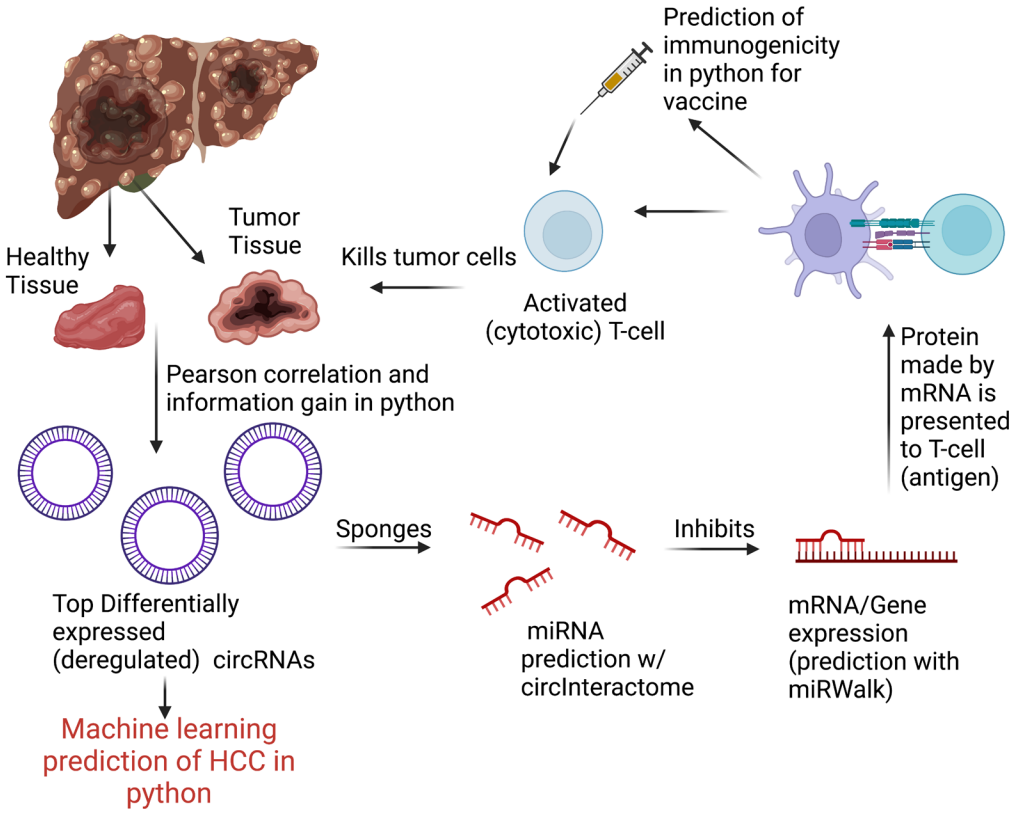

Figure 4. Experimental outline showing the analysis pipeline. Created with BioRender.com.

Machine Learning in Python & Examples: KNN

Hamming Distance

$$D_H = \sum_{i=1}^{k} |x_i - y_i|$$
$$x = y \Rightarrow D = 0$$
$$x \neq y \Rightarrow D = 1 \qquad (2)$$

An example of building a machine learning model in Python is shown in Figure 5. This classification model is K-Nearest Neighbor. First, data is split into 70% for training the model and 30% for testing using the train-test split from SKlearn. The KNN model from SKlearn is imported, and the model is fit (learns patterns in data using the KNN algorithm) for the training data. The trained model is tested on the other 30% of the data, and the prediction accuracy is displayed. A confusion matrix is displayed (explained in detail in results section). The K-nearest neighbor classifies data by looking at similarity based on Hamming distance (equation 2), or the neighboring points (if other points near the point at question fall under a certain class, then the point at question will most likely fall into that class). The Hamming Distance function is shown on the left (used for categorical prediction -https://www.saed-sayad.com/k_nearest_neighbors.htm), where K determines the number of neighbors. The code for building the other machine learning models is repeated in a similar way.

```
#create a dict to keep track of the accuracy of each model
accuracies = {}
#X will contain circRNA values, y will contain target variable - tumor
X = circ.values
y = circ['Tumor'].values
X = np.delete(X,0,axis=1)
#split the dataset into 70-30 for training and testing
x_train, x_test, y_train, y_test = train_test_split(X,y,test_size = 0.3,random_state=1)
scoreList = []
# KNN Model
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors = 2)
knn.fit(x_train, y_train)

print("{} NN Score: {:.2f}%".format(2, knn.score(x_test, y_test)*100))
knn_predictions = knn.predict(x_test)
scoreList.append(knn.score(x_test, y_test))
acc = max(scoreList)*100
accuracies['KNN'] = acc
confusion_matrix(y_test, knn_predictions)
```

```
2 NN Score: 83.33%

array([[15,  1],
       [ 4, 10]])
```

```
knn_predictions
```

```
array([1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0,
       0, 1, 0, 0, 0, 1, 1, 0])
```

Figure 5. K-Nearest Neighbor (KNN) classification model in Python.

Random Forest & Decision Tree: Decision trees learn by splitting dataset into smaller subsets to predict a target value (each condition is called a node, and possible outcomes are called "branches"), hence forming a tree. Random forest consists of many individual decision trees that operate as an ensemble (multiple learning algorithms). Decision Trees usually classify using Gini impurity, which gives a probability of misclassifying an observation by randomly picking a datapoint and randomly classifying it according to the distribution of the dataset (equation 3). The equation is shown on below, where C is the total possible classes, p(i) is the probability of picking a datapoint with class i (in this study, C is 2 for Tumor/No-tumor prediction, and p(i) would be 0.5).

$$\mathrm{G} = \sum_{i=1}^{C} p(i) * (1 - \mathrm{p(i)}) \tag{3}$$

Logistic Regression: Logistic Regression solves binary classification problems although it is a regression function. The basis of logistic regression is the sigmoid function shown in equation 4:

$$y = \frac{1}{1+e^{-x}} \tag{4}$$

The function can take any real value number and map it to a value between 0 and 1 (continuous circRNA expression values can be mapped in this way). The sigmoid curve can be graphed as in Figure 6 (https://towardsdatascience.com/logistic-regression-explained-9ee73cede081). With this method, the location of the point on the curve can determine the classification of that point. In this study, the x-axis would be the circRNA expression, and the y-axis would be the probability of tumor vs no-tumor.

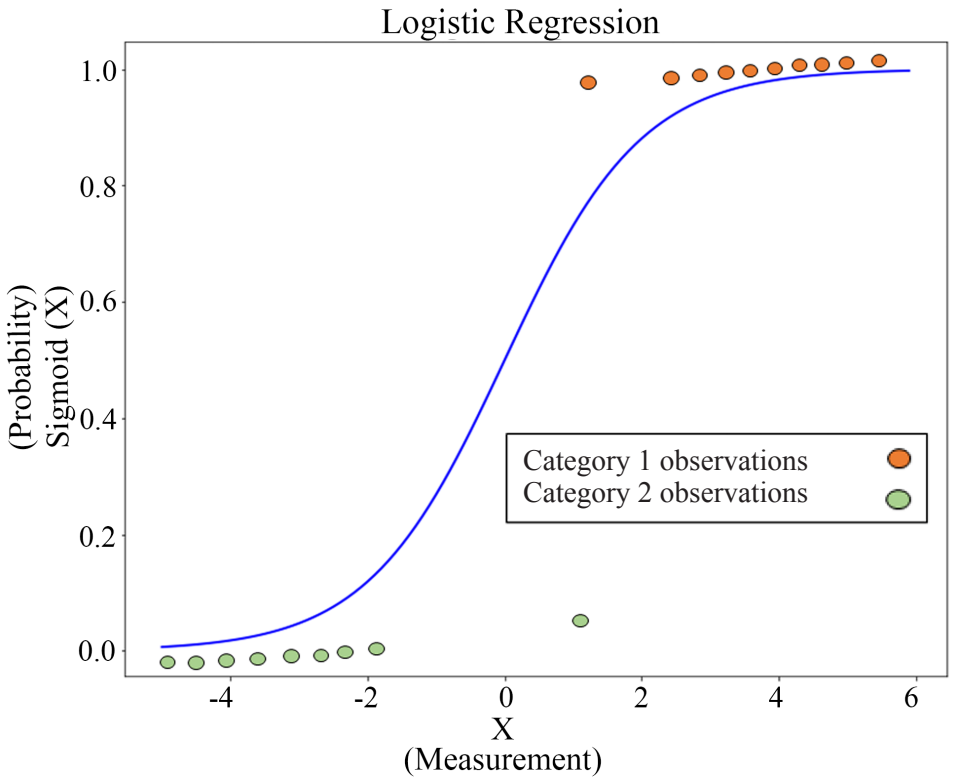

Figure 6. Sigmoid curve based on Logistic Regression showing CircRNA expression values (0 and 1; 0 being no tumor and 1 being tumor).

CircRNA-miRNA-Gene network

First, the top 2 circRNAs from Pearson correlation filtering ($|r| > 0.7$) were found (further explained below). These circRNAs were entered into the circInteractome website to predict the miRNA sponges for each of the two circRNAs. Figure 7 (based on circInteractome analyses - https://circinteractome.nia.nih.gov/mirna_target_sites.html) shows the result for one of the circRNAs (hsa_circ_0005284). Only the top 3 miRNAs were selected based on the context+ score (smallest numbers are ideal). miRNAs selected for hsa_circ_0005284 were hsa-miR-558, hsa-miR-639, hsa-miR-626. For hsa_circRNA_089372, miRNAs selected were hsa-miR-1247-5p, hsa-miR-1289, and hsa-miR-1184. These miRNAs were fed into the mirWalk algorithm to predict the gene (mRNA) targets that these miRNAs inhibit. Figure 8 shows an example for gene target search with mirWalk for hsa-miR-1184. These analyses were done based on mirWalk (http://mirwalk.umm.uni-heidelberg.de/). Genes were filtered for a score higher than 0.85, only CDS (coding sequence genes), and finally validation from miRTarBase.

miRNA Predictions

miRNA	#Sites
hsa-miR-1265	1
hsa-miR-495	1
hsa-miR-526b	1
hsa-miR-558	1
hsa-miR-582-3p	1
hsa-miR-587	1
hsa-miR-622	1
hsa-miR-626	1
hsa-miR-639	1

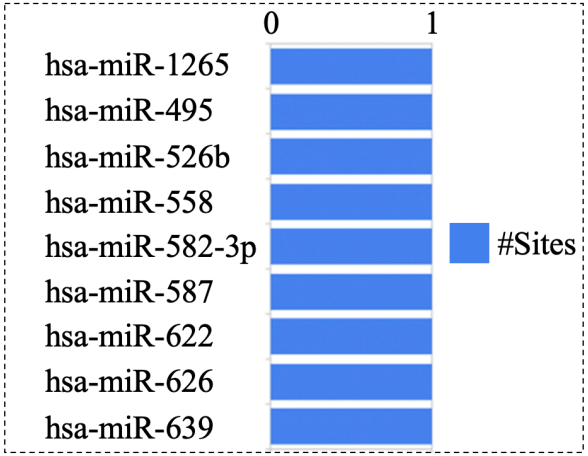

Figure 7. CircRNA-miRNA-mRNA (gene) network for top two significantly altered CircRNA based on circInteractome analyses.

TargetScan miRNA predictions

CircRNA Mirbase ID	CircRNA (Top) - miRNA (Bottom) pairing	Site Type	circRNA Start	circRNA End	3' pairing	Local AU	Position	TA	SPS	Context+ Score	Context+ Score+%
hsa_circ_0005284 (5' ... 3') hsa-miR-1265 (3' ... 5')	GGAUACAGCCUGUGCACAUCCUG ||||||| UUGUUGUGAACUGGUGUAGGAC	7mer-m8	213	219	0.003	0.064	-0.061	-0.011	-0.002	-0.127	86
hsa_circ_0005284 (5' ... 3') hsa-miR-495 (3' ... 5')	UCUGAGAAGGGACCAGUUUGUUG ||||||| UUCUUCACGUGGUACAAACAAA	7mer-m8	44	50	-0.007	0.053	-0.056	0.038	0.071	-0.021	93
hsa_circ_0005284 (5' ... 3') hsa-miR-526b (3' ... 5')	AUGCGACUGAGACAGCUCAAGAG ||||||| UGUCUUUCACGAAGGGAGUUCUC	7mer-m8	92	98	0.003	0.047	-0.049	0.003	0.016	-0.1	86
hsa_circ_0005284 (5' ... 3') hsa-miR-558 (3' ... 5')	UACAUGCGACUGAGACAGCUCAA |||||| UAAAACCAUGUCGUCGAGU	7mer-1a	89	95	0.004	0.020	-0.040	0.010	-0.052	-0.132	80
hsa_circ_0005284 (5' ... 3') hsa-miR-582-3p (3' ... 5')	GAAAUCUGAGAAGGGACCAGUUU ||||||| CCAAGUCAACAAGUUGGUCAAU	7mer-m8	40	46	0.012	0.057	-0.056	-0.009	0.004	-0.112	82
hsa_circ_0005284 (5' ... 3') hsa-miR-587 (3' ... 5')	AAUCUUAAACCAAGAAUGGAAAC |||||| CACUGAGUAGUGGAUACCUUU	7mer-1a	184	190	0.004	0.024	-0.046	0.023	0.021	-0.048	81
hsa_circ_0005284 (5' ... 3') hsa-miR-622 (3' ... 5')	NNNAAUAAUCAUGGGCCAGACUGG ||||||| CGAGGUUGGAGUCGUCUGACA	7mer-m8	14	20	too_close	too_close	too_close	too_close	too_close	too_close	NA
hsa_circ_0005284 (5' ... 3') hsa-miR-626 (3' ... 5')	GAGUACAUGCGACUGAGACAGCU |||| ||||||| UUCUGUAAAAG---UCUGUCGA	7mer-m8	86	92	-0.025	0.023	-0.050	-0.005	-0.037	-0.214	91
hsa_circ_0005284 (5' ... 3') hsa-miR-639 (3' ... 5')	AACCAAGAAUGGAAACAGCGAAG |||||| UGUCGCGAGCGUUGGCGUCGCUA	7mer-1a	191	197	0.008	0.010	-0.047	-0.080	-0.048	-0.231	73

miRNAID | GeneSymbol | 0.85 | CDS | | | miRTarBase | set filter

miRNA	RefseqID	GeneSymbol	Duplex	Score	Position	Binding Site	Au	Me	N Pairings	Targetscan	Mirdb	Mirtarbase
hsa-miR-1184	NM_001256695	PRDM11	details	0.92	CDS	3292,3330	0.37	-5.18	22	—	—	MIRT790030
hsa-miR-1184	NM_001271282	DICER1	details	1.00	CDS	5623,5675	0.52	-5.495	20	—	Link	MIRT558376
hsa-miR-1184	NM_001271608	LASP1	details	0.92	CDS	596,619	0.34	-6.828	18	—	—	MIRT735893
hsa-miR-1184	NM_001282878	LAX1	details	1.00	CDS	824,846	0.48	-6.372	18	—	—	MIRT643447
hsa-miR-1184	NM_001284390	VGLL4	details	1.00	CDS	878,898	0.27	-8.492	18	—	Link	MIRT464214
hsa-miR-1184	NM_001284391	VGLL4	details	1.00	CDS	780,800	0.27	-8.492	18	—	Link	MIRT464214
hsa-miR-1184	NM_001286265	MRS2	details	0.92	CDS	978,1001	0.53	-6.352	18	—	—	MIRT640711
hsa-miR-1184	NM_001287489	OTOF	details	1.00	CDS	3440,3474	0.34	-5.269	19	—	—	MIRT718119
hsa-miR-1184	NM_001287489	OTOF	details	1.00	CDS	5364,5392	0.46	-5.763	20	—	—	MIRT718119
hsa-miR-1184	NM_001287489	OTOF	details	1.00	CDS	2559,2586	0.35	-6.484	21	—	—	MIRT718119
hsa-miR-1184	NM_001290145	POLDIP2	details	1.00	CDS	954,999	0.37	-10.845	17	—	Link	MIRT566361
hsa-miR-1184	NM_001297576	PEA15	details	0.92	CDS	512,534	0.48	-4.171	18	—	Link	MIRT735898
hsa-miR-1184	NM_031459	SESN2	details	1.00	CDS	1077,1139	0.31	-8.382	18	—	—	MIRT505717
hsa-miR-1184	NM_032206	NLRC5	details	0.92	CDS	4773,4799	0.47	-8.948	19	—	—	MIRT725310
hsa-miR-1184	NM_032206	NLRC5	details	1.00	CDS	1929,1953	0.43	-5.771	19	—	—	MIRT725310
hsa-miR-1184	NM_032313	NOA1	details	1.00	CDS	1872,1908	0.56	-6.251	21	—	—	MIRT719531
hsa-miR-1184	NM_032816	CEP89	details	0.92	CDS	950,979	0.53	-4.957	19	—	—	MIRT626489
hsa-miR-1184	NM_080550	SYNRG	details	1.00	CDS	1030,1062	0.4	-6.484	18	—	—	MIRT725121
hsa-miR-1184	NM_133635	POFUT2	details	0.92	CDS	182,205	0.43	-10.139	18	—	—	MIRT761200
hsa-miR-1184	NM_180976	PPP2R5D	details	0.92	CDS	1753,1772	0.38	-10.526	15	—	—	MIRT735899

Figure 8. mirWalk analyses showing mRNA (hub gene) targets for the top miRNA from Figure 7. Based on this, it is expected that these mRNAs are inhibited by microRNAs that are in turn sponged by CircRNA (from Figure 6 above). The identified miRNA-mRNA pairs were validated by miRTarBase.

A.

	allele	peptide	measurement_value	measurement_inequality	measurement_type	measurement_kind	measurement_source	original_allele
0	BoLA-1*21:01	AENDTLVVSV	7817.0	=	quantitative	affinity	Barlow - purified MHC/competitive/fluorescence	BoLA-1*02101
1	BoLA-1*21:01	NQFNGGCLLV	1086.0	=	quantitative	affinity	Barlow - purified MHC/direct/fluorescence	BoLA-1*02101
2	BoLA-2*08:01	AAHCIHAEW	21.0	=	quantitative	affinity	Barlow - purified MHC/direct/fluorescence	BoLA-2*00801
3	BoLA-2*08:01	AAKHMSNTY	1299.0	=	quantitative	affinity	Barlow - purified MHC/direct/fluorescence	BoLA-2*00801
4	BoLA-2*08:01	DSYAYMRNGW	2.0	=	quantitative	affinity	Barlow - purified MHC/direct/fluorescence	BoLA-2*00801

B.

```
total = 0
lists = []
cluster = []
for index, row in peps.iterrows():
    lists.append(row["peptide"])
for x in lists:
    if x in e:
        total+=1
        clus = e.find(x)
        cluster.append(clus)
print(total)
```

C.

```
         10         20         30         40         50
MNLFRFLGDL SHLLAIILLL LKIWKSRSCA GISGKSQVLF AVVFTARYLD
         60         70         80         90        100
LFTNYISLYN TCMKVVYIAC SFTTVWLIYS KFKATYDGNH DTFRVEFLVV
        110        120        130        140        150
PTAILAFLVN HDFTPLEILW TFSIYLESVA ILPQLFMVSK TGEAETITSH
        160        170        180        190        200
YLFALGVYRT LYLFNWIWRY HFEGFFDLIA IVAGLVQTVL YCDFFYLYIT
        210
KVLKGKKLSL PA
```

Figure 9. (A) Potential T-cell epitope in hub genes (mRNAs identified from Figure 6 and 7) were compared against known T-cell epitopes from global T-cell epitope dataset. The table shows some few examples of T-cell epitopes (peptides), their score and the assays used to confirm immunogenicity. (B) Python program used to find the immunogenicity of each protein sequence. (C) Epitope clustering is determined by the location of epitopes in the protein sequence. Note that the actual number (and not normalized) of T-cell epitopes in each candidates gene targets are identified.

Immunogenicity Prediction

Figure 9A shows the head of the T-cell epitope dataset, which contained about 200,000 T-cell epitopes. First, epitopes were filtered for only human alleles of MHC proteins (HLA). Next, they were filtered for the best nanomolar affinity between each epitope and peptide (most often an IC50 or inhibit concentration value). The smaller value indicates a higher MHC binding affinity and therefore a better epitope. Each IC50 value was normalized to a value between 0 and 1 using 1 – log (min(IC50, 50000))/log(50000). The epitopes that scored less than 0.35 on the normalized scale were discarded. Another dataset on eluted-MHC ligands was downloaded, which contained epitopes that are identified to bind MHCs via immuno-precipitation. Finally, an overlap between epitopes from each dataset was created, and the epitopes (Figure 9A) would be checked in the hub gene peptide sequences.

Figure 9B was the program used to find the immunogenicity of each protein sequence. "e" was the peptide sequence, and x was each T-cell epitope. The location of each epitope found was also predicted. Location is given by the position of the epitope amino acid in the full proteins sequence. For example, the full proteins sequence for KDELR1 hub gene is shown in Figure 9C. The epitope clustering in Figure 9C was determined by the location of epitopes in the protein sequence.

CircRNA correlation (Supplementary Finding): Some circRNAs are highly correlated with each other, rather than with tumor occurrence. The researcher hypothesized that circRNAs that are highly correlated with each other come from the same gene. For example, ASCRP3008985 (hsa_circ_0008539) and ASCRP3009102 (hsa_circ_0031027) have a strong correlation (r = 0.95) on the heatmap. Arraystar data on the two circs show they come from the same gene (TMCO3). Example graphs are shown below (Figure 10), and actual correlations are in Figure 11.

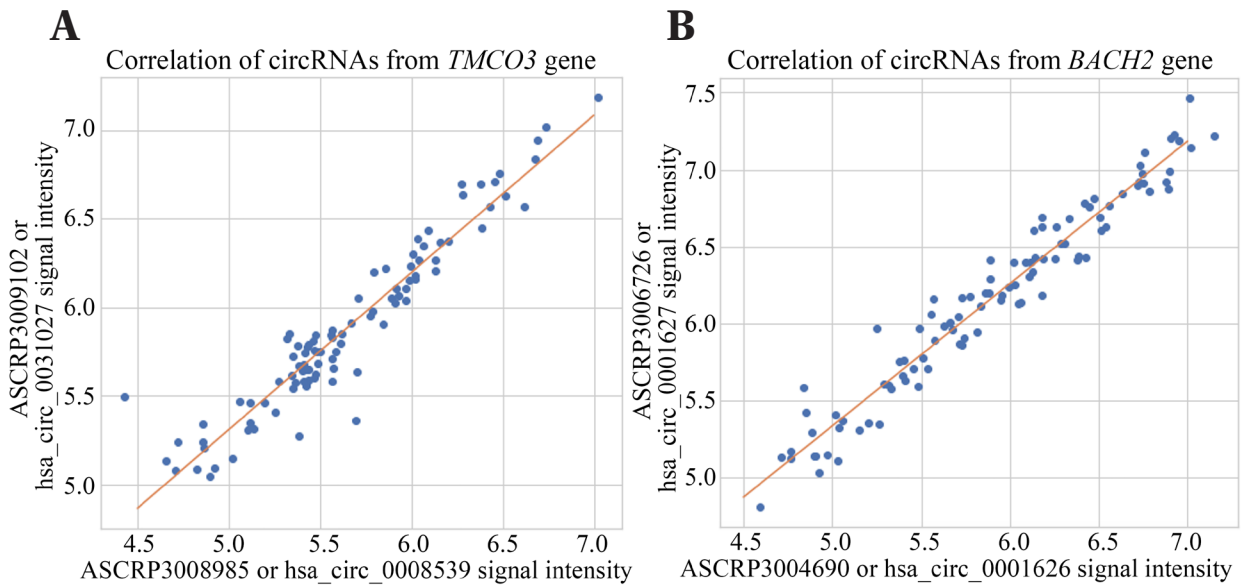

Figure 10. Pearson correlation for hsa_circ_0008539 and hsa_circ_0008539 for hub gene TMCO3 (A); and has_circ_0001627 and has_circ_0001626 for hub gene BACH2 (B). Pearson correlation coefficient for panel A – 0.95; Pearson correlation coefficient for panel B – 0.97.

3 RESULTS

The results are divided into following sections: CircRNA & Tumor Correlation, Machine Learning, Cytoscape Network Analysis, Validation & Survival Analysis, and Immunogenicity Prediction.

CircRNA & Tumor Correlation

Figure 11 shows the Pearson correlation heatmap generated in Python of top 12 deregulated circRNAs. Purple regions indicate strong negative correlation, and green areas indicate strong positive correlation. The most significant column is the first column, as it shows the correlation with tumor occurrence, which is this study's purpose. CircRNAs that are strongly negatively correlated ($r<0$) with tumor occurrence are downregulated (less expression in tumor tissue compared to healthy tissue), and circRNAs that are positively correlated ($r>0$) are upregulated (more expression in tumor tissue compared to healthy tissue). Upregulated circRNAs are associated with tumor growth, while downregulated circRNAs are not associated with tumor growth. The top 12 deregulated circRNAs can also be visualized similarly with violin plots below (Figure 12).

Violin Plots of Top 12 Deregulated CircRNAs

Figure 12 shows the violin plots for the top 12 deregulated circRNAs. The violin plot graphs the general distribution of the data as well as the median signal intensities for both Tumor and No-Tumor samples (0 indicates No-Tumor, 1 indicates Tumor). The shape of each plot follows a probability density function, where the widest point indicates a high probability that each sample will have the given circRNA expression value. A uniform shape shows a normal distribution, where data points are concentrated around the median. Dots shown in the middle of the graph represent each of the 98 tissue samples from the dataset, and more concentrated/clustered data points will widen the probability density function/distribution curve. Upregulated circRNAs show a blue violin plot higher than the red (signal intensity/expression for tumor samples is higher than for non-tumor samples), and the opposite for downregulated circRNAs. Distributions of both Tumor and Non-Tumor plots show generally normal distributions with some plots having an irregular distribution.

After restricting the Pearson correlation values to $[r] > 0.7$, the top 2 circRNAs were ASCRP3001251 (formal alias is hsa_circ_0005284) and ASCRP3001458 or hsa_circRNA_089372. The hsa_circ_0005284 is upregulated in tumor with a correlation value of 0.71 and hsa_circ_089372 is downregulated with correlation value of -0.82 (strongest out of all circRNAs). Figure 13A & 13B are graphs of a sigmoid function fit for the top 2 circRNAs.

From the graphs, high hsa_circ_0005284 expression values are mostly associated with tumor occurrence, while high hsa_circRNA_089372 expression values are indicative of no tumor. This shows that the circRNAs are opposites of each other, and the sigmoid curves are also pointing opposite ways for each circRNA. hsa_circRNA_089372 (Figure 13A) has a more uniform sigmoid curve and higher Pearson value than hsa_circ_0005284 (outliers at the top of the graph). The sigmoid curve in Figure 13A is close to 1, which shows a near 100% tumor probability for low circRNA expression. In Figure 13B, the probability curve is not as certain and even high signal intensity will only guarantee ~90% tumor probability. Overall,

from sigmoid curve analysis and Pearson correlation, hsa_circ_089372 is more correlated with lower tumor probability and is relatively a better biomarker than hsa_circ_0005284 albeit both show strong, yet opposite, correlations (hsa_circ_089372 – negative correlation and hsa_circ_0005284 – positive correlation) with tumor probability.

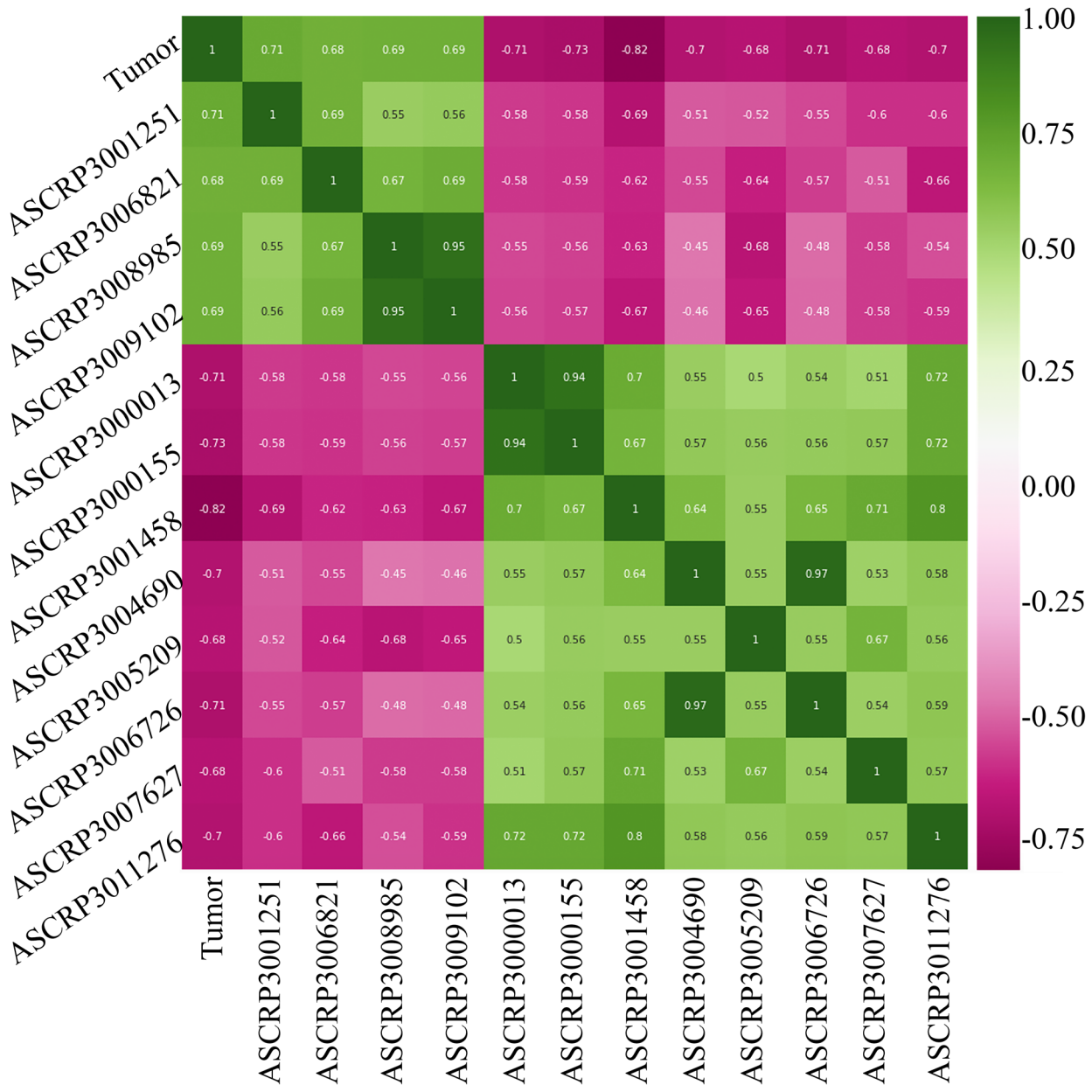

Figure 11. Pearson correlation heatmap generated with intensity of expression of top 12 deregulated circRNAs in healthy liver/non-tumor and tumor regions. X and Y-axis – circRNA type, and first column depicts Pearson correlation value for each of top 12 circRNA for tumor. Other columns correlations between circRNAs based on their expression intensity. Data generated in Python of top 12 deregulated circRNAs. Purple regions indicate strong negative correlation, and green areas indicate strong positive correlation. The most significant column is the first column, as it shows the correlation with tumor occurrence.

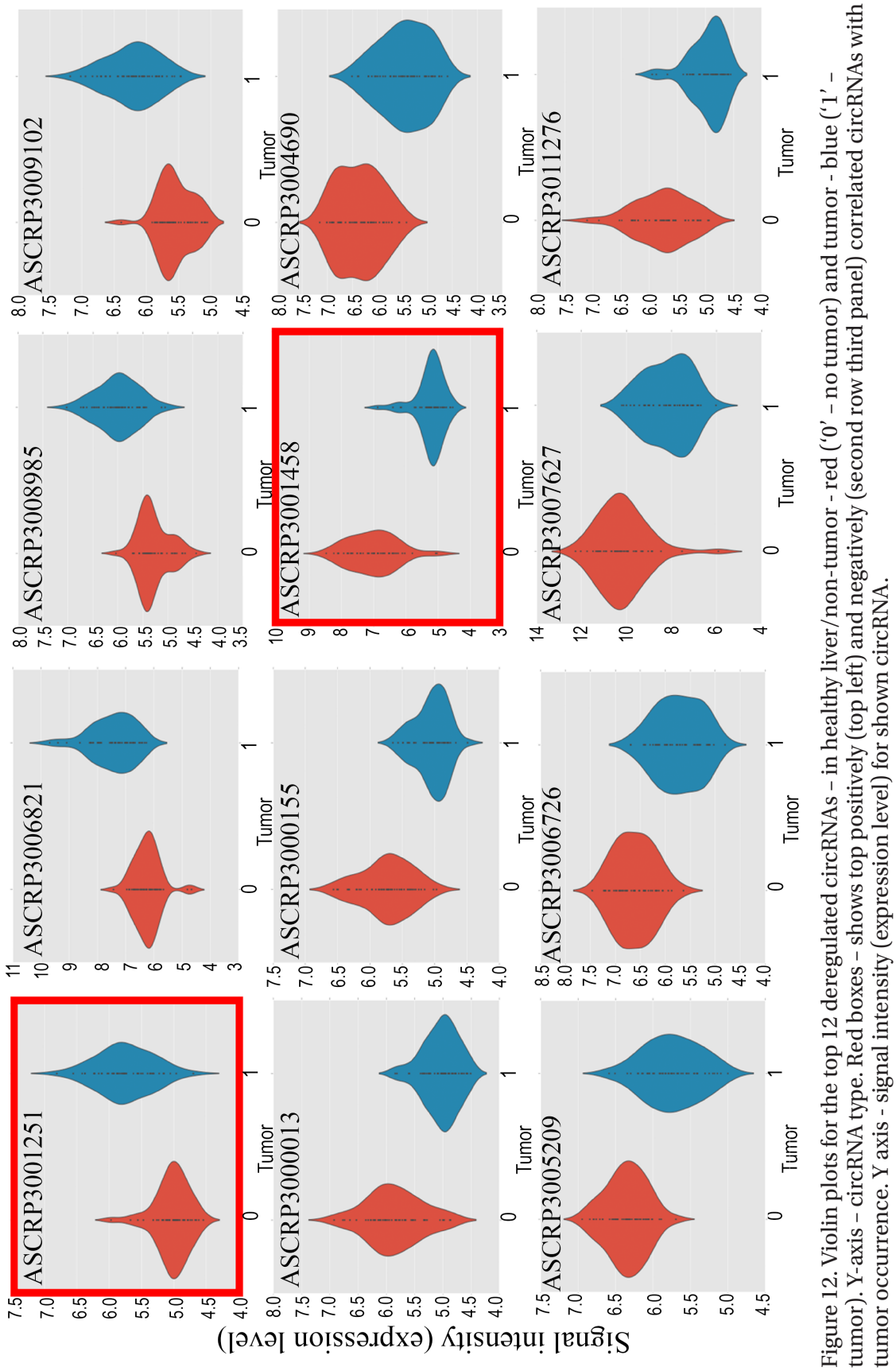

Figure 12. Violin plots for the top 12 deregulated circRNAs – in healthy liver/non-tumor - red ('0' – no tumor) and tumor - blue ('1' – tumor). Y-axis – circRNA type. Red boxes – shows top positively (top left) and negatively (second row third panel) correlated circRNAs with tumor occurrence. Y axis - signal intensity (expression level) for shown circRNA.

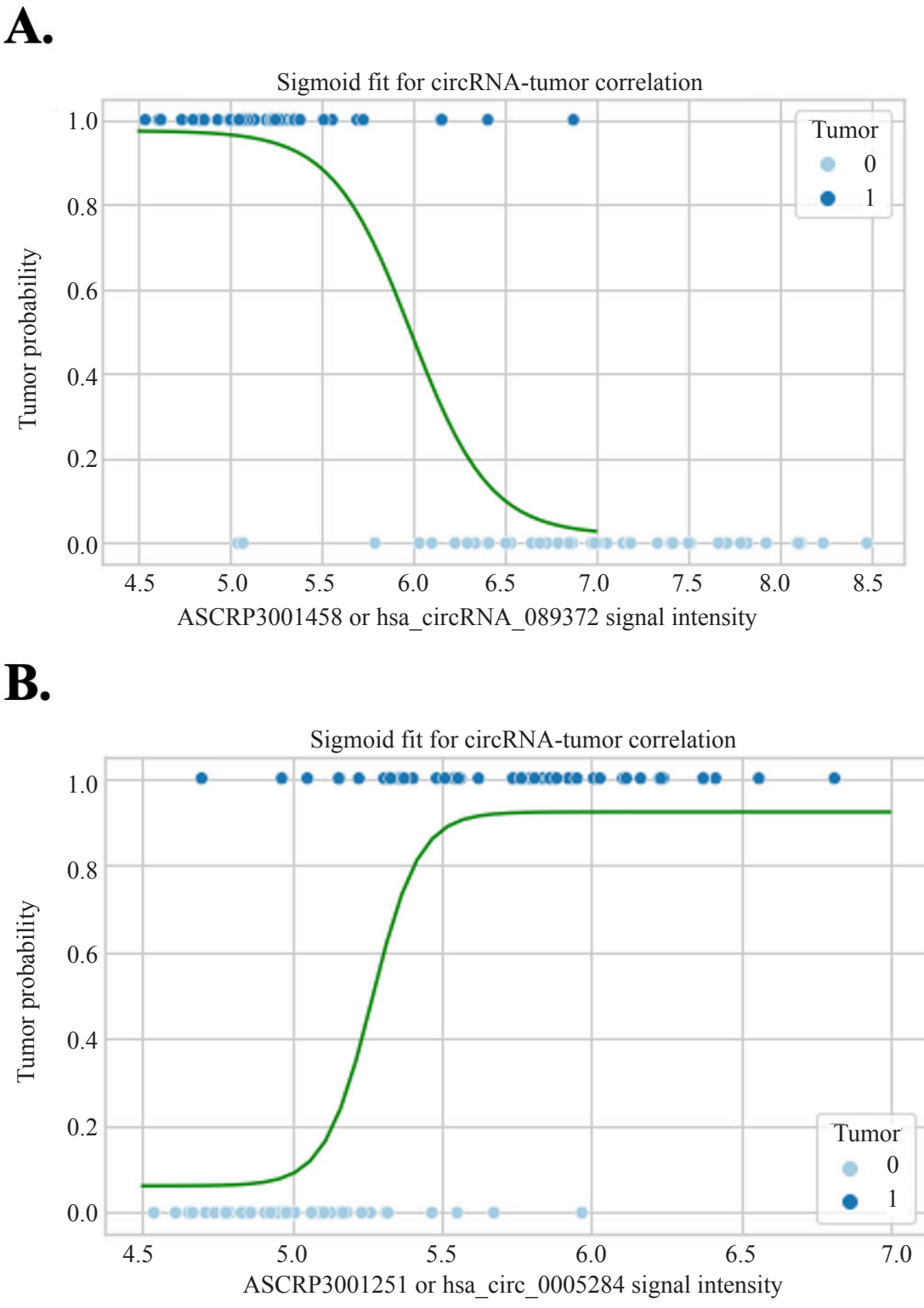

Figure 13. Sigmoid curve based on Pearson correlation for circRNA and tumor occurrence. (A) has_circRNA_089372 vs tumor probability. (B) has_circRNA_00005284 vs tumor probability.

Machine learning and model accuracies

Figure 14A shows the accuracies of seven machine learning classification models after trained by the dataset. The logistic regression scored the highest with 100% accuracy. Model accuracies for other machine learning algorithms were 'KNN': 83.33, 'Random Forest': 96.67, 'Decision Tree': 73.33, 'SVM': 96.67, 'Naive Bayes': 96.67, 'Gradient Boosting': 90.0 (Figure 14). There is a chance that the model might be overfitting (generalizing patterns in training data but performing poorly in test data), but through cross-validation this can be prevented. This was further confirmed by confusion matrix analyses (Figure 14B).

Figure 14B shows the confusion matrix for each of the seven machine learning models. Confusion matrices evaluate model performance by showing the number of predictions of each class. Table 1 shows what each number means in each quadrant of the confusion matrix:

	Predicted: Tumor	Predicted: No Tumor
Actual: Tumor	True positive	False Positive
Actual: No Tumor	False Negative	True negative

Table 1.

True Positive: # of correctly predicted tumor occurrences
False Positive: # of incorrectly predicted tumor occurrences
False Negative: # of incorrect predictions of no tumor (healthy)
True Negative: # of correct predictions of no tumor (healthy)

To better visualize the confusion matrix, a receiver operating characteristic (ROC) curve was plotted for all seven machine learning algorithms. An ROC curve illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. Overall, many of the machine learning algorithm tested showed >0.9 sensitivity compared to random prediction, whose area under the curve (AOC) was at 0.5 (Figure 15). Model performances were analyzed using ROC curves and AUROC (Area under ROC curve). The ROC curve plots the True positive rate (TPR) vs False positive rate (FPR) (equations 5 and 6).

$$TPR\ or\ Recall\ or\ Sensity = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{5}$$

TPR = (# of True positives)/(True positives + False Negatives)

(False positive rate = (# of False Positives)/(True Negatives + False Positives)

$$FPR\ = 1 - specificity = \frac{\mathrm{FP}}{\mathrm{TN+FP}} \tag{6}$$

The AUROC is the area under the ROC curve. A good AUROC is near 1, which shows a good measure of separability (Tumor vs No Tumor). A poor model will have an AUROC near 0, which indicates the model gets every prediction incorrect. A model with AUROC of 0.5 cannot make any separation (cannot differentiate between Tumor and Non-Tumor circRNA samples). Figure 15 shows the ROC curve for all models. Most models had an AUROC near 1 and are therefore not visible (cross into the border). Logistic Regression, Support Vector Machine, and Random Forest had the highest AUROC scores.

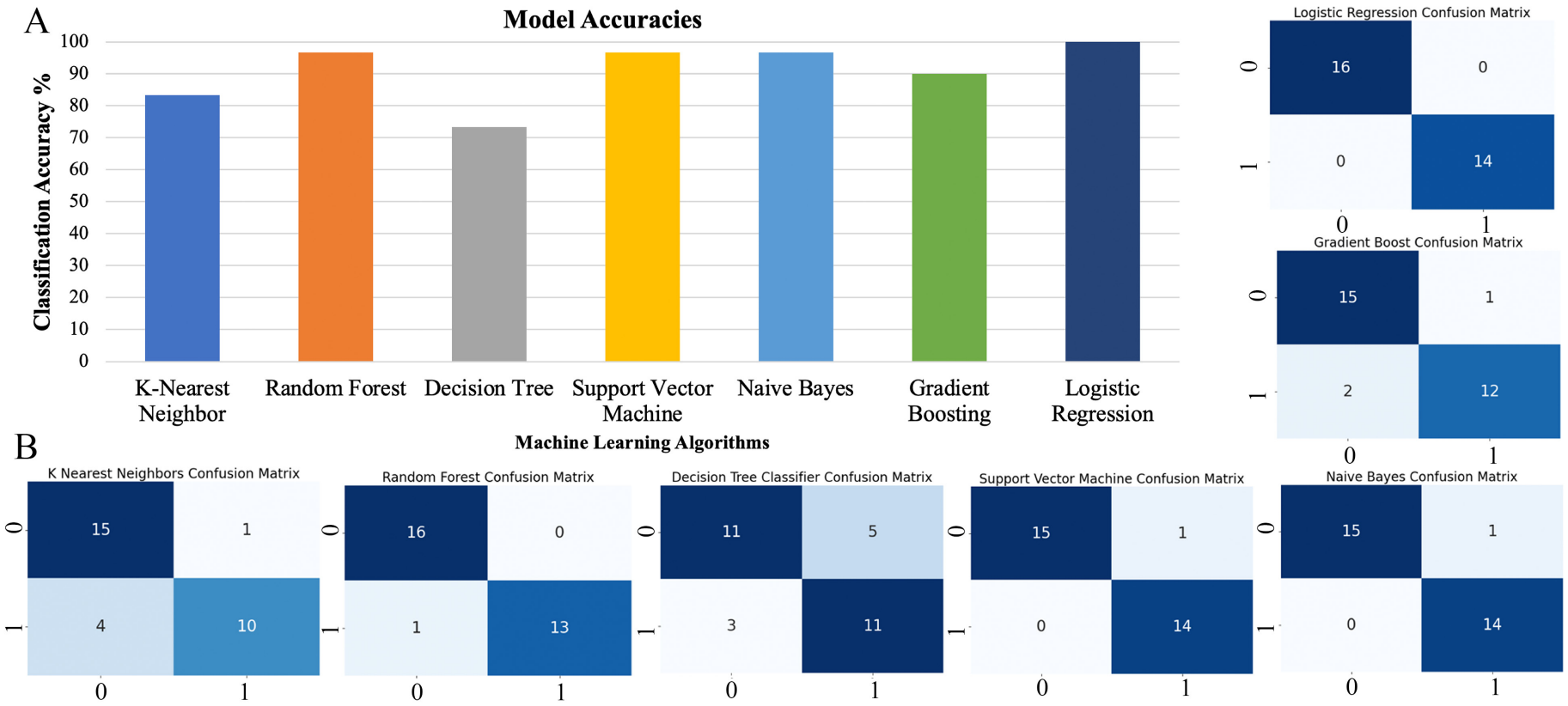

Figure 14. Graph showing ML model accuracies and confusion matrix. (A) Shows the accuracies of 7 machine learning classification models after training. Note that majority of the models (except decision tree) showed >80% accuracy with logistic regression showing 100% classification accuracy. (B) shows the confusion matrix for each of the 7 machine learning models. Left label (Y-axis) – predictive value; bottom (X-axis) – actual values. Positive value (1) and negative value (0). True positive are bottom right, false positive are bottom left, true negative are top left and false negative are top right.

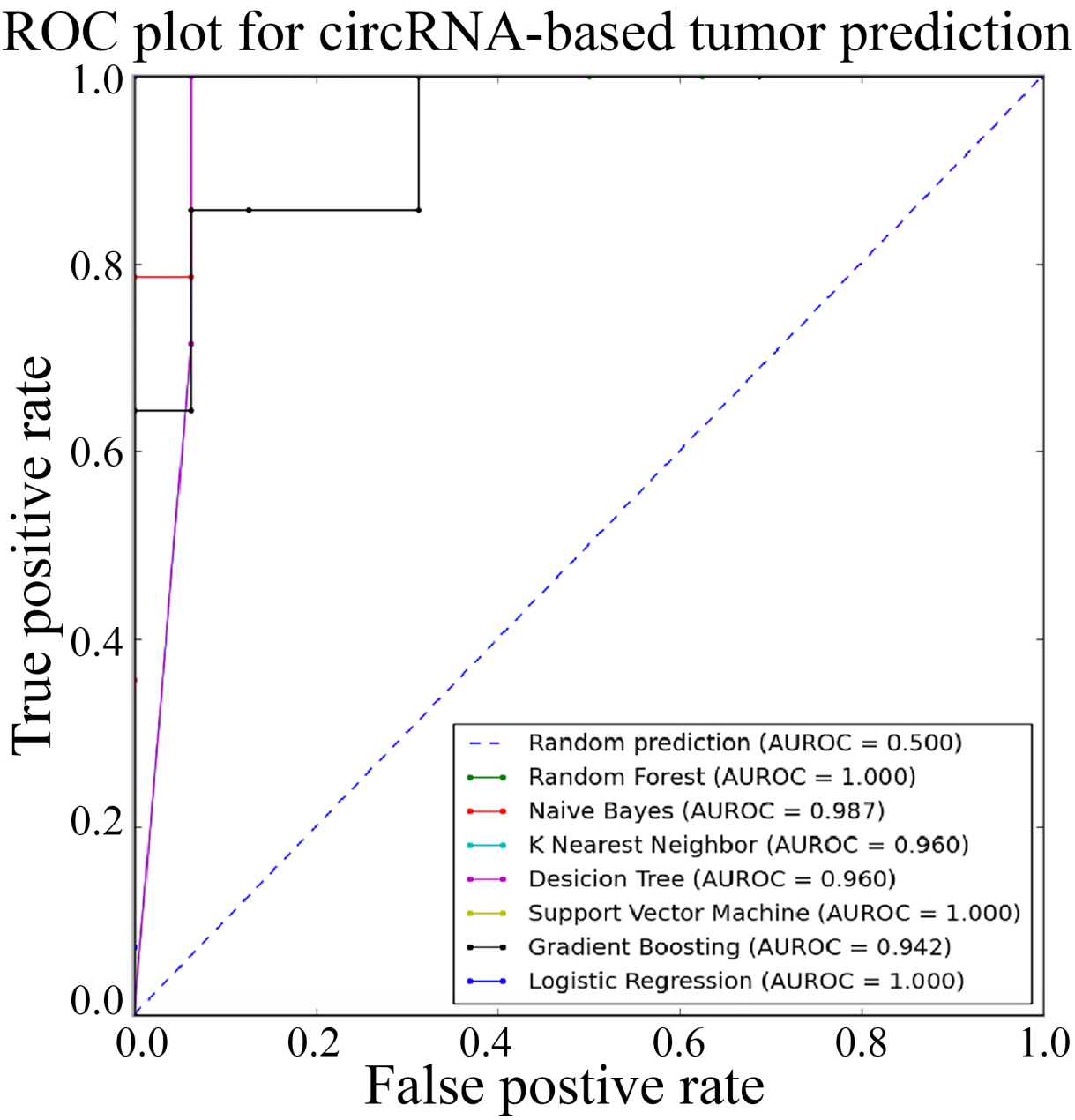

Figure 15. ROC curve for 7 different machine learning algorithms.

Stratified K-Fold Cross Validation: Next, Stratified K-fold Cross validation was performed in Python. In regular machine learning, the data was split using the train test split (70% of the data used to train models, and the other 30% used to test the model accuracy). Splitting the data with a train-test split has more bias and can lead to overfitting, so Stratified K-fold splits the data in a different manner (Figure 16A). K-fold cross validation involves splitting data into k equal folds (Figure 16). The first k-1 folds are used for training, and the remaining are used for testing. This is repeated for all k-folds, and the mean of the accuracies of each k-fold is returned. Stratified k-fold is similar, but involves splitting the data into folds not randomly, but based on the number of each class (if fold 1 has 15 tumor samples and 15 non-tumor samples, then fold 2 should have a roughly equal amount). This helps eliminate biases that

come with randomly splitting the data. Model accuracies for Stratified K-Fold: KNN: 87.784, Random Forest: 95.928, Decision Tree: 88.826, Support Vector Machine: 95.928, Naïve Bayes: 95.928, Gradient Boosting: 90.846, Logistic Regression: 95.928

Figure 16B shows the model accuracies of each model after performing Stratified-K-Fold Cross Validation. Logistic Regression, Naïve Bayes, Support Vector Machine, and Random Forest had the highest accuracy of nearly 96%.

Next, original accuracy prior to cross validation vs the accuracy after cross validation was plotted for all machine learning models (Figure 17). After Stratified K-Fold cross validation, the KNN, Decision Tree, and Gradient boosting algorithm accuracies increased while the Random Forest, Support Vector Machine, Naive Bayes, and Logistic Regression models

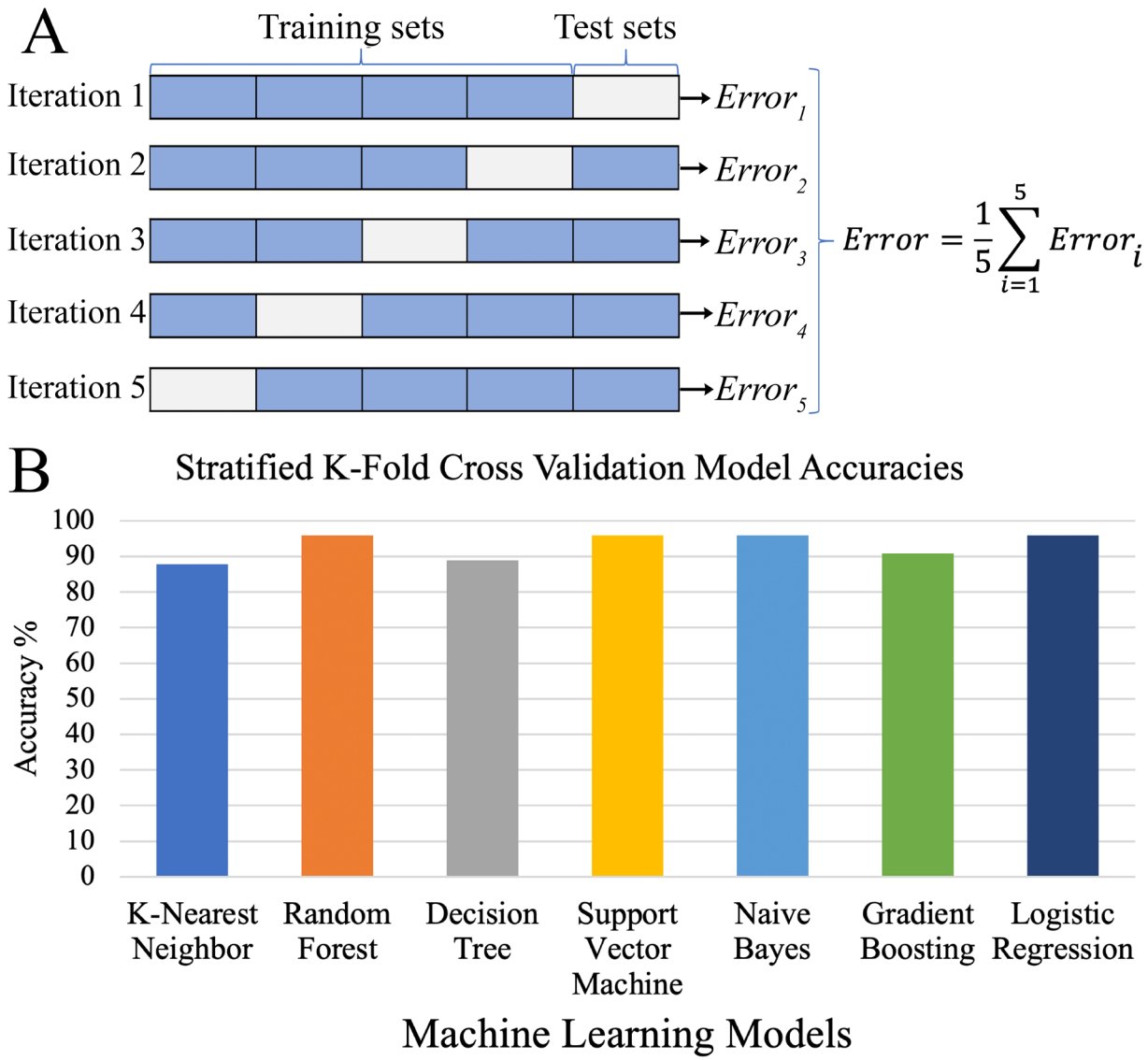

Figure 16. Cross-validation of ML algorithm (https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b). (A) Stratified K-fold Cross validation was performed in Python. (B) Shows the cross validation of 7 machine learning classification models after training.

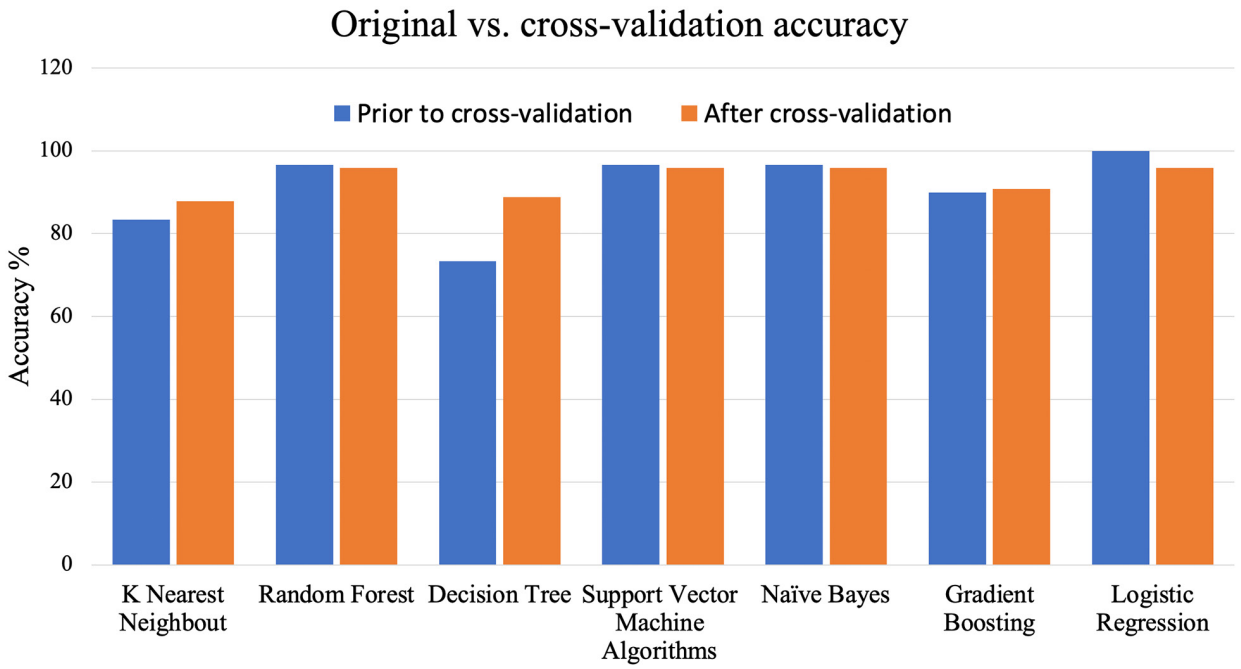

Figure 17. Comparison between original accuracy and validation accuracy of ML algorithm.

decreased. It is likely that the algorithms that decreased in accuracy were slightly overfitting prior to cross validation.

Finally, re-plotting ROC and AUROC scores for each model for Cross Validation showed that the Naive Bayes and Logistic Regression had the largest AUROC scores of 0.992 (Figure 18).

Cytoscape Network Analysis: Cytoscape is an open-source bioinformatics analyses software tool/program that helps in visualization of gene and protein interaction networks. Using Cytoscape, the entire network of two circRNAs that are downregulated (hsa_circRNA_089372) or upregulated (hsa_circ_0005284) in tumor was plotted along with their corresponding miRNAs and mRNAs or the "hub genes".

Figure 19 is a circRNA-miRNA-gene network created in Cytoscape. “Bindingp” is a metric predicted by miRWalk which shows the probability/score of an interaction between a miRNA and mRNA/gene. Top two circRNAs are shown in two orange octagons, miRNA is shown in purple circles, and genes/mRNAs are shown in squares. Fewer validated and high-scoring gene targets were found for hsa-mir-626 and hsa-mir-639 as shown. Because of the vast quantity of genes predicted by mirWalk, the researcher wanted to filter all predicted genes to the top 4 for each circRNA. These are hub genes and usually play critical roles in HCC progression and development.

To find the hub genes, first a PPI (Protein-Protein Interaction) or gene interaction network was constructed using the STRING algorithm in Cytoscape for each circRNA (Figure 20). Note that genes in hsa_circ_0005284 (upregulated in tumor as shown in Figure 20A) have no significant interaction with other genes and were not investigated in this study.

On the other hand, the PPI network of hsa_circRNA_089372 showed far more gene inter-

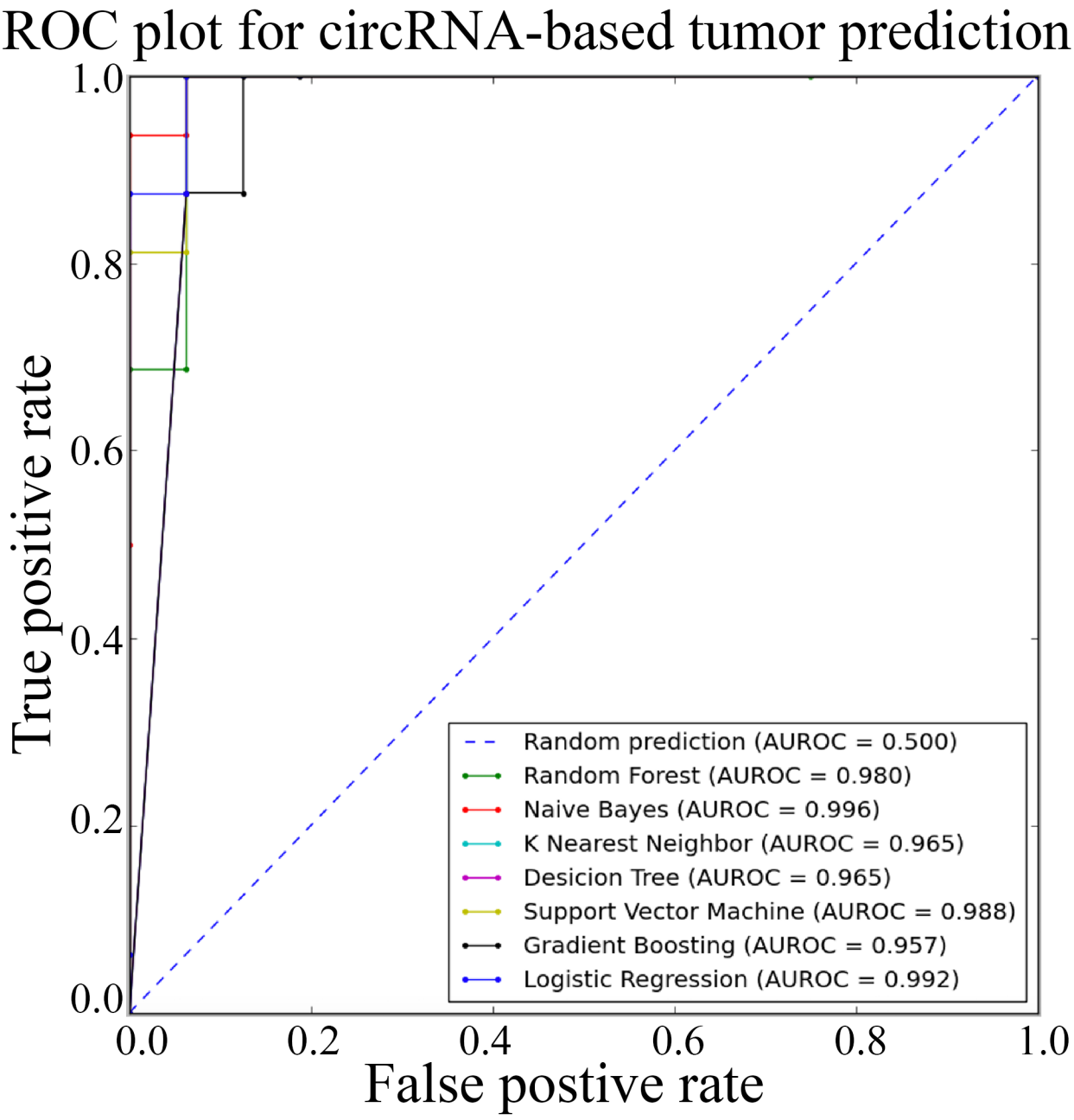

Figure 18. Shows the ROC plot and AUROC scores for each model for Cross Validation.

actions compared to hsa_circ_0005284 (Figure 20B), which could mean that it likely plays a more critical role in HCC progression and development (also proven by higher correlation value w/ tumor). Next, the MCODE algorithm (finds clusters of genes in the network) was performed for both PPI networks to determine the final hub gene. These clusters contain hub genes, which was used for immunogenicity analysis. After MCODE algorithm, the resulting hub genes involvement in cancer were validated using Human Protein Atlas (https://www.proteinatlas.org/).

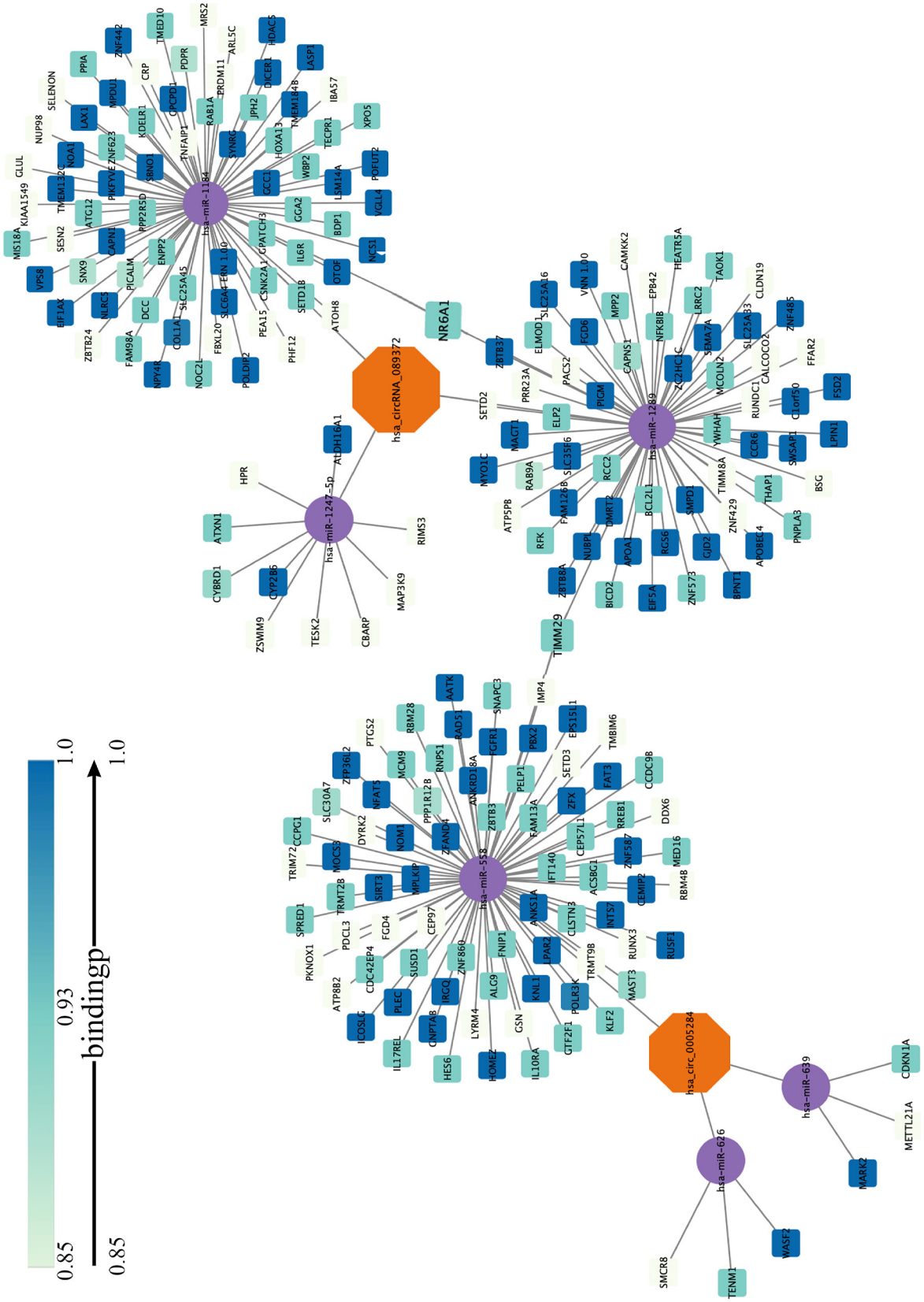

Figure 19. CircRNA-miRNA-gene network created in Cytoscape. CirRNA (orange octagon); miRNA (purple circles; mRNA (hub genes) (blue or cyan squares).

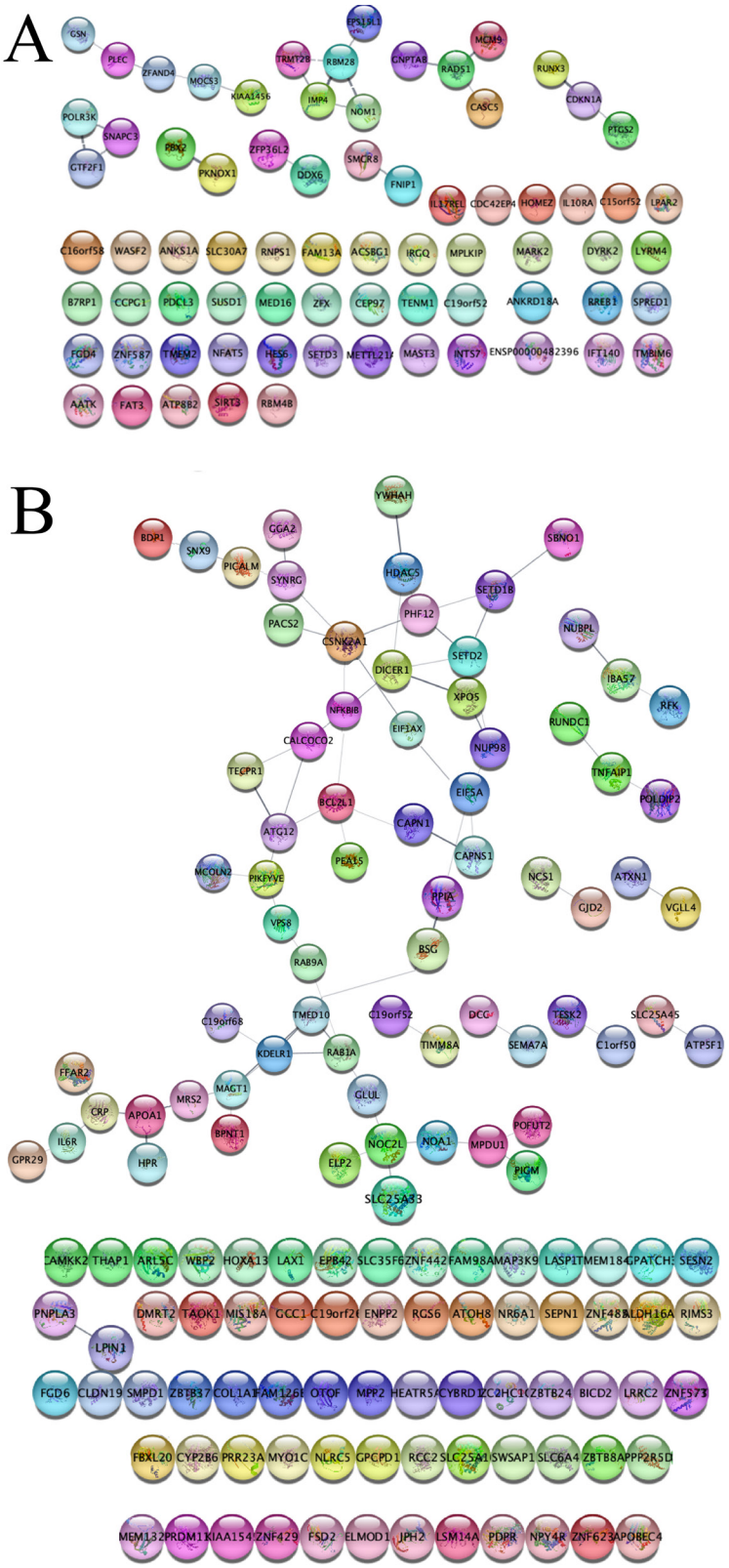

Figure 20. Protein-protein interaction (PPI) network of hub genes regulated by (A) hsa_circ_0005284 (upregulated in tumor) and (B) hsa_circRNA_089372 (downregulated in tumor) based on STRING algorithm in Cytoscape.

Validation and Survival Analysis: As seen in Table 2 (below), 7/8 of the hub genes are known to play an unfavorable or favorable prognostic markers in a variety of cancers based on the human protein atlas. Further, 3/8 of the hub genes are known to show an unfavorable prognosis in liver cancer (HCC), although additional papers outside of the Human Protein Atlas were found that validated hub genes other than these 3 for liver cancer (for example *RAB1A*). Next, human protein atlas survival analysis was performed for these 3 hub genes validated for liver cancer, and the graphs are shown below.

hsa-circ-0005284 hub genes(below)	**Prognosis and Cancer type (from Human Protein Atlas)**
TRMT2B	Prognostic marker in renal cancer (favorable) and breast cancer (unfavorable)
RBM28	Prognostic marker in liver cancer (unfavorable), endometrial cancer (unfavorable) and melanoma (unfavorable)
IMP4	Prognostic marker in liver cancer (unfavorable)
NOM1	Not prognostic
hsa-circ-089372 hub genes (below)	
MAGT1	Prognostic marker in breast cancer (unfavorable), pancreatic cancer (unfavorable) and melanoma (unfavorable)
KDELR1	Prognostic marker in liver cancer (unfavorable) and head and neck cancer (unfavorable)
TMED10	Prognostic marker in thyroid cancer (unfavorable)
RAB1A	Prognostic marker in head and neck cancer (unfavorable)

Table 2. Hub genes predicted by MCODE and Human Protein Atlas validation.

Survival Analysis: Table 3 shows the survival analysis graphs for the 3 circRNA targeted hub genes that were validated for liver cancer. Out of the three hub genes, *RBM28* expression levels can most significantly determine survival probability, then *KDELR1* and finally *IMP4* (*RMB28* had lowest p-value). In all three graphs, the survival probability for high gene expression is lower than low gene expression, which shows unfavorable prognosis for high gene expression. The expression cutoff is highest in *KDELR1* (highest *FPKM*), which could mean that *KDELR1* has higher median gene expression levels. It can also be noted that *RBM28* has the lowest 5-year survival rate for high expression (0% survival rate).

Hub gene	**Survival analyses score**
RBM28	Median follow up time (years) :1.63 P score: 2.7e-11 5-year survival for high expression: 0% 5-year survival low expression: 54% High vs low expression is determined by the cutoff 1.52 Fragments Per Kilobase of transcript per Million mapped reads (FPKM). High expression is above this cutoff, low expression is below this cutoff.

IMP4	Median follow up time (years) :1.63 P score: 0.00047 5-year survival for high expression: 35% 5-year survival low expression: 57% High/Low expression cutoff: 15.27 FPKM
KDELR1	Median follow up time (years) :1.63 P score: 0.00016 5-year survival for high expression: 40% 5-year survival low expression: 54% High/Low expression cutoff: 78.45 FPKM

Table 3. Survival analyses values for all three genes.

Immunogenicity Predictions: Immunogenicity predication can be done by a variety of ways. Table 4 is the graph for the number of HLA T-cell epitopes (detailed in appendices) found in the peptide sequences for each of the circRNA targeted hub genes. Hub genes targeted by hsa_circ_089372 have a higher number of T-cell epitopes in general. Because each hub gene has a different length (number of amino acids), the number of T-cell epitopes were normalized to a length of 1000 amino acids using (equation 7):

$$\frac{\#\ of\ T-cell\ epitopes}{(length\ of\ protein\ sequence) * 1000} \tag{7}$$

The number of T-cell epitopes predicted for each hub genes is shown in Table 4 (below). For a negative control, a random, fake protein sequence was created the same size of each hub gene to see how many epitopes were predicted. In all random protein sequences, 0 epitopes were found.

Gene	# Of T-cell epitopes predicted for each hub gene
TRMT2B	2
RBM28	18
IMP4	78
NOM1	20
MAGT1	110
KDELR1	52
TMED10	151
RAB1A	151

Table 4. # of T-cell epitopes predicted for each hub gene (protein)

Next, epitope clustering histograms for the top 4 genes were created. Regions in the protein sequences where there are a high number of T-cell epitopes have higher immunogenicity and are great targets for a cancer vaccine. These graphs show the raw number of epitopes (without normalization). Figure 21A-D shows the number of T-cell epitopes in each region of the protein sequences for the top 4 hub genes with the most T-cell epitopes. Peptide regions where there is a high epitope concentration have high immunogenicity.

The figure below shows the highest immunogenicity regions for each of the 4 hub genes, and the peptide sequence in this region of the protein are great targets for a peptide cancer vaccine.

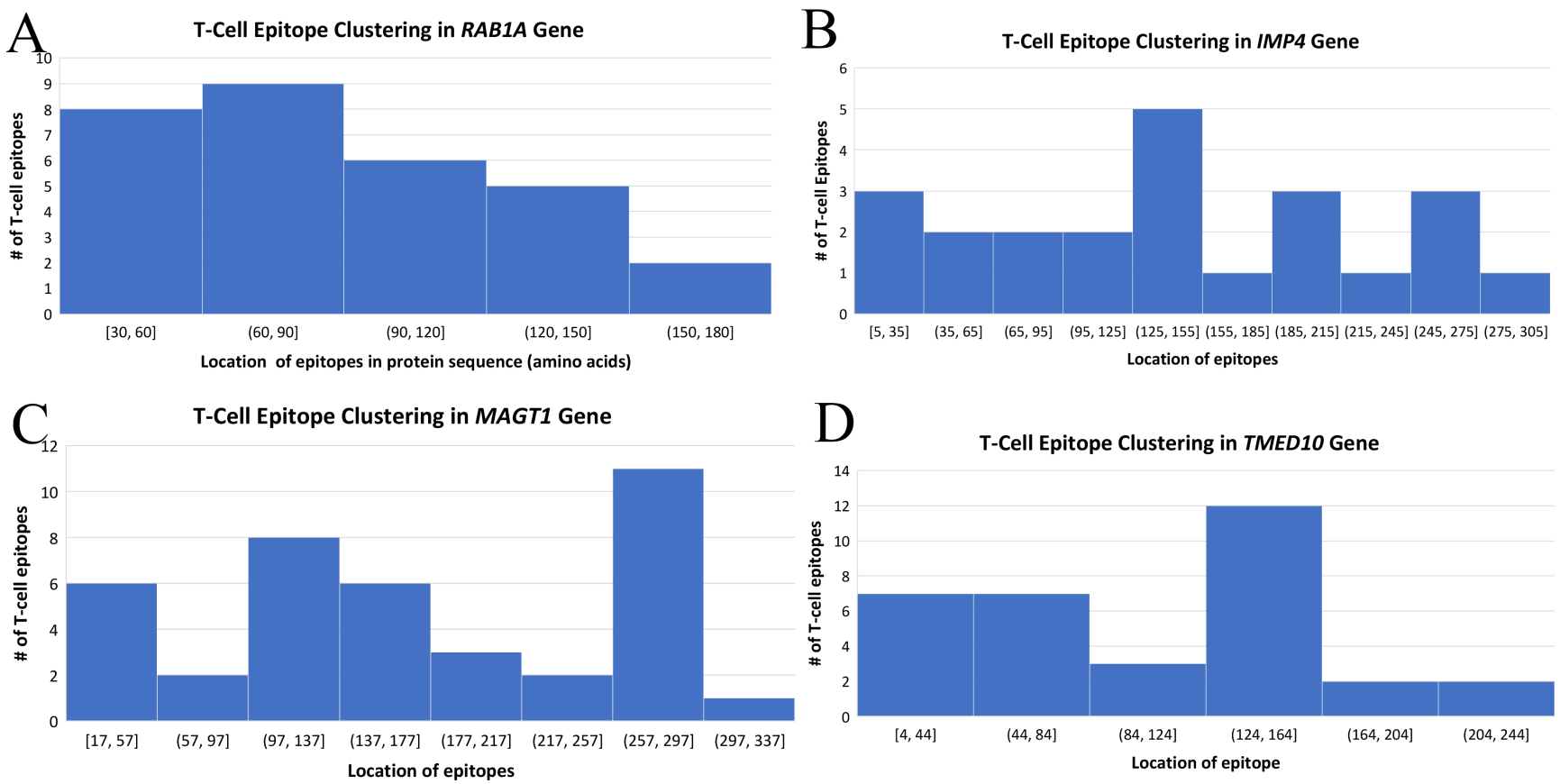

Figure 21. Shows the number of T-cell epitopes in each region of the protein sequences for the top 4 hub genes with the most T-cell epitopes. Peptide regions where there is a high epitope concentration have high immunogenicity. (A) *RAB1A;* (B) *IMP4* (C) *MAGT1* (D) *TMED10*. Note that the actual number (and not normalized) of T cell epitopes in each candidates genes shown.

4. DISCUSSION

This study aimed to design and build multiple machine learning models to predict the occurrence of hepatocellular carcinoma in patients and to predict the immunogenicity of deregulated circRNA gene targets for a potential immunotherapy or cancer vaccine. With Pearson correlation and information gain statistical methods, the top deregulated circRNAs in tumor tissue were determined. All machine learning models used could predict the occurrence of tumors with circRNA expression data with >85% accuracy and many reached nearly 100% classification accuracy. The top machine learning models upon stratified k-fold cross validation were logistic regression and naive bayes. Correlation analysis for the top 2 deregulated circRNAs showed that hsa_circ_0005284 is strongly upregulated in tumor, and hsa_circrna_089372 is downregulated in tumor. From Cytoscape network analysis, miRWalk, and circInteractome, the top hub genes were determined, and 7 of the 8 hub genes were validated to be cancer-prognosis biomarkers with the human protein atlas. Finally, with t-cell epitope analysis in hub gene peptide sequences, it was determined that IMP4, MAGT1, TMED10, and RAB1A have immunogenic potential due to high epitope count and concentration and are good candidates for vaccine targeting. IMP4 was the most immunogenic out of the hub genes validated to be a prognostic marker in liver cancer (the rest were RBM28, IMP4, and KDELR1), but RAB1A and TMED10 showed the highest immunogenicity and can still be potential immunotherapeutic candidates for their respective cancers. However, a study by Yang, et al. (2015) showed and validated RAB1A to be prognostic in hepatocellular carcinoma, although not from the human protein atlas. This indicates some genes can still be prognostic of liver cancer even if they are not validated from the protein atlas. Additionally, novel, unexplored oncogenes for other cancers could be predicted with the pipeline developed in this study since it was validated to work and predict oncogenes by the human protein atlas. Recent reviews provide a comprehensive review on the role of circRNAs in liver and other cancers (Liu et al., 2020; Shen et al., 2021; Su et al., 2019).

In future studies, other methods of machine learning could be explored, including neural networks, hierarchical clustering, and more. These methods could find new trends in circRNA expression data which could not be picked up by classification models. Additionally, pharmacogenomic therapies could be explored such as drug targeting and drug repurposing for the hub genes found in this study. More immunotherapeutic options can also be explored such as immune checkpoint inhibitors and monoclonal antibodies. CircRNA expression data can also be explored for other cancers and the same machine learning pipeline created in this study can be applied to all other cancers.

One of the biggest limitations to this project is data size. The circRNA expression data was for 98 tissue samples. In the future, the researcher might try to find larger datasets on circRNA expression and compare model accuracies. Another limitation is immunogenicity prediction. Multiple methods can be used for immunogenicity prediction and T-cell epitope concentration is only one of them. It is possible that epitope concentration may not be able to fully gauge the immunogenicity of hub gene peptide sequences, so more methods could be explored in the future.

The main future prospects of this project, if validated in biological system, is potential development of active immunotherapy (vaccine) /passive immunotherapy (monoclonal anti-

bodies) therapies for any cancer if circRNA data is given. The novel methodology developed in this project (circRNA and immunogenicity) can help find immunotherapy for any cancer with circRNA data. The machine learning models can be a valuable tool for healthcare professionals because they were over 90% accurate in predicting/diagnosing cancer using circRNA data, potentially replacing existing diagnosing methods such as expensive MRI/CT scans/machines).

AUTHOR INFORMATION

***Corresponding Author**

Aditya K. Koushik
La Cueva High School
7801 Wilshire Ave NE
Albuquerque, NM 87122
Email: adityakoushik1234@gmail.com

ACKNOWLEDGMENT

I would like to thank Dr. Nikolaos Mellios Associate Professor at the University of New Mexico for answering my questions on circRNA data, and my parents for their encouragement and helping me with my project.

REFERENCES

Balogh, J., Victor, D., 3rd, Asham, E. H., Burroughs, S. G., Boktour, M., Saharia, A., . . . Monsour, H. P., Jr. (2016). Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma, 3*, 41-53. doi:10.2147/JHC.S61146

Cavanagh, M., and Findlay, E. M. . ((n.d.)). T-cell activation. *British Society for Immunology*. Retrieved from https://www.immunology.org/public-information/bitesized-immunology/systems-processes/t-cell-activation

Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salmanidis, M., Phillips, C. A., . . . Goodall, G. J. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell, 160*(6), 1125-1134. doi:10.1016/j.cell.2015.02.014

de Candia, P., Prattichizzo, F., Garavelli, S., & Matarese, G. (2021). T Cells: Warriors of SARS-CoV-2 Infection. *Trends Immunol, 42*(1), 18-30. doi:10.1016/j.it.2020.11.002

De Groot, A. S., Moise, L., Terry, F., Gutierrez, A. H., Hindocha, P., Richard, G., . . . Martin, W. D. (2020). Better Epitope Discovery, Precision Immune Engineering, and Accelerated Vaccine Design Using Immunoinformatics Tools. *Front Immunol, 11*, 442. doi:10.3389/fimmu.2020.00442

Dudekula, D. B., Panda, A. C., Grammatikakis, I., De, S., Abdelmohsen, K., & Gorospe, M. (2016). CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol, 13*(1), 34-42. doi:10.1080/15476286.2015.1128065

Garnelo, M., Tan, A., Her, Z., Yeong, J., Lim, C. J., Chen, J., . . . Chew, V. (2017). Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma. *Gut, 66*(2), 342-351. doi:10.1136/gutjnl-2015-310814

Greene, J., Baird, A. M., Brady, L., Lim, M., Gray, S. G., McDermott, R., & Finn, S. P. (2017). Circular RNAs: Biogenesis, Function and Role in Human Diseases. *Front Mol Biosci, 4*, 38. doi:10.3389/fmolb.2017.00038

Henderson, S. R. R. (2021). What are T cells? Retrieved from https://www.news-medical.net/health/What-are-T-Cells.aspx

Liu, J., Zhang, X., Yan, M., & Li, H. (2020). Emerging Role of Circular RNAs in Cancer. *Front Oncol, 10*, 663. doi:10.3389/fonc.2020.00663

Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., . . . Finn, R. S. (2021). Hepatocellular carcinoma. *Nat Rev Dis Primers, 7*(1), 6. doi:10.1038/s41572-020-00240-3

Muhammad, S. A., Zafar, S., Rizvi, S. Z., Imran, I., Munir, F., Jamshed, M. B., . . . Zhang, Q. (2020). Experimental analysis of T cell epitopes for designing liver cancer vaccine predicted by system-level immunoinformatics approach. *Am J Physiol Gastrointest Liver Physiol, 318*(6), G1055-G1069. doi:10.1152/ajpgi.00068.2020

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504. doi:10.1101/gr.1239303

Shen, H., Liu, B., Xu, J., Zhang, B., Wang, Y., Shi, L., & Cai, X. (2021). Circular RNAs: characteristics, biogenesis, mechanisms and functions in liver cancer. *J Hematol Oncol, 14*(1), 134. doi:10.1186/s13045-021-01145-8

Sticht, C., De La Torre, C., Parveen, A., & Gretz, N. (2018). miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One, 13*(10), e0206239. doi:10.1371/journal.pone.0206239

Su, M., Xiao, Y., Ma, J., Tang, Y., Tian, B., Zhang, Y., . . . Wang, W. (2019). Circular RNAs in Cancer: emerging functions in hallmarks, stemness, resistance and roles as potential biomarkers. *Mol Cancer, 18*(1), 90. doi:10.1186/s12943-019-1002-6

Yang, Y., Hou, N., Wang, X., Wang, L., Chang, S., He, K., . . . Huang, C. (2015). miR-15b-5p induces endoplasmic reticulum stress and apoptosis in human hepatocellular carcinoma, both in vitro and in vivo, by suppressing Rab1A. *Oncotarget, 6*(18), 16227-16238. doi:10.18632/oncotarget.3970

Zhang, X., Wang, S., Wang, H., Cao, J., Huang, X., Chen, Z., . . . Xu, Z. (2019). Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Mol Cancer, 18*(1), 20. doi:10.1186/s12943-018-0935-5

2022 New Mexico Research Symposium

ABOUT THE RESEARCH SYMPOSIUM

After two years online, the 2022 New Mexico Research Symposium returned in person, welcoming over 150 people from across New Mexico. Held at The University of New Mexico Student Union Building in Albuquerque on Saturday, November 5th, the 2022 programming included oral sessions, nearly 40 posters, and the addition of the NM Research SLAM, a flash presentation competition for postdoctoral students organized by NM EPSCoR, the Air Force Research Laboratory, Los Alamos National Lab, and Sandia National Laboratories.

Hosted by NM EPSCoR in collaboration with the New Mexico Academy of Science and the Central New Mexico Local Section Chapter of the American Chemical Society, the Symposium closed with an awards ceremony honoring the two 2022 Outstanding New Mexico Science Teacher Award recipients, Colleen Fordyce from La Cueva High School in Albuquerque and Hope Cahill from El Dorado Community School in Santa Fe, the NM Research SLAM winners, and the student poster competition winners.

SYMPOSIUM WELCOME FROM 2022 NMAS PRESIDENT

On behalf of the New Mexico Academy of Science, I would like to announce the 2022 edition of the New Mexico Journal of Science. NMAS is pleased to partner with New Mexico EPSCoR, the UNM Center for Water and the Environment, the American Chemical Society, and New Mexico Space Grant Consortium to sponsor this journal and our annual conference to promote science and science education in our community. This year, EPSCoR did an excellent job in organizing the in-person symposium, after the two previous years of virtual gatherings due to the pandemic. Big compliments are due to NM EPSCoR's resourceful and hard-working staff led by Sara Pichette and Brittney Van Der Werff.

This year, archaeologists and anthropologists worldwide set their eyes on New Mexico in an important way. Previously, the earliest known presence of humans in the Americas, New Mexico's Clovis people, showed that humans came to North America around 13,000 years ago. A recent discovery shows that humans inhabited White Sands, New Mexico, around 22,000 years ago. That discovery was made possible by radiological dating of footprints and organic remains of ancient flora and fauna at White Sands. The biologist who discovered the oldest-ever human footprints in North America, Mr. David Bustos of the National Park Service, was the Keynote Speaker at our conference.

In the new Rio Grande Research SLAM, scientists from the Air Force Research Laboratory, Los Alamos National Laboratory, Sandia National Laboratories, and New Mexico EPSCoR gave three-minute presentations and were judged to receive prizes. As always, the symposium also presented the awards for Outstanding Teachers from our state. Student posters were presented and judged, with awards given to the authors of the best posters. Also new this year is the annual NMAS Poster Award named in honor of the late Prof. Igor Sevostianov of New Mexico State University.

Anton Sumali, PhD, NMAS President

KEYNOTE SPEAKER: DAVID BUSTOS

David Bustos is a wildlife biologist and resource program manager at White Sands National Park. He is a graduate of New Mexico State University. He is responsible for the management of the park's cultural and natural resources and has investigated ancient human and megafauna for nearly two decades. His opening keynote address, "Life Before the Ice Melted: Racing to Preserve Fossil Footprints and Traces of People From 23,000 Years Ago" summarized ongoing research around 61 ancient human footprints found in 2020 preserved in layers of gypsum soil at the park that fundamentally changed the timeline of North American human habitation.

ABOUT THE SPONSORS

New Mexico Academy of Science

Founded in 1902, the New Mexico Academy of Science (NMAS) has been in continuous existence since 1915. NMAS is a member of the National Association of Academies of Science (NAAS) and an affiliate of the American Association for the Advancement of Science (AAAS). NMAS works with teachers, state agencies, and the legislature to establish appropriate standards for the teaching of the sciences. NMAS goals are to foster scientific research and scientific cooperation, increase public awareness of the role of science in human progress and human welfare, and promote science education in New Mexico. Visit www.nmas.org to learn more.

New Mexico EPSCoR

The New Mexico Established Program to Stimulate Competitive Research (NM EPSCoR) is funded by the National Science Foundation (NSF) to build the state's capacity to conduct scientific research. The infrastructure and activities of NM EPSCoR's current project, New Mexico SMART Grid Center are designed to support shared–use equipment, engage new research and community college faculty, and support the STEM pipeline by training teachers, undergraduate and graduate students, and postdoctoral fellows. Research findings are communicated broadly through various outlets, including local museums. Visit www.nmepscor.org to learn more about NM EPSCoR, and visit www.nsf.gov/epscor to learn more about the NSF EPSCoR initiative and other jurisdictions.

American Chemical Society

The American Chemical Society (ACS) is the world's largest scientific society and one of the world's leading sources of authoritative scientific information. A nonprofit organization, chartered by Congress, ACS is at the forefront of the evolving worldwide chemical enterprise and the premier professional home for chemists, chemical engineers and related professions around the globe. The Central New Mexico Local Section of the American Chemical Society was founded in 1946 and generally serves the northern two-thirds of the state of New Mexico. The Local Section specifically includes the following New Mexico counties: Bernalillo, Los Alamos, Rio Arriba, San Miguel, Sandoval, Santa Fe, Socorro, Taos, Torrance, and Valencia.

UNM Center for Water & the Environment

The mission of the Center for Water and the Environment at the University of New Mexico is to increase the participation of underrepresented minorities in science, technology, engineering and math professions while conducting cutting-edge research into technological and engineering-based solutions to problems with water and the environment, in a framework that considers the social, economic, policy, regulatory, and legal implications. Practical solutions to problems related to water availability in arid environments and in times of drought, and problems associated with energy generation and consumption are particularly relevant, in light of the criticality of these issues to the state of New Mexico, the southwestern United States, and their global importance. Learn more at www.cwe.unm.edu.

New Mexico NASA EPSCoR

NM NASA EPSCoR provides competitive seed funding and advice to faculty who are conducting research that aligns with NASA Mission Directorates and/or NASA Field Centers (including JPL). The goal is to provide funding to develop competitive research and technology projects and programs for the solution of scientific and technical problems of importance to NASA. NM NASA EPSCoR also contributes to the overall research infrastructure, science and technology capabilities, and/or economic development of the state. Various funding opportunities are available from NASA EPSCoR such as the Research Cooperative Agreement Notice, International Space Station Flight Opportunity, Suborbital Flight Opportunity, and Rapid Research Response. Learn more at www.nmspacegrant.com.

Outstanding Science Teacher Award

ABOUT THE AWARDS

Since 1968, the New Mexico Academy of Science (NMAS) has awarded the Outstanding Science Teacher Award to honor New Mexico science and math educators. This award recognizes teachers who provide opportunities for students to succeed. Nominations are open to all preK-12 teachers and informal science educators throughout New Mexico. NMAS presents a plaque and a monetary award to each teacher. The American Chemical Society also presents a monetary award to the winning teachers.

2022 Recipients

The 2022 New Mexico Outstanding Science Teacher Award winners are Colleen Fordyce from La Cueva High School in Albuquerque and Hope Cahill from El Dorado Community School in Santa Fe.

Colleen Fordyce is a biology and AP biology teacher at La Cueva High School in Albuquerque. She earned a PhD in Biomedical Sciences from the University of New Mexico and pursued postdoctoral studies at the University of California, San Francisco where she applied insights from her doctoral studies to describe cellular responses to telomere shortening which were associated with increased risk for breast cancer. While pursuing a career as a research scientist Dr. Fordyce also sought out opportunities to teach. Ultimately, Dr. Fordyce decided to hang up her pipettes and pursue teaching full-time. She has taught at La Cueva High School in Albuquerque since 2017. Dr. Fordyce is motivated by several factors, including a desire to prepare her students to be empowered science consumers and to inspire them to pursue STEM fields in college. She is humbled and inspired by her amazing students who laugh as hard as they work.

She brings her experience as a research scientist to her classroom and has developed unique labs that expanded the hands-on experience for students while linking to real-world problems. She uses her ties to researchers to bring guest speakers to her students and during 2019 when APS schools closed, she created a virtual scientific conference with a series of weekly virtual seminars for her students presented by researchers from across the country. During the summer of 2020, Dr. Fordyce co-authored an entire NGSS-aligned Biology unit which could be delivered virtually for APS schools. She is the La Cueva HS Science Bowl Mentor and a member of several teams of teachers at La Cueva HS who support struggling students and identify areas and strategies for school-wide improvement.

Hope Cahill is a 6th grade science teacher at El Dorado Community School in Santa Fe. She began as a Language Arts and Creative Writing teacher. Since 2012, she has been a science teacher for 6th and 7th grades who occasionally picks up a section of language arts. She has fostered connections between the NM content standards and curriculum and New Mexico's rich geologic and scientific history. She has integrated literacy and technology into her instruction and has created a curriculum that accommodates a variety of abilities and learners within the classroom setting. She authored the article "Place-Based Educational Activities Inspired by the Socorro Magma Body" in the Summer 2021 issue of The Earth Scientist, a quarterly journal of the National Earth Science Teachers Association (NESTA). Over the years she has assumed additional responsibilities, including the remote-learning Science Design Team, Middle School Science Fair Coordinator, Global Warming Express mentor, Mentor for Santa Fe High School Supercomputing Challenge team, and STEM Pathways for Girls Conference planning committee and group guide. She is a 2020 recipient of the "Presidential Awards for Excellence in Science Teaching" and a 2021 recipient of the "Partners in Education Teachers Who Inspire" Award. She was a Thornburg Corporate Giving Program grant recipient in 2021.

Making it Click: Synthetic Tools for Multi-Stage Diversification

Brian Gold, University of New Mexico

Modularity enables diversity. Since its inception two decades ago,"click chemistry" has both embodied this idea and driven innovation throughout the molecular sciences. The key to this strategy, success lies in its simplicity: molecular building blocks are clicked together using "near-perfect" reactions. We continually seek to harness fundamental principles of chemical structure and reactivity in the pursuit of modular synthetic tools. Recent progress will be discussed.

Colorimetric and Fluorometric Sensing of Various Carboxyl Anions with dinuclear Ni(II) and Cu(II) Complexes of Polyamine Macrocycles Following Indicator Displacement Assay

Md Mhahabubur Rhaman, Eastern New Mexico University

There are various carboxyl anions that play important roles in biological systems, biomedical science, and the food industry. For example, the quantitative information of oxalate in urine is commonly used in the diagnosis of several diseases including hyperoxaluria and vulvodynia. Glutamate can excite cells to death in a process now referred to as "excitotoxicity". So, it should be present at the right concentration at the right time at the right places. The amount of it in food should be monitored because it is used as a flavor and taste enhancer in food. Citrate is another carboxyl anion whose presence in urine is considered to inhibit the crystallization of calcium salt. Its high concentration in urine indicates the growth of kidney stones and urological diseases such as nephrolithiasis and hypocitraturia. Therefore, the easy and cheap method of quantitative detection of carboxyl anions is important. In our research, several polyamine macrocycles were synthesized by Schiff's base reaction between 2,2'-diamino-N-methyldiethylamine and various aromatic dicarboxaldehyde in high dilution conditions followed by NaBH4 reduction. Then, these macrocycles were converted to dinuclear Ni(II) and Cu(ii) complexes and studied for recognition of carboxyl-containing anions by indicator displacement assay using several commercially available dyes Eosin Y, Pyragallol red, and pyrocatechol violet.

Acknowledge: The project was funded by the US Department of Defense (Grant number W911NF-19-1-0006). The National Science Foundation supported the purchase of the diffractometer (Grant CHE-0130835).

Effect of Processing Route on Properties of NbMoTaWVTix Refractory High Entropy Alloys

Surya Bijjala, University of New Mexico

The four core effects of high entropy alloys (HEAs) intrigued the researchers towards the exploration of HEAs, in particular, HEAs made from refractory elements, as high-temperature materials. Of all the refractory high entropy alloys (RHEAs) studied, NbMoTaWVTix RHEAs have shown good strength retention and excellent thermal stability. However, while these RHEAs have the potential to achieve desirable performance targets, manufacturing these alloys is challenging as casting produces segregations due to differences in the elemental melting temperatures. The powder metallurgy approach eliminates these effects but suffers from porosity and impurities. We have studied different PM routes like pressure-less sintering, Hot isostatic pressing, and Spark plasma sintering with variable parameters for manufacturing dense RHEAs. The effect of the processing route on the properties of RHEAs is presented.

This research is based upon work supported by New Mexico NASA EPSCoR office under a NASA Rapid Respond Research (R3) Cooperative Agreement No. NM-80NSSC21M0171.

Pill Bug-Inspired Robot for Lunar Exploration

Chase Dunaway, New Mexico Institute of Mining and Technology

Through the Artemis Program, NASA intends to conduct prolonged scientific research and exploration of the lunar surface with the eventual goal of establishing a long-term human presence on the moon. Specific interests have been placed on lunar lava tubes, caves, and craters as potential sites for habitation. As a result, NASA is pursuing possibilities utilizing advancements in autonomous navigation and swarm robotics to explore the lunar surface in the near future.

The need for robust and versatile robotics systems is necessary to traverse the foreign lunar environment efficiently and conduct scientific operations. Proposing the Fengaribot, a robot capable of ascending steep, rocky terrain while having the ability to transform its shape to roll down inclines for a controlled, energy-efficient descent. Biologically inspired by the pillbug, the Fengaribot's frame is based upon a segmented shell-like approach with a twelve-legged locomotion system. Where a heavy-bodied rover may have been blocked, the Fengaribot's low center of gravity allows the robot to climb unleveled terrain without flipping. Furthermore, the transformation into a rolling state allows for the controlled descent of a sloped surface while protecting sensitive components and preventing a scenario in which the robot topples into an irrecoverable state. The robot's electronics are mounted separately in the shell segments with the capability of storing various scientific instruments depending on the task.

This research is based upon work supported by the New Mexico Space Grant Consortium (NMSGC) Space Grant Scholarship through a NASA Cooperative Agreement No. NM-80NSSC20M0034

Efficient Microgravity Heat and Mass Transfer with No Moving Parts

Joel Gates, University of New Mexico

In our work with NASA EPSCOR, we propose to demonstrate an efficient, noncontact method of transporting heat via fluids in microgravity. This method does not use any moving parts and only requires a modest power supply that drives three mutually orthogonal, low-impedance Helmholtz coils. This differs from current methods of heat transfer in microgravity which traditionally require mechanical mechanisms, such as a pump, to drive artificial convection. Extensive testing of artificially created convection currents with this method has been carried out on Earth and there is evidence that using such technology could result in a more efficient heat transfer process in microgravity. This will be tested through the implementation of an orbital payload using an infrared camera for visualization and quantification of flow as driven by Helmholtz coils in the previously described configuration acting on magnetic nanoparticles inside a fluid capsule. The demonstration of efficient heat transfer via this method allows for the development of low-power heat transfer systems for potential use in microgravity environments where traditional mechanical systems have the potential for failure.

This research is based upon work supported by New Mexico NASA EPSCoR office under a NASA International Space Station (ISS) Cooperative Agreement No. NM-80NSSC20M0146.

Predictive Analysis of Genetic Radiation Resistance Mechanisms

Katheryn Perea-Schmittle, New Mexico Institute of Mining and Technology

Extremophilic bacteria and archaea are able to tolerate traditionally uninhabitable environments. Understanding the cellular mechanisms necessary for this survival has implications ranging from medicine to bioremediation to astrobiology. One such extreme is survival to high levels of ionizing radiation. These highly resistant species are candidates for model organisms in fields where radiation resistance has a large influence. Over the last couple of decades, a large body of work has focused on molecular mechanisms for radiation resistance. While a wide variety of mechanisms contributing to radiation resistance have been identified, how they specifically result in radiation resistance is largely unknown. However, metagenomic and transcriptomic studies have uncovered a multitude of genetic trends related to radiation resistance. My PhD work centers on the creation of a machine-learning model-based computational tool that, using these trends, will:

- Identify which known genomes a naïve genome is most similar to based on suspected genetic markers for radiation resistance.
- Determine the impact of previously identified genetic trends on the prediction of radiation resistance.
- Determine the degree of radiation resistance of a microorganism based on its genomic composition.
- Identify new genetic trends associated with radiation resistance

This work required the creation of a specialized database focused on molecular radiation resistance mechanisms, the development of a tool featuring machine learning models designed for both classification and regression-based analysis, and the validation of the tool using both new and established data sets.

This research is based upon work supported by the New Mexico Space Grant Consortium (NMSGC) Space Grant Fellowship through a NASA Cooperative Agreement No. NM-80NSSC20M0034.

Examining Spatial Heterogeneity and Tourism Potential of Water Resources in New Mexico

Jason Banegas, New Mexico State University

Increasing outdoor tourism in New Mexico and the growing demand for water-based tourism amenities requires developing innovative strategies for resource allocation and service integration among many existing and potential recreation locations. Identifying tourism potential and optimizing resource allocation has been addressed in studies using a multitude of models, methods, and algorithms; however, variations between the influential factors of tourism locations are best analyzed with a Geographical Information System (GIS) based research model that can examine diverse spatial relationships in these environments. This study uses a Geographically Weighted Regression (GWR) method to illustrate that variables related to drivers of tourism are locally distinct. This tool is used to examine the spatial non-stationarity of relevant water-based tourism variables. Spatial discrepancies are shown to exist in the explanatory variables useful for developing tourism resources in rural and urban locations in the New Mexico study area. This GWR analysis can strengthen the decision process and allow policymakers to identify distinct influential factors and areas where resources for expanding water-based tourism should be focused.

Dual Turn-On/Turn-Off Sensing of Acetylacetone and Turn-On Sensing of Water in Organic Solvents

Alisha Gogia, New Mexico Highlands University

Metal-organic framework (MOF)-based sensors for the detection of various analyte molecules have been a subject of absolute importance. However, most of these sensors rely on the turn-off (quenching) transduction response, while those reporting turn-on responses are very rare. In this article, we have synthesized two new MOF-based sensors, $\{[Zn_2(oxdz)_2(tpbn)]\neg\Sigma 14H_2O\}_n$ (1) and $\{[Zn_2(oxdz)_2(tpxn)\neg\Sigma 10H_2O{\cdot}2C_2H_5OH\}_n$ (2), via the self-assembly of Zn(II) metal ions, a fluorogenic $(oxdz)_2$ linker, and bis(tridentate) ligands (tpbn and tpxn) under ambient conditions. Their formation from such a self-assembly process has been evaluated on the basis of the geometry around the five-coordinated Zn(II), preferential meridional binding of the bis(tridentate) ligands, and diverse binding of the carboxylate groups in $(oxdz)_2$. Although 1 and 2 are isostructural, a difference in the transduction mechanism for the sensing of acetylacetone in organic solvents (turn-on for 1 and turn-off for 2) is observed and can be attributed to the spacer in the bis(tridentate) ligands. We have demonstrated the competing effect of the nonradiative interactions and photoinduced electron transfer toward the sensing mechanism. The results are well-supported by the Fourier transform infrared spectroscopy study, intensity versus concentration plots, spectral overlap measurements, time-resolved fluorescence studies, and MM2 and density functional theory calculations. Furthermore, we have showcased the utilization of 1 for the sensing of trace amounts of water in organic solvents.

Comparing the Growth of Xeric Trees in Soil vs. Aquaponics

Conner Wood, Santa Fe Community College

In my experiment, I was comparing the growth of xeric trees that grow in the southwest in soil with traditional methods vs. aquaponics. I felt as though this application could provide a potential solution to fighting climate change, saving massive amounts of water, as well as contributing to habitat restoration efforts. As I will also have shown in my poster, I examine in greater detail factors that are common in the field of aquaponics such as PH, EC, $N0_3$ concentration, and PAR. I also provide graphed data on growth rate comparison for the height and width of each tree species. I also provide and discuss data on root mass comparisons between the soil control group and the aquaponic group. This is a study that took place over one year from the summer of 2021 to the summer of 2022. I break apart each trial in detail and discuss the different findings of each. I am comparing different methods of cultivation so within each species, all of the seeds were sown and dated at the same time. This established a fair starting point to begin the measurement of growth comparison between the soil and aquaponic groups. I will demonstrate how I transplanted them into the aquaponics system as well as how I recorded data and cared for them. There are more than several species of trees that grow in New Mexico as well as the surrounding states and Southwest that are included and I examine each one of their growth rate comparisons. I also identify and explain similar growth rate patterns identified in the aquaponics group. I will also discuss methods for a successful transplanting of the trees from a water-based media into a soil media including specific media used to help the transition. Overall after all three trials, I answered my original questions which were, Can the growth rate of trees be accelerated with aquaponics vs. growing them in soil with traditional methods? As well as, Can larger root masses be cultivated while growing trees in aquaponics? The answer was Yes, for both questions and all of the species in the three trials. I believe that my findings of achieving faster growth rates of trees with aquaponics have the potential to help produce trees faster which has many applications. I will discuss how this method can be utilized to speed up habitat restoration and possibly help curb the effects of climate change by speeding up tree production which was previously thought would take much longer to produce. Aquaculture can be a sustainable form of food in areas that greatly limit food production. I will suggest combining aquaponics and xeric trees could be a game changer for sustainability habitat restoration and sustainable aquaculture.

Analysis of Asynchronous Communications on Microgrid State Estimation

Mohammad Afrazi
New Mexico Institute of Mining and Technology

Average consensus algorithms in microgrids are becoming attractive concepts to meet the increasing demands for energy. The goal of the average consensus is to calculate the average of initial values across dispersed nodes by sharing data via communications. The Average Consensus Theorem relies on the local information of agents to guarantee that important information is shared in a distributed way. It is known that synchronization is necessary to obtain an accurate average across the distributed nodes. Due to several technical issues, however, such as the requirement of a global clock and random packet drops/delays, which may occur during the implementation stage, the synchronous model is highly limited in reality. An asynchronous model, a synchronous one's counterpart, naturally takes into consideration the problems mentioned above and is thus more practical. The only downside of the asynchronous model is that it may result in inaccuracies in the consensus value which requires more thorough investigations. To this end, the performance of different network topologies has been investigated in this study for asynchronous systems with 16 different agents and random initial values. We analyzed three important performance factors: the convergence speed, error bound width with respect to the exact value, and the error variance.

Use of Waste Bioproducts as a CO_2 Sink in Sustainable Structural Composites

Marcos Hernandez, Los Alamos National Laboratory
Marcos M. Hernandez, Adalee R. Witt, Yuliia Trujillo, Cesar Raul Gonzalez Esquer, Babetta Marrone, Joseph H. Dumont

By identifying useful waste streams, bio-manufacturing technologies such as sustainable composite materials can be developed to offset CO_2 emissions. Waste biomass offers many material advantages such as biosequestration, biomineralization, and improved mechanical properties. By incorporating bio-waste into cementitious, pressed, and hydrogel materials we aim to develop bio-composites for applications in structural materials, 3D printing, and fuel pellets. Samples were prepared and characterized using a wide range of chemical, mechanical, and spectroscopic techniques. From this study, we evaluated the effects of waste biomass in sustainable composites for CO_2 mitigation, reduced construction cost, and waste utilization potential. LA-UR-22-26869.

IoT Enabled Healthcare Monitoring System

Md Apu Sayeed
Eastern New Mexico University

To address the drawbacks of conventional healthcare and satisfy the increasing demand for high-quality care, a large amount of research is presently focused on smart healthcare. This work presents an IoT-enabled healthcare monitoring system for building smart healthcare. This work consists of three parts: a sensor unit, a detection unit, and an IoT unit. The sensor unit continuously monitors and records patients' health activity using sensors. Detection unit machine learning algorithm to analyze signals and detects abnormal patterns. IoT unit enables a physician or authorized personnel to keep track of a patient's health status through the internet. The system displays real-time information on a person's status via IoT. The last part is contributing the disclosed information for remote surveying. Remote surveying of the data enables a doctor or guardian to monitor a person's health improvement away from the person's premises. This device allows a person to be monitored continuously. The system makes use of an Arduino to achieve this functionality. It makes use of sensors like temperature sensors, pulse oximeter sensors, and GSR sensors to keep track of a person's health. This project also helps the person to get a medical alert and a fall detector is used to give alerts to the patient's guardian using IoT when elderly people fall from the bedside. The sensors are secured to a microcontroller of the Arduino to track the condition which is in turn connected to a wi-fi module to transfer the real-time notifications through the internet. The system shows detail of a person's status over IoT in real time. Thus IoT builds a health tracking system that beneficially uses the internet to observe health conditions and protect individuals in due time.

Learning to Run a Power Grid with High Penetration of Renewables

Di Shi, New Mexico State University

Renewables are transforming power systems, posing grand challenges to system operations. Increasing data with critical operational insights, on the other hand, offers new opportunities in addressing these challenges. However, there are no effective approaches that can collect and synthesize massive measurements from tens of thousands of smart sensors over wide areas to make timely decisions on how to best allocate energy resources. This talk presents a data-driven and learning-based framework for autonomous power systems, and a corresponding software platform named Grid Mind. As recently Grid Mind has been deployed at six grid dispatch centers, results from the real world will be discussed.

UNDERGRADUATE STUDENT POSTER ABSTRACTS

Poster session abstracts are listed alphabetically by last name of registered presenter. An asterisk (*) indicates the poster received an outstanding undergraduate poster award at the 2022 New Mexico Research Symposium.

Examining the Accessible Chromatin Regions Between Duplicated Gene Copies in the Tetraploid *Leucaena trichandra*

Katie Banga, New Mexico State University Donovan Bailey, New Mexico State University

Chromosomal and genetic duplications are key sources of genetic novelty in evolution. When duplication events occur, two copies of the genes are usually not necessary, which results in one being lost through mutational processes. However, duplicate copies can also be retained, going on to gain new functions or subfunctions between the daughter genes. For these reasons, the duplicated regions are often expressed differently than the original copy. The plant genus *Leucaena,* which contains 19 species with tetraploid genomes, provides exceptional opportunities to investigate the evolutionary changes brought forth by chromosomal and genic duplications. Species in the group are good models because they grow relatively quickly for trees and they have many uses as multipurpose crops in the tropics. Here we use the genome of the tetraploid, *L. trichandra* because it is a parent of other octoploid species within the genus. Chromatin accessibility studies, such as the one conducted here, include methods used to identify heterochromatic DNA regions open to transcription factors and therefore expression. We used an Assay for Transposase-Accessible Chromatin (ATAC-seq) to determine the changes to the accessible chromatin regions between different copies of the chromosomes in the *L. trichandra* genome. This is possible because a Tn5 transposon is inserted into open chromatin regions during the library preparation. The results of the ATAC-seq analysis will provide an important annotation for future research and allow us to make conclusions about the accessible regions that have experienced gene expression differences since the duplication of the chromosomes. The data obtained from ATAC-seq will identify regions of accessible and inaccessible chromatin, providing key information regarding how chromatin accessibility varies among the different copies and the relative fate of duplicated gene copies.

Innovating a Three-Level Boost Converter Design for Enhanced Efficiency and Component Longevity in Photovoltaic Systems

Luis Carranza, New Mexico State University Sijo Augustine, New Mexico State University

The idea of this project was to innovate a Three-Level Boost Converter which will be implemented into a Photovoltaic system. Although there have already been many different designs of DC Boost Converters, we have taken it upon ourselves to create a design that will be more effective to avoid the replacement of an entire Printed Circuit Board due to any shorts in the circuit or other malfunction in components such as voltage measuring devices or the Gate Driver component. Although I am sure there have been designs implemented where they have separated the voltage sensors and Gate driver components, we believe that there can be a more efficient way. This can help reduce error as well as a cost during the manufacturing process as well as in the long term use of the DC Boost Converter. During this research we took an existing design of a DC Boost Converter and decided to innovate by creating a PCB that is strictly the Power Circuit, we also proceeded to create another board for auxiliary power, a board for the gate driver and controller, and a series of different boards that can be used as voltage sensor (measurement devices). Overall the project is still in process but we have discovered that we will have a more efficient design that will increase the efficiency of the converter as well as elongate the life of the components.

Comparing Semi-Supervised Anomaly Detection Algorithms that use PMU Data to Detect Cyber Attacks in Smart Grids

Crystal Cevering, University of New Mexico
Allison Kilpatrick, University of New Mexico
Larry Maes, Eastern New Mexico University
Damian Lovato, New Mexico Tech

There are many different machine learning algorithms capable of detecting cyber attacks. Previous methods have used supervised algorithms that use both normal and attack data to train a detection model. Supervised learning may have poor performance in detecting cyber-attacks because it is difficult to get various instances of attack data. Whereas semi-supervised anomaly detection algorithms rely only on normal event data to train a detection model, making it more capable of detecting unknown attack types. Data is gathered from the Phasor Measurement Unit (PMU), which is a sensing device that provides real-time measurements of the power system state. The data from the PMU is then used to train and improve the accuracy of learning algorithms.

Bacterial Communities Surrounding the Isleta Division Dam on the Rio Grande: Implications for Crop Safety in the Southwestern United States

Alexa Davis, UNM-Valencia Campus
Marissa Vigil, UNM-Valencia Campus
Benjamin Flicker, UNM-Valencia Campus

The Rio Grande stretches across the lower southwest region of the United States. It goes near the farmlands in Alamosa Colorado, right through New Mexico, parts of Texas, and ending in the Gulf of Mexico. The Rio Grande has 15 dams; we chose to study the bacterial communities found around the Isleta Division Dam, New Mexico (34.90602 N, 106.68634 W). The Isleta Diversion Dam diverts water from the river into irrigation canals that run through the Isleta pueblo. In the Rio Grande, there can be harmful bacteria that grow and get transferred into our crops. Our project is centered around using 16s sequencing which is used to find specific genetic codes for bacteria with over 89% accuracy (Illumina, 2022). We sampled around the dam upstream in the morning and evening and downstream in the morning and evening. We took eight samples altogether, we took four samples at 6 am and four samples at 6 pm: Sample U1AM & U1PM, upstream was set 5-10 feet upstream from the dam. Sample D1A & D2P, Downstream was set around 50 feet downstream from the dam. In our results, we found the 7 most common phyla of bacteria both upstream and downstream included *Proteobacteria, Bacterioidetes, Actinobacteria, Verrucomicrobia, Firmicutes, Cyanobacteria*, and *Chlorflexi*, with *Proteobacteria*, have the highest number of reads.

Signals Features for Classification of Power System Disturbances using PMU Data

Daniel Diaz, New Mexico State University
Alexa Templeton, University of New Mexico
Patrick Jojola, New Mexico State University
Jeremy Nichols, University of New Mexico

Disturbance classification in our power systems is vital in ensuring that energy, which powers our communities and homes, stays protected and under control. With the usage of PMU data collected from the Phasor Data Concentrator (PDC), we are able to implement machine learning techniques to accurately classify the status of relays or disturbances. This poster shows that S3 feature extraction is the better method overall by capturing the trend of the sequence by calculating the changes in sequence values. This, in turn, helps classify signal disturbances.

Growth of Long Emission Wavelength (1.3 – 2 μm) InAs Quantum Dots using Metamorphic Buffers

Carter Heinrich, University of New Mexico
F. Ince, University of New Mexico
M. Moreno, University of New Mexico
G. Balakrishnan, University of New Mexico

InAs Stranski-Krastanov (SK) quantum dots have been used for single photon emitters. The SK growth mode involves the growth of a planar wetting layer of InAs on a substrate like GaAs, and subsequently, the high mismatch drives the growth three-dimensional leading to an ensemble of nanoscale islands or quantum dots. This growth phenomenon has been observed in a few material systems; however, InAs on GaAs has been the most widely used embodiment of this growth resulting in lasers, solar cells, and detectors. The coverage of InAs that is used for quantum dots ranges from 1 ML which is a pure wetting layer to up to 2.7 MLs. Beyond this, the dots experience Oswald ripening and result in photonically inactive quantum dots. This translates to a wavelength range of ~ 1000 nm to 1250 nm. While these quantum dots are typically very high in density at > 10^10 dots/cm^2, through growth methods the density can be reduced, and using advanced microscopic techniques a region with a few quantum dots can be probed. In this paper, we attempt to increase the emission wavelength of InAs quantum dots to telecommunication wavelengths of 1.33 μm and subsequently to 1.55 μm. To do this, we must be able to grow larger quantum dots and that will require reduced mismatch between the quantum dot and the underlying substrate. We, therefore, employ a metamorphic buffer on GaAs comprised of step-graded AlGaAsSb. The results presented will include XRD and TEM-based characterization of the buffer along with AFM and photoluminescence-based analysis of the quantum dots.

A Machine Learning Approach for Optimizing Antenna Design*

Matthew Kube, Eastern New Mexico University
Sarbagya Ratna Shakya, Eastern New Mexico Uni.

With an increasing demand for smarter antenna design in advanced technology applications like 5G, IoT (Internet of Things), and WLAN applications, the development of antenna design has been a complicated process, which requires multiple iterations and extensive testing to produce a final product. With recent development in Machine Learning (ML) algorithms and the availability of data for antenna design, we investigated different ML algorithms, for optimizing the output strength of an antenna by analyzing different dimensions of the patch antenna. We utilized Regression based ML models to learn the behaviors and efficiency of the antenna and predict the output strength (S11) for a range of frequencies. The preliminary comparative study of different algorithms shows that Random Forrest (RF) has higher accuracy as compared to other ML algorithms. The application of this study will help to analyze the design using ML algorithms and will provide a quicker and smarter antenna design as compared to traditional design methods.

Concerns of the 2016 WHO Diagnosis Criterion for Polycythemia Vera

Sophia Lim, Albuquerque Academy
Amalia S. Parra, University of New Mexico
Christopher A. Johnston, University of New Mexico

Polycythemia Vera is a Myeloproliferative disorder involving an increase in red blood cells and white blood cells. Increased red blood cells result in hyperviscosity of the blood causing it to clot and interfere with blood flow through small blood vessels. In 2016, the WHO altered its guidelines regarding the diagnosis of PV from 2008 due to the emergence of "masked PV" for which patients fared worse outcomes, which mimicked Essential Thrombocythemia, increasing the risk of thrombosis due to delayed diagnosis. Modern diagnosis status for PV requires the mutation status of JAK2V617F. 3 While more accurate the 2008 guidelines, there are still issues due to the limitations in JAK2 testing caused by a lack of a broadly-used method. These can include qPCR, PCR,

melting curve analysis, pyrosequencing, RFLP, and Sanger sequencing as well as the intellectual property rights of Ipsogen which influence the choice of method. In this poster, I will highlight the issues of the multiple methods and the prospect of endogenous erythroid colony formation, which is highly specific for JAK2 mutations and could be an effective addition to diagnostic criteria.

Using Bisulfite Sequencing to Characterize the Fate of Duplicate Gene Copies

Mike Lopez III, New Mexico State University
Donovan Bailey, New Mexico State University
Katie Banga, New Mexico State University

The field of Epigenetics is a relatively new field of scientific study and research, with the focus being on the effect that nature has on an organism's development. However, there are still specific areas of epigenetics that are not fully understood. The current literature describes several mechanisms that play key roles in how an environment can cause changes in an organism's genetic structure over time, such can be described as the fate of duplicate gene copies; Pseudogenization, conservation of gene function, subfunctionalization, or neofunctionalization are all ways the genome deals with duplicated genes. We are focused on the methylation patterns in the tetraploid Leucaena trichandra genome as an example of what can happen to genic methylation following genome duplication. These patterns are identified by a methyl group attached to the phosphate group of a nucleotide in the genome; methylation (presence of a methyl group on a nucleotide) is commonly found on cytosines, with some adenines being methylated as well. The goal of our project is to use bisulfite sequencing (a type of gene library that accounts for methylation) to detect the methylcytosine patterns in the *Leucaena trichandra* genome to identify gene copies that are likely to be available (unmethylated) or unavailable (methylated) for expression and function. The expected results of our experiment include a description of the Leucaena genome that shows three classes of methylation patterns for duplicated genes, including only one copy methylated (one copy retains function), neither copy methylated (both copies retain function), or both copies methylated (neither copy retain function).

Model and Power Flow Analysis of a Distribution Feeder using Data Recorded from Buildings, Facilities, and a Photovoltaic System*

Adrian Maez, New Mexico Tech
Casie Taylor, New Mexico Tech
Dawn Walaitis, New Mexico Tech
Kevin Wedeward, New Mexico Tech

This poster presents the model and power flow analysis of a distribution feeder developed for academic studies related to the optimization, control, and integration of distributed energy resources such as sources of renewable energy and energy storage systems. The 16-bus, three-phase model was created to provide a realistic test feeder the size of a small campus where radial underground cables connect the utility, a primary source of power to loads and a photovoltaic system. Measurements of power collected from actual buildings, facilities, and a photovoltaic system serve as the basis for the model and enable computer simulations to be performed with realistic profiles over time. A number of models of distribution test feeders are available from IEEE and others, but few are on the scale of a small campus, and fewer yet have associated time-series data. The proposed model consists of underground cables represented by total phase impedance and shunt admittance matrices, seven loads (five buildings, heat plant, and chiller plant) that have a peak total of approximately 2,000kW and have time-series values for average and reactive power, and a 346kW-rated photovoltaic system that has time-series values for average power. To provide options and confirm results, the model and associated power flow equations were solved using hourly data over multiple days in three software packages (Gridlab-D, OpenDSS, and MATLAB). The outcome of this effort is a new test distribution feeder with time-series solutions that may be used by researchers to develop more sustainable and resilient power distribution systems in the future.

Survival Motor Neuron Uncovers a New Link Between Splicing and Defense Response in *Arabidopsis thaliana*

Jayden Montoya, Northern New Mexico College
Regina Bedgood, Northern New Mexico College
Xaun Zhang, University of Georgia
Bob Schmitz, University of Georgia
Tristan Maestas, Northern New Mexico College
Haidong Yan, University of Georgia
Michael Gonzales, University of New Mexico
Mario Izaguirre-Sierra, Colorado State Uni.

Spinal Muscular Atrophy (SMA) is an autosomal recessive genetic disorder found in metazoans that are associated with the loss of motor neurons caused by mutations within the survival motor neuron (SMN) gene. A complete knockout of motor neurons and the progressive degeneration of muscle tissue. SMA is of the SMN gene that results in embryonic lethality in humans as well as animals such as mice and flies. The SMN protein is a muscle-based protein, which is encoded in two genes in the human genome; plants contain only one SMN gene although they lack muscular structures. Interestingly, SMN mutant plants are nonlethal. SMN protein is necessary for the assembly and transport of small nuclear ribonucleoproteins (snRNPs) which are required for the spliceosome in all eukaryotes. Our preliminary data showed that several defense response genes are over-expressed in our SMN mutant background. Specifically, my goal is to use molecular genetics and bioinformatics to understand the role of SMN and alternative splicing in defense response. Therefore, I am studying the chromatin behavior during and after a pathogen attack using Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq). Investigating the chromatin accessibility profiles in SMN mutant plants during infection will allow us to understand the genomic response during pathogenesis.

Comparison of Bacterial Communities in the Middle of the Rio Grande

Aida Nevarez, UNM - Valencia campus
Ben Flicker, UNM - Valencia Campus
Jarret Attell, UNM - Valencia Campus
Victor French, UNM - Valencia Campus

While soil and aquatic bacteria are critical to ecosystem function, their abundance, and diversity in river ecosystems, particularly the Rio Grande, have not previously been studied. The impact of wastewater treatment effluent discharge on these bacterial communities is also not well understood. In this study we used the 16s rRNA sequencing method to elucidate the diversity of bacteria in three different regions of the middle Rio Grande in New Mexico: Los Lunas upstream of the municipal effluent discharge; the site of the Los Lunas wastewater treatment effluent discharge into the river; and Belen, south of the Los Lunas municipal effluent discharge. Both water and soil samples were collected for whole genome extraction. We also compared the three sites in abiotic factors such as dissolved oxygen, conductivity, and temperature. We will present the bacteria found using these sequencing data and make comparisons of the species diversity between these three communities based on statistical analyses. We will also provide comparisons of the species diversity in collections from the fall of 2019 through the winter of 2022.

The Effects of Forest Fires on Bacterial Communities

Savanna Parada, University of New Mexico
Julia Aragon, University of New Mexico
Fatima Martinez, University of New Mexico
Benjamin Flicker, UNM - Valencia Campus

Wildfires can cause a significant impact on the surrounding ecosystem. It can be difficult to determine the effect these wildfires can have on microorganisms. For our experiment, we collected four total samples. Each pair was collected along the Rio Grande in Jarales, New Mexico (34.5518045 N, -106.5819789 W), which experienced a wildfire in June of 2021. Visibly burned soil was collected for samples A1 and A2 while visibly unaffected soil was collected for samples B1 and B2. We used 16s sequencing to determine the diversity of the classes of bacteria found. The results were consistent with the most abundant classes being *Actinobacteria, Alphaproteobacteria,*

Gammaproteobacteria, Deltaproteobacteria, and *Betaproteobacteria* for all samples. These results do not offer any evidence that burned soil has less diversity in the bacteria found compared to regular soil.

Phytoremediation

Serena Prather, Santa Fe Community College

The Ambrosia Lake disposal site is the home of a nonoperational Uranium ore processing facility in New Mexico. The facility stopped operations in 1982. Algae polycultures that were isolated at the site have been shown to remove radionuclides from the water. Our team intends to take the next step in this research project by turning the experiments into a realistic solution for remediating the water on the disposal site and developing an economic model to understand other impacts of this solution for the region. The current land steward for Ambrosia Lake is hauling in water from outside of the area for dust control and other uses at the Ambrosia Lake disposal site. The water available at the site has levels of uranium and radium that exceed the EPA standards for drinking water but not the NRC standards for radioactive material. The contamination is in the water in the mines and aquifers in the area. This problem is not isolated to Ambrosia lake, and the solution could be applied to many affected areas. For example, uranium plumes have been documented in the San Andreas, Glorieta, and Alluvial Aquifers at the Bluewater, New Mexico, disposal site. Two hundred and fifty uranium mining sites have operated across New Mexico alone, and over a hundred of these sites have no record of reclamation. The Navajo Nation contains five hundred and twenty-one abandoned uranium mines. The polycultures from the Ambrosia Lake site at Santa Fe Community College have shown successful results in remediating water samples collected from the disposal site. They quickly and effectively remove radionuclides from the contaminated groundwater. The algae are naturally occurring in the site's ecosystem and probably have an exopolysaccharide coating that adsorbs the uranium, but critical questions have yet to be answered. Our team proposes to create a system of bioremediation of heavy metals that can be implemented to help restore lakes and lands to usable levels. Scientific evidence has proven some algae to be incredibly productive in phytoremediation. Discovering practical implementation of algae's abilities in phytoremediation can undo the harm caused to our waterways and land by mining and contamination across the globe.

Algae's superpower of remediation has been overlooked for far too long. Scientific evidence has proven it to be incredibly productive in phytoremediation. It can help repair generations of thoughtless damage caused to the earth and help maintain healthy, thriving ecosystems for future generations to come. Moreover, we can help address environmental justice depravity that has existed for too long.

Use of Machine Learning in Combination with Analytical Tools for Medical Screening

Luis E. Tafoya, University of New Mexico
Jameel K. Remtulla, University of New Mexico
Xander Augustson, University of New Mexico
Sarah Lavelle, University of New Mexico
Silas Bussmann, University of New Mexico
Robert M. Taylor, University of New Mexico
Virginia Severns, University of New Mexico
Andrea Howard, University of New Mexico
Justin T. Baca, University of New Mexico

Gas Chromatography (GC), Mass Spectrometry (MS), or a combination of both, are used in many industries to identify substances and contaminants in complex mixtures. Analyzing GC and MS signatures using Artificial Intelligence (AI) allows for improved signal classification and predicting the composition of unknown mixtures. This research focuses on training Machine Learning (ML) algorithms capable of predicting SARS-CoV-2 infection status through GC analysis of exhaled breath. 150 human subjects were recruited for feasibility studies in an effort to develop a rapid, point-of-care test for SARS-CoV-2 infection based on GC-ML signal analysis. 100 had positive PCR tests for SARS-CoV-2 and 50 had negative PCR tests for SARS-CoV-2. Of the subjects with positive

PCR results, 50 reported active symptoms while 50 were asymptomatic. After informed consent, subjects were provided 2.0 L of breath into a Tedlar collection bag secured via a mouthpiece connected to high-efficiency particulate air (HEPA) filters. Demographics were also collected. Retention time signatures were measured over a run time of <800 seconds. ML was performed using Scikit-learn libraries in python predicting positive, negative, and positive-asymptomatic individuals using GC intensity and retention time data. Models with metrics are generated and compared with the use of different algorithms as logistic regression. Based on preliminary data, we expect the breath analysis with GC, in combination with ML, will allow the prediction of SARS-CoV-2 cases with an accuracy >70%. This approach may be applicable to other screening strategies such as for diabetes or lung cancer.

Analyzing the Golden Age of Science Fiction: A Topic Modeling Approach*

Stephen Villanueva, Eastern New Mexico University
Edgar Eduardo Ceh Varela, Eastern New Mexico University

Science Fiction is a popular literature gender, yet it lacks a clear definition.

Even literature scholars have different interpretations of what Science Fiction is.

Several Machine Learning (ML) techniques exist to perform literature analysis. Similarly, Natural Language Processing (NLP) comprises a set of tools and algorithms to analyze, manipulate, generate, and understand human language. A typical analytical method for evaluating data is topic modeling. Topic modeling has been used as a research method in many different disciplines. These studies examine enormous volumes of text data to categorize document subjects objectively and examine the trends within each topic. In this study, we apply different ML and NLP techniques to a corpus of more than 2,000 Science Fiction novels from the so-called "Golden Age of Science Fiction" (i.e., the 1930s to 1960s). Our main goal is to find hidden topics in these documents to understand the Science Fiction genre better. In text analysis, we assume that documents from the same group must cover the same topic. Therefore, in this project, we use two different approaches to compare and contrast the topics extracted by each of them. In the first approach, we use a Latent Dirichlet Allocation (LDA) model for topic extraction. LDA is a generative probabilistic model considering the documents as random mixtures over latent topics. For the second, we use Doc2Vec. This model uses neural networks to represent each Science Fiction novel as an n-dimensional vector. Then, using a KNN algorithm, we use the generated vectors to find groups of similar documents. We obtained three important topics using LDA and five groups of documents using Doc2Vec with KNN. Some of the topics we found are related to "Horror and creatures," "Crimes," "Galaxy and space," and "Robots and Machines."

Similarly, we analyze these topics' evolution during the "Golden Age" period. This analysis shows how different topics increased or decreased in popularity over this period. Applying NLP techniques to a corpus of literature helped us discover latent topics that give us insight into the most important themes written during this important period for the Science Fiction genre.

Comparing the Growth of Xeric Trees in Soil vs. Aquaponics

Conner M. Wood, Santa Fe Community College R. Charlie Schultz, Santa Fe Community College
Stephen M. Gomez, Santa Fe Community College

In my experiment, I was comparing the growth of xeric trees that grow in the southwest in soil with traditional methods vs. aquaponics. I felt as though this application could provide a potential solution to fighting climate change, saving massive amounts of water, as well as contribute to habitat restoration efforts. I also examine factors that are common in the field of aquaponics such as PH, EC, N03 concentration, and PAR. Aquaponics has been around for thousands of years however, we as a society today still have a lot of research into what answers to current problems aquaponics can give us. Today there is not a lot of knowledge about growing Xeric trees in aquaponics so far, this is a unique study. This is a study that took place over one year from the summer of 2021 to the summer of 2022. There were three trials over the course of the study. I originally started with seeds that were in plug trays and were ready to be transplanted. There are more than several varieties of trees in the study. Since I was comparing the growth rate of the trees in various methods, all of the seeds per each specie were sown in the trays at the same time. This offers a fair and sets a starting point for all of the trees to be fairly compared in terms of height, width, and root mass comparison. I will demonstrate how I transplanted them into the aquaponics system as well as how I transitioned them out and eventually hardened them off without harming them. I will also discuss methods for a successful transplanting of the trees from water-based media into soil media including specific media used to help the transition. Overall after all three trials, I answered my original questions which were, Can the growth rate of trees is execrated with aquaponics vs. growing them in soil with traditional methods? As well as, Can larger root masses be cultivated while growing trees in aquaponics? The answer was Yes, for both questions and all of the species in the three trials. I believe that my findings of achieving faster growth rates of trees with aquaponics have the potential to help produce trees faster which has many applications. This can be used to speed up habitat restoration and possibly help curb the effects of climate change by being able to speed up tree production which was previously thought would take much longer to produce. Aquaculture can be a sustainable form of food in areas that greatly limit food production capabilities. Growing trees indigenous to the specific region could offer a great product for habitat restoration, combining these two concepts with aquaponics could be a potential game changer for restoration and sustainability.

GRADUATE STUDENT POSTER ABSTRACTS

Poster session abstracts are listed alphabetically by last name of registered presenter. An asterisk (*) indicates the poster received an outstanding graduate poster award at the 2022 New Mexico Research Symposium.

RapidPhage: a Microfluidic Based Phage Isolation Platform to Combat Antibiotic Resistant Pathogens

George A. Abernathy, Bioscience Division, Los Alamos National Laboratory
Anand Kumar, Bioscience Division, Los Alamos National Laboratory

In 2019 alone, antibiotic-resistant (AR) bacterial infections impacted over 3 million Americans, with a death toll of 48,000. By 2050, researchers predict that AR pathogens will globally cause over 300 million deaths, resulting in $100 trillion in direct and indirect losses. To avoid such a catastrophe, researchers have increasingly explored antibiotic alternatives. Bacteriophage, aka phage, is a virus that specifically infects bacteria and has historically been used to treat bacterial infections before antibiotics. However, conventional methods to isolate such phages are slow, tedious, and inconsistent with modern drug discovery processes. We are developing a high-throughput microfluidic platform to rapidly isolate effective phages in a fraction of the time, named "RapidPhage". Thus far, we have established conventional methods for the isolation of phages against Methicillin-Resistant Staphylococcus aureus (M.R.S.A.) from sewage samples. Irrespective of the sewage sample origin, conventional methods produced approximately 103 plaque-forming units against MRSA. Our developing RapidPhage platform will be used to compare against the phage titer efficacy of conventional methods. Since microfluidic co-capturing efficiency is very high, we expect that the RapidPhage platform will improve phage titer values by over 100-fold. Additionally, this developed platform should also recover more diverse MRSA-specific phages compared to established conventional methods. To exploit these viruses as therapeutic agents, we also plan to determine the pharmacodynamics of the most robust phages, such as reproducibility factors and infection turnover rates. In conclusion, the developed RapidPhage will be a swift, robust, and highly efficient method for phage isolation to combat the rise of AR.

Residue Quality Affects Litter Decomposition Dynamics and Carbon and Nitrogen Mineralization in a Semi-arid Agroecosystem

Amrit Dhoj Adhikari, Eastern New Mexico Uni.
Rajan Ghimire, New Mexico State University
Prabha Shrestha, Eastern New Mexico University
Zhiming Liu, Eastern New Mexico University
Pramod Acharya, New Mexico State University

Although cover cropping is promoted to improve soil health in arid and semi-arid environments, information on its residue decomposition and soil carbon (C) and nitrogen (N) dynamics is very limited. This study aimed to evaluate decomposition dynamics and nutrient release from different winter cover crops in a semi-arid agroecosystem. Twenty grams of triticale (Triticale hexaploid Lart), turnip (*Brassica rapa* subsp. *rapa L),* and pea (*Pisum sativum* subsp. *arvense L*) cover crops residues were kept in litter bags and deployed in the field. After cover crop termination forage sorghum (*Sorghum bicolor L*) was planted as a subsequent cash crop. The treatments were arranged in a randomized complete block design (RCBD) with four replications. Litter bags were removed every other week (a total of 12 destructive samplings) to evaluate the remaining dry matter and C and N contents. Also, a 10-week long incubation was set in laboratory conditions to understand the decomposition of triticale, turnip, pea, wheat, sorghum, and native grass residues by estimating potential mineralizable carbon (PMC) each week. After 98 days, the residue decomposition in litter bags was higher with pea (47 %) followed by turnip and triticale (both decomposed 40%). The residue analysis showed that legumes' lower C: N ratio favored faster decomposition. Residue types and quality both affected the decomposition rates. Cover crop termination and cash crop

planting times should be managed based on the quality of crop or cover crop residue for effective soil nutrient cycling and their uptake by subsequent crops.

Analysis of Asynchronous Communications on Microgrid State Estimation

Mohammad Afrazi, New Mexico Tech Kooktae Lee, New Mexico Tech

Average consensus algorithms in microgrids are becoming attractive concepts to meet the increasing energy demands. The goal of the average consensus is to calculate the average of initial values across dispersed nodes by sharing data via communications. The Average Consensus Theorem relies on the local information of agents to guarantee that important information is shared in a distributed way. It is known that synchronization is necessary to obtain an accurate average across the distributed nodes. Due to several technical issues, however, such as the requirement of a global clock and random packet drops/delays, which may occur during the implementation stage, the synchronous model is highly limited in reality. An asynchronous model, a synchronous one's counterpart, naturally takes into consideration the problems mentioned above and is thus more practical. The only downside of the asynchronous model is that it may result in inaccuracies in the consensus value which requires more thorough investigations. To this end, the performance of different network topologies has been investigated in this study for asynchronous systems with 16 different agents and random initial values. We analyzed three important performance factors: the convergence speed, error bound width with respect to the exact value, and the error variance.

NMHU-BioPACIFIC MIP PREM Website

Justice Ainsworth, New Mexico Highlands University

There are different ways to build a website, which require knowledge of multiple programming languages. To structure a website the most common languages used are HTML, JavaScript, Python, SQL, and PHP. The most common languages used for styling the website are CSS and Bootstrap. To determine which language was best for this website each programming language was researched and the four languages that were chosen were HTML, JavaScript, CSS, and Bootstrap 5. These languages allow for a smooth process for building the NMHU-BioPACIFIC MIP PREM website.

Examining Spatial Heterogeneity and Tourism Potential of Water Resources in New Mexico

Jason Banegas, New Mexico State University

Increasing outdoor tourism in New Mexico and the growing demand for water-based tourism amenities requires developing innovative strategies for resource allocation and service integration among many existing and potential recreation locations. Identifying tourism potential and optimizing resource allocation has been addressed in studies using a multitude of models, methods, and algorithms; however, variations between the influential factors of tourism locations are best analyzed with a Geographical Information System (GIS) based research model that can examine diverse spatial relationships in these environments. This study uses a Geographically Weighted Regression (GWR) method to illustrate that variables related to drivers of tourism are locally distinct. This tool is used to examine the spatial non-stationarity of relevant water-based tourism variables. Spatial discrepancies are shown to exist in the explanatory variables useful for developing tourism resources in rural and urban locations in the New Mexico study area. This GWR analysis can strengthen the decision process and allow policymakers to identify distinct influential factors and areas where resources for expanding water-based tourism should be focused.

Wall-Modeled Large Eddy Simulations of Turbulent Boundary Layer over a Flat-Plate for Aero-Optical Distortion Analysis*

Pedro Castillo Gomez, New Mexico State University
Andreas Gross, New Mexico State University
Nathan E. Miller, Sandia National Laboratories
Kyle P. Lynch, Sandia National Laboratories
Daniel R. Guildenbecher, Sandia National Laboratories

Density fluctuations due to compressible turbulent boundary layers cause significant aero-optical distortions that deteriorate the performance of sensors and directed energy applications. As part of a multi-institutional research team led by Sandia National Laboratories, compressible wall-modeled large-eddy simulations of turbulent flat plate boundary layer flows were carried out for Mach 3.5, 7.87, and 13.64. The Mach 3.5 conditions are the same as for a reference direct numerical simulation (Miller et al., 2021). For the Mach 7.87 case, the Sandia Hypersonic Wind Tunnel freestream values were matched. The Mach 13.64 case was modeled after the freestream conditions inside the Arnold Engineering Development Complex Hypervelocity Tunnel 9. Despite the relatively low grid resolution compared to reference direct numerical simulations, the mean velocity, temperature, and Reynolds-stress profiles obtained from the wall-modeled simulations are in good agreement with available reference data. In addition, the normalized root-mean-square optical path difference obtained from the present simulations compares well with reference simulations and experiments. This suggests that wall-modeled large-eddy simulations can provide accurate and efficient predictions of the aero-optical path distortions at a drastically lower computational cost than direct numerical simulations. A semi-analytical relationship by Notre Dame University that was developed for low-Mach applications underpredicts the path difference for the higher Mach numbers. This mismatch provided the motivation for evaluating the underlying assumptions of the theoretical model for hypersonic boundary layer flows. The present results contribute to the development of updated models for the path difference that will assist the design of high-speed aero-optical applications.

Examining the Rights-of-Way for Navajo Allotment Lands in Connection to the Navajo-Gallup Water Supply Project

Bernadette Fontenelle, University of New Mexico

This research examines the Right-of-Way (ROW) process for the Navajo Nation, USA, allotment lands. Today, there are 573 Indian nations. Each Indian nation has its history relating to Indian allotment lands. During the 1880s, allotment lands were created through a federal Indian policy. The policy approved various Indian reservation lands into allotment lands. Today, these allotment lands have become a ROW issue. This research examines a real example of the ROW issues impacting the Navajo-Gallup Water Supply Project (Project). Land access for allotment lands is questionable. Water access for the Project was secured and supplied through the Navajo Nation San Juan River Water Rights Settlement. The Project pipeline alignment will cross six types of land. Each type of land has its ROW process. Several methods were applied to determine the ROW process for the Navajo allotment lands. A literature review of existing federal Indian policy finds that the Navajo Nation does not have authority over the allotment lands. Two federal sister agencies, the U.S. Bureau of Indian Affairs and the Bureau of Reclamation are challenged by the administrative ROW process. One has authority over these allotment lands, the other must initiate the ROW application of the Project. The Bureau of Indian Affairs will approve or disapprove of the ROW easement. After examination, the research identifies areas of improvement for the current ROW process and provides recommendations, starting with a better framework to implement into the ROW process for Navajo allotment lands. Water is Life!

Remote Activation Strategy for Increased Reactivity and Stability in Cycloalkynes: Incorporation of an Endocyclic Sulfate within a Dibenzocyclononyne Scaffold

Michael Holzmann, University of New Mexico
Pavel Yamanushkin, University of New Mexico
Namrata Khanal, University of New Mexico
Brian Gold, University of New Mexico

Cycloalkynes and their utilization in cycloaddition reactions enable modular strategies spanning the molecular sciences. Strain-imparted by deviation from linearity-enables sufficient alkyne reactivity without the need for a catalyst (e.g., copper), although the design and synthesis of stable reagents with suitable reactivity remain an ongoing challenge. We report the incorporation of an endocyclic sulfate within a dibenzocyclononyne scaffold to generate a cyclononyne displaying remarkable reactivity and stability. Through computational analyses, we revealed that the endocyclic sulfate group shares nearly half of the total strain energy, providing an activation strategy that reduces alkyne bending. Rehybridization of alkyne carbons in the formation of the heterocyclic product relieves strain both at the reactive site and in the transannular sulfate group. This mode of remote activation enables rapid reactivity while minimizing distortion and strain at the reactive site (the alkyne). The result: a design strategy for a new class of cycloalkynes with increased stability and reactivity.

Late-Stage SuFEx Diversification of HIV Protease Inhibitor

Bipin Khanal, University of New Mexico

Click chemistry- a nature-inspired approach for rapid synthesis of useful compounds , since its dawn has dominated organic synthesis. Modern development of near-perfect reactions enables modular manipulation of building blocks. Our group has recently reported a mild route to SuFExable pyrazoles. We are now poised to rapidly explore unknown chemical space via controllable diversification of these densely functionalized heterocycles. Recent progress in late-stage modification of FDA-approved pharmaceuticals, especially for the treatment of HIV/AIDS, to unlock new properties will be presented.

A Deep Spatial-Temporal-Channel Attention Network for sEMG-based Hand Gesture Recognition

Qingqing Li, New Mexico Tech
Zhirui Luo, New Mexico Tech
Ruobin Qi, New Mexico Tech
Zheng Jun, New Mexico Tech

Hand gesture recognition has become more attractive in the field of human-computer interaction (HCI) due to the rich information and the natural and intuitive communication mode provided by hand gestures in a human dialog. Many computer vision-based approaches have been applied to hand gesture recognition but the strict requirements of background invariance and lighting insensitivity limit their accuracy and robustness. Compared with vision-based approaches, methods using surface electromyography (sEMG) signals are more robust and have better usability. In this work, we proposed a new deep learning model for hand gesture recognition using multi-channel sEMG signals, which consists of a ResNet50 network and a novel triple attention module. The ResNet50 network serves as an implicit automated feature extractor to retrieve features from the sEMG signals. The triple attention module combines spatial attention, temporal attention, and channel attention to signify important features for hand gesture recognition. The performance of the proposed model is evaluated with two popular datasets, Ninapro-DB1 and Ninapro-DB2. The results demonstrate that the proposed model significantly outperforms the existing state-of-the-art methods.

Deep Feature Extraction for Semi-supervised Electricity Theft Detection in AMI with Observer Meters*

Zhirui Luo, New Mexico Tech
Jun Zheng, New Mexico Tech
Ruobin Qi, New Mexico Tech
Qingqing Li, New Mexico Tech

Electricity thieves can launch cyberattacks against advanced metering infrastructure (AMI) to reduce their electricity bills which damage the integrity of the smart grid and cause huge economic losses to utilities. Traditional supervised learning-based methods rely on representative fraudulent data to train good detection models. However, it's hard to collect representative electricity theft data in the real world and a trained supervised model may not be able to identify new false data injection attacks. On the other hand, semi-supervised learning only employs normal data to train detection models which makes it suitable for detecting unknown attacks. To this end, we propose a semi-supervised deep learning method for electricity theft detection in AMI. The method uses ratio profile data calculated from observer meter data as the input. Deep features are extracted from the ratio profile data using continuous wavelet transform (CWT), deep representation learning, and principal component analysis (PCA). The extracted features are then fed into an autoencoder network for semi-supervised anomaly detection. The performance of the proposed method is evaluated with real-world smart meter datasets to demonstrate its validity for electricity theft detection.

A Deep Attention Network for Non-intrusive Building Occupancy Detection Using Smart Meters

Zhirui Luo, New Mexico Tech
Ruobin Qi, New Mexico Tech
Qingqing Li, New Mexico Tech
Jun Zheng, New Mexico Tech
Sihua Shao, New Mexico Tech

Occupancy information is gaining more attention for efficient energy management in the building sector because it provides significant potential for energy reduction. The massive smart meter data collected by the advanced metering infrastructure (AMI) facilitates inferring occupancy status in a non-intrusive way, but the non-linear relationship between power consumption data and building occupancy status hampers the effectiveness of detection techniques. Current machine learning as well as deep learning models are restricted in manual feature extraction and hard to scale up when facing massive smart meter data. In this work, we propose a deep attention network called ABODE-Net which utilizes a novel parallel attention block to infer real-time building occupancy status from raw smart meter data and corresponding time information in an end-to-end manner. We compared our method with a set of state-of-the-art shallow machine learning and deep learning models using two smart meter datasets widely used for building occupancy detection. The results show that ABODE-Net has a significantly better performance in all experimental cases. This study proves the validity of the proposed ABODE-Net as a solution for non-intrusive building occupancy detection using smart meter data.

Hydrogels as Support Medium for the Culture of Cyanobacteria

Brannette Mejia, Los Alamos National Laboratory

Cyanobacteria are prokaryotic cells that have the ability to perform photosynthesis. Spirulina (*Arthrospira platensis*) is a common cyanobacterium that can be a food source for animals and humans, as well as possess characteristics such as biosorption and bioaccumulation capacity. Hydrogels are hydrophilic polymers that do not dissolve in water, instead, they tend to absorb water, by possessing hydrophilic functional groups. They can be formed by utilizing natural polymers and adding crosslinkers when together creating a solution with high viscosity and consistency. Hydrogels have a vast variety of applications, mostly in the biomedical field such as contact

lens production, wound dressings, 3D printing, and agarose. In this project, we study the use of poly(vinyl) alcohol as a polymer matrix in combination with borax as a compatible crosslinker. In this work, we first explore various parameters to produce self-standing hydrogels that can be cast, or 3D printed into various shapes. We also investigate the viability of *A. platensis* in a hydrogel medium. A wide variety of techniques are employed to determine the microscopic features as well as the mechanical and chemical properties of the hydrogels. We show that after forming six different hydrogels with a variety of concentrations, the level of PVA and borax that is added to the solution has a direct effect on the hydrogel behavior. Our results identified that the 5% PVA and 4% Borax hydrogel showed the utmost characteristics and properties that matched the expectations of hydrogel behavior. LA-UR-22-30293

Isoreticular Synthesis of HKUST-Type Metal-Organic Framework Using Imine Functionality

Edgar Novik, New Mexico Highlands University Alisha Gogia, New Mexico Highlands University
Tatiana Timofeeva, New Mexico Highlands University

Metal-organic frameworks (MOFs), which is a class of hybrid organic and inorganic porous materials with 2D and 3D networks, have emerged as important materials which can be used for different purposes such as gas separation and storage, in the pharmaceutical industry as drug delivery compounds, and in environmental applications such as photoluminescent sensors. One of the main advantages which MOFs offers is the easy tunability of functionality in the resultant MOF structures by slight changes in the organic ligand. By choosing the right geometry of ligands and metal nodes it is possible to construct the target MOFs with high porosity and pore volume. One highly advantageous functionality which can be incorporated into the MOF assembly is the imine group. The imine group tends to have extensive H-bonding interactions with the guest/analyte molecules. This allows the sensing and sorption of various important molecules using MOFs. In addition, the imine functional group provides the advantage of post-synthetic modification to convert imine into amine for different applications such as water and gas separation, drug delivery, and luminescent sensing. In the current work, we have designed and synthesized an imine-based tricarboxylic ligand, 4,4,(((1E,1,E,1E)-benzene))tris(azaneylylidene)) tribenzoicasid (H3bttta), to incorporate in the HKUST-type MOF assembly using Cu(II) metal center as the node. HKUST was chosen as the target structure and isoreticular synthesis was done using the synthesized H3bttta for it offers large hexagonal pores which can encapsulate a bunch of analyte molecules. Furthermore, in addition to the functional benefits, we have also focused on the structural variations brought in by the flexibility of the H3bttta ligand to the resultant MOF structure. The study showcases the standardizations involved in the synthesis of the organic ligand and the MOF, and the characterization of the bulk powders.

Synthesis of Carbonyl-Functionalized p-Phenylenes for Probing Electron Delocalization*

Chimezie Onukwuli, Eastern New Mexico University
Juchao Yan, Eastern New Mexico University

Organic pi-conjugated p-phenylenes have attracted enormous current interest, owing to their optoelectronic applications including organic solar cells (OSCs), light-emitting diodes, and organic lasers. The design of OSCs with an improved power conversion efficiency demands a proper understanding of electron delocalization and its influence on the energetics and dynamics of electrons. In literature, nitrile-functionalized ladder-type oligo (p-phenylenes) have been synthesized and used to probe electron delocalization by time-resolved infrared spectroscopy coupled with pulse radiolysis. While these rigid coplanar compounds are infrared-responsive and thermally stable, nitrile groups are not commonly incorporated into solar-related molecules. Herein, we propose

the incorporation of a carbonyl group infrared reporter into ladder-type oligo (p-phenylene) architecture to further probe electron delocalization. The proposed compounds are being synthesized via palladium reductive carbonylation using N-formyl saccharin as a source of carbonyl. The target compounds are affordable and easy to manufacture, which is critical for the commercial production of donor materials for OSCs. In this poster, we focus on our molecular design, syntheses, and characterizations.

Novel Manganese-Halide 1-D and 2-D Frameworks and their Emission Properties

Michael O. Ozide, New Mexico Highlands Uni. Rayen Gonzales, New Mexico Highlands Uni.
Tatiana Timofeeva, New Mexico Highlands Uni. R. Castaneda, New Mexico Highlands Uni.

In the past years, metal-halide perovskites (MHPs) attracted much attention as they are exceptional candidates for novel photovoltaic materials. MHPs are not limited to photovoltaics and our goal at NMHU is to study manganese-halide perovskites (MnHPs) for their potential magnetic, ferroelectric, and light emission properties. These properties imply that these compounds can be used for LEDs and detectors. Manganese-halide compounds normally crystallize with isolated tetrahedral MnX4 units, however, they can crystallize in 1-D chains, or 2-D layers when organic ligands are present. Our group synthesized five new MnHPs, four of them with 1-D chains, and one of them with 2-D layers with 3-aminopyridine as the organic ligand. Our results show these MHPs can be modified by the amount of water present, and by the solvent used, leading to 1-D chains with different angles between manganese octahedrons.

Characterization of a Modified Carbon Fixation Mechanism in Synechocystis sp PCC 6803

Sara Pacheco, Los Alamos National Laboratory

Cyanobacteria are a promising platform for the production of carbon-neutral biofuels, plastics, and other commodities. However, these cyanobacterial-based technologies must improve in biomass yield and biochemical composition to compete with the economics of petroleum-based products. We seek to increase biomass production by improving the CO_2 fixation capacity within the cyanobacterial cell. Towards this goal, we introduced a copy of the ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) type IC from the soil bacterium Cupriavidus necator H16 into the cyanobacterium Synechocystis sp. PCC 6803, which naturally expresses Rubisco-type IB within a carboxysome. We hypothesize that differences in Rubisco, IC cellular location, catalytic turnover, and substrate specificity in the transgenic PCC 6803 strains will complement its native CO_2 concentrating mechanism, resulting in improved photosynthetic efficiency and biomass accumulation. We will test this through growth (spectrophotometrically at 730 nm), photosystem II fluorescence, oxygen evolution, and flow cytometry (dye-assisted evaluation of polyhydroxybutyrate accumulation) in cultures under various CO_2 concentrations. Our success will improve the carbon fixation capacity of the cell, increase biomass production, and increase cyanobacteria's economic value.

iCAD: information-Centric network Architecture for DDoS Protection in the Smart Grid

Sharad Shrestha, New Mexico State University George Torres, New Mexico State University

With the proliferation of differently-abled and heterogeneous devices in the smart grid, Denial of Service (DoS) is becoming an even more potent attack vector than before. This poster demonstrates the ease with which an adversary can orchestrate DoS and distributed DoS (DDoS) attacks on the grid. We propose iCAD, an information-centric architecture that extends the iCAAP architecture we previously proposed, complete with mitigation strategies built for DoS/DDoS resilience. We present our architecture and demonstrate the architecture and the

mitigation technique's effectiveness in mitigating DoS/DDoS attacks in the face of significant attack load from the distributed agents.

High-Throughput Synthesis and Modification of Peptides

Samuel Takyi, University of New Mexico
Brian Gold, University of New Mexico
Pavel Yamanushkin, University of New Mexico
Mark Aldren M. Feliciano, Uni. of New Mexico

Expanding chemical space beyond the genetically encoded amino acids endows biomolecules with unique properties. As nature harnesses chemical (post-translational) modifications, we seek to incorporate synthetic modifications into peptide scaffolds. We are currently utilizing high-throughput techniques (i.e., phage display) to fabricate modified peptides to understand the role of peptide structure on employed techniques and synthesize biomaterials with utility in various applications.

POSTDOCTORAL POSTER ABSTRACTS

Using Machine Learning for Magnetic Hysteresis of Metal-Organic Frameworks

Marine S. Foucher, New Mexico Highlands Uni.
Patrik Boloz, New Mexico Highlands Uni.
Saige E. Martinez, New Mexico Highlands Uni.
Mike S. Petronis, New Mexico Highlands Uni.
Gil R. Gallegos, New Mexico Highlands Uni.
Tatiana Timofeeva, New Mexico Highlands Uni.

Magnetic hysteresis data provide critical information for Material Science applications to characterize the magnetic behavior of synthesized metal-organic frameworks (MOFs), organic metal halide materials, and natural materials. The analysis of magnetic hysteresis yields a wide range of information on the magnetic domain state, and magnetic grain size (e.g., single domain vs. vortex state vs. multidomain in ferromagnetic materials), and quantifies the elemental chemistry of materials. The interpretation of the magnetic hysteresis data is typically done by analyzing distribution curves individually (e.g., their shape, the saturation levels) and requires a good understanding of material magnetism. However, projects using magnetic hysteresis measurements typically involve tens to hundreds of specimens whose interpretation is time intensive. At New Mexico Highlands University, two of the PREM Research Thrusts are dedicated to (1) characterizing the magnetic properties and (2) integrating machine learning and material modeling for Material Sciences. Machine learning is another great tool already used in material and geophysical sciences. Unfortunately, only a handful of studies have attempted to integrate machine learning into magnetic material characterization. For this research, we initiated a project dedicated to facilitating the data reduction and interpretation of magnetic hysteresis data using image processing and Euclidean distance algorithms. We will show 1) the first stage of the project involving the construction of a magnetic hysteresis catalog and algorithms development and 2) our ongoing plans for the deployment of this machine learning program to evaluate hysteresis data and the interpretation of the results.

About the New Mexico Academy of Science

The New Mexico Academy of Science (NMAS) was founded in 1902 to foster scientific research and scientific cooperation, to increase public awareness of the role of science in human progress and welfare, and to promote science education in New Mexico.

Membership in NMAS is open to anyone interested in science, formal and informal science education, or the other goals and programs of NMAS. Individuals engaged in scientific research or teaching at all levels are particularly encouraged to become members. Applications for membership as well as more information about NMAS and its programs can be found at www.nmas.org.

CONTACT INFORMATION

The New Mexico Academy of Science
c/o The New Mexico Museum of Natural History and Science
1801 Mountain Road NW
Albuquerque, New Mexico 87104
nmas@nmas.org www.nmas.org

OFFICERS AND EXECUTIVE BOARD 2022

Officers

President: Dr. Anton Sumali
President-Elect: Dr. Babu Chalamala
Vice-President: Dr. Diane Peebles
Treasurer: Dr. Deb Thrall
Secretary: Malva Knoll
Past President: Dr. Gretchen Gürtler

Directors

Jayne C. Aubele: NMAS Education Awards, NMMNHS Museum Liaison, Lectures Committee
Lynn Brandvold: New Mexico Junior Academy of Science; SWARM, AAAS, NAAS Delegate
Dr. Diane Peebles: Public Relations/Communications/Outreach
Dr. Cecilia Hernandez: Liaison for NMSU
Christy Kranek: Liaison for NM Public Education Department
Dr. Rod Rock: Liaison for NM Highlands University
Margaret Showalter: Director at Large, National Youth Science Camp
Dr. Zhiming Liu: Director at Large
Vladislav Sevostianov: Editor-in-Chief, New Mexico Journal of Science
Dr. Anton Sumali: Community Relations
Ellen Loehman & Malva Knoll: Newsletter

Directors Emeritus

David Duggan
Dr. David Hsi
Harry Pomeroy (NMAS Fellow)
Mona Pomeroy
Dr. Maureen Romine


NMAS
NEW MEXICO
ACADEMY OF SCIENCE