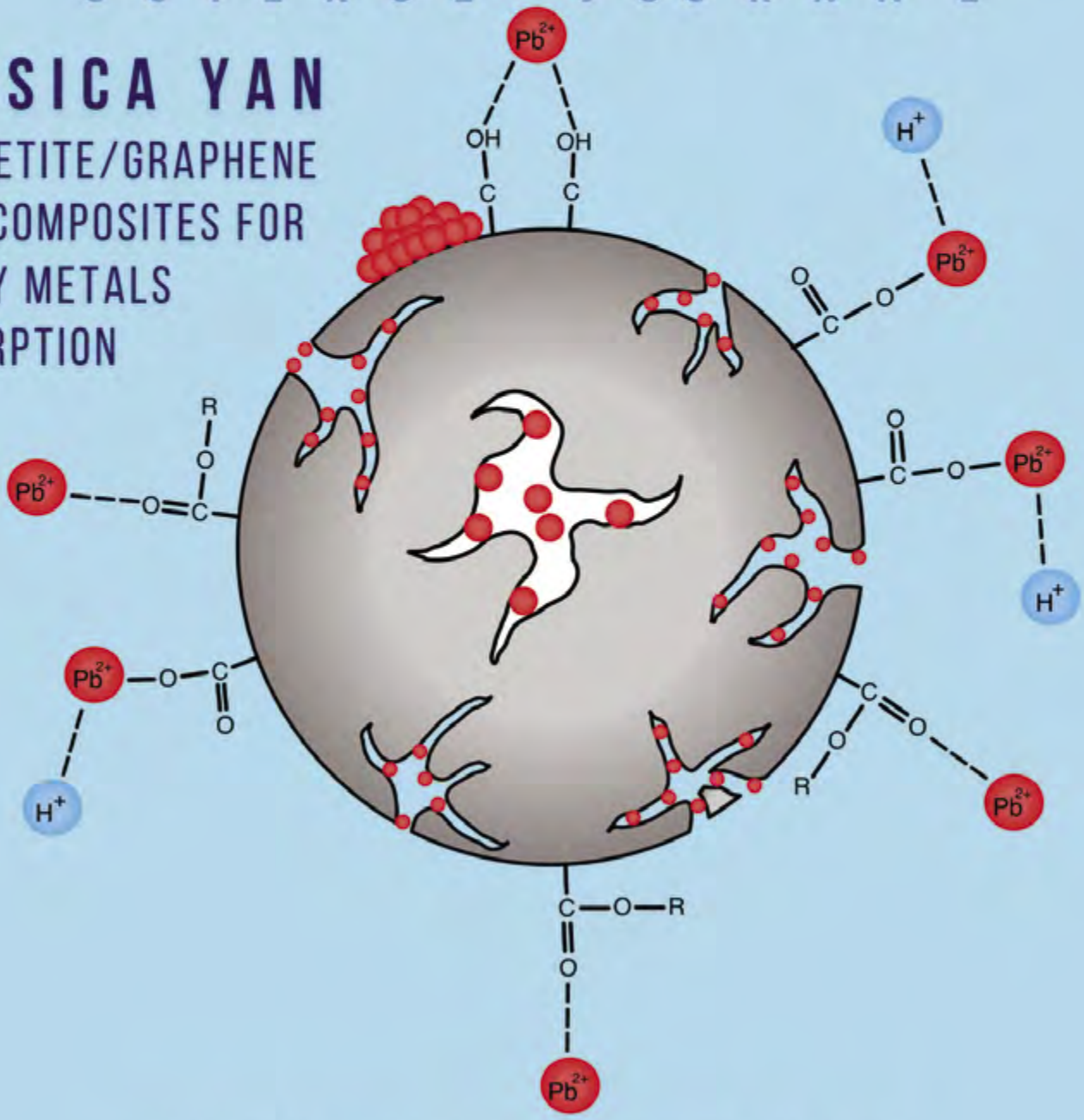
ISSN 2692-191X

MSMS

SCIENCE JOURNAL

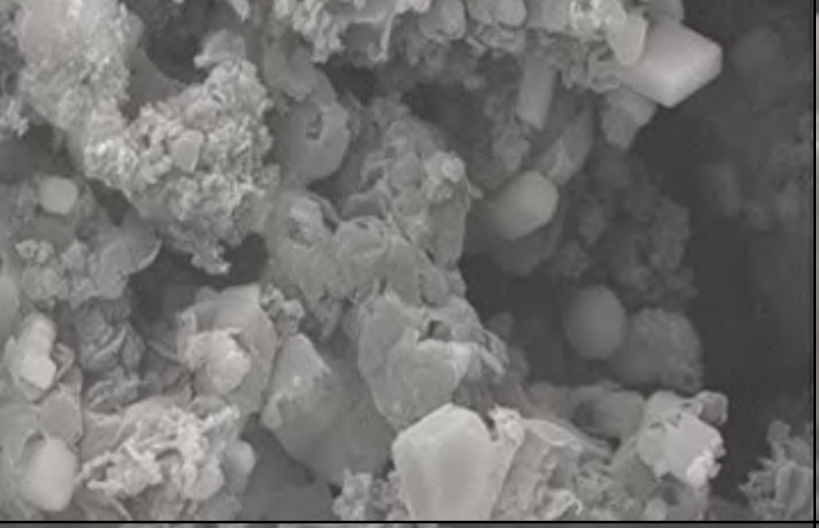
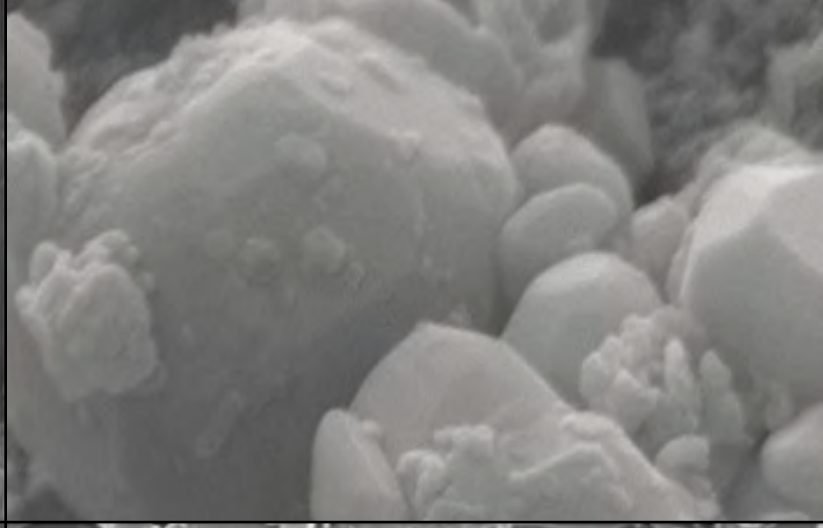
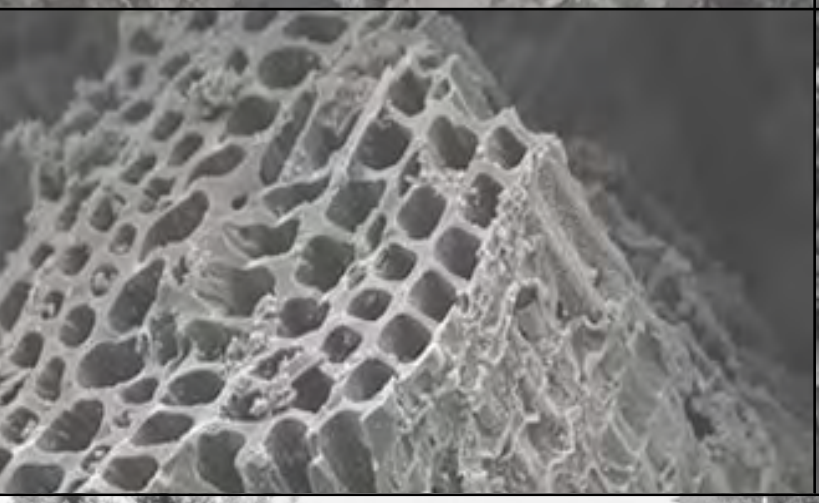
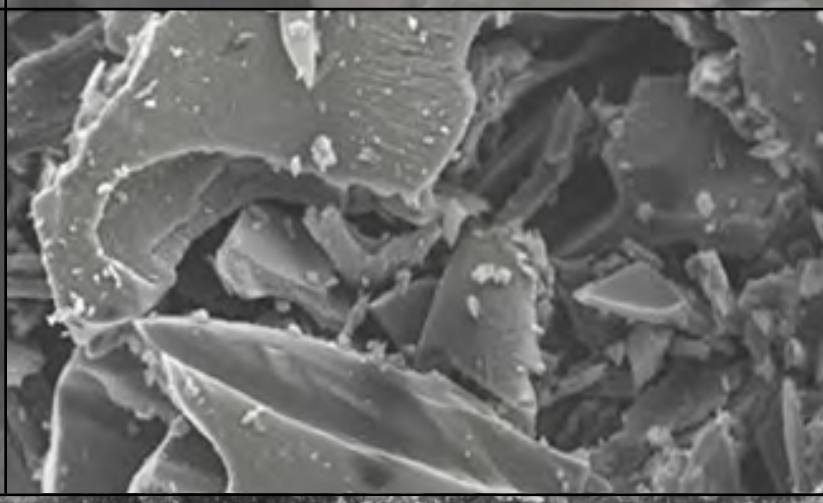
JESSICA YAN

MAGNETITE/GRAPHENE NANOCOMPOSITES FOR HEAVY METALS ADSORPTION



WAVES OF SCIENCE

MAY 2022 | VOL. 3 | ISSUE 1 | EST. 2020


Cover Diagram: Jessica Yan (2022)

Faculty Advisor: Dr. Tina Gibson

Cover Design: Jessica Yan

Photography: Caleb Jenkins (2022)

Editor: Jessica Yan

Co-Editors: Madison Echols and Dia Kher

CONTENTS

Magnetite/Graphene Nanocomposites for Heavy Metals Adsorption

Jessica Yan

Abstract

It is challenging to produce economical magnetic graphene-based adsorbents on an industrial scale for heavy metal ions removal. Here, magnetite/graphene nanocomposite embedded in activated carbon matrix (magnetite/G-AC) was synthesized via in situ catalytic graphitization of iron-impregnated biochar to obtain graphene encapsulated iron nanoparticles (GEINs) embedded in biochar (BC) matrix and followed by steam activation of GEINs-BC. Steam activation aimed to upgrade biochar to activated carbon with oxygen functional groups, crack encapsulated graphene shell to graphene nanosheets, and obtain magnetic Fe_3O_4 by oxidation of iron, thereby improving the adsorption capacity of magnetite/G-AC-800 (153.2 mg/g) four times higher than that of GEINs-BC. The parameters on the adsorption capacity were investigated using Pb(II) ions as a typical pollutant as a function of solution pH (3–7), contact time (5–300 min), initial Pb(II) concentration (50–400 mg/L), and adsorbent dosage (0.05–0.25 g). This research developed a novel, low-cost magnetic adsorbent with the advantage of simple large-scale production and excellent adsorption capacity per unit cost for remediating wastewater.

Introduction

Inorganic pollutant heavy metals observed in wastewater effluents are highly toxic to human health, aquatic life, and the environment. Current methods for wastewater treatment include adsorption, chemical precipitation, ion exchange, and others. Furthermore, there are various adsorption materials, including biochar, activated carbon, and metal sorbents. In pursuance of reduced cost, the use of inexpensive carbonaceous materials as adsorbents has gained attention. Biochar derived from the waste biomass has emerged as a promising adsorbent due to its economic efficiency, waste recycling, abundant availability, and ease of production (Li et al., 2017a; Xiang et al., 2020). Alternatively, the catalytic graphitization of biochar can directly synthesize biochar-based graphene encapsulated nanoparticles (Tang et al., 2015; Tokoro et al., 2004; Yan et al., 2018; Zhang et al., 2021; Zhao et al., 2012), which is a simple and inexpensive method. Therefore, this study focuses on the adsorption of heavy metal ions using a novel magnetic composite adsorbent of Fe3O4/graphene nanocomposites in biochar-based activated carbon.

Methodology

Preparation of biochar from biomass

Biochar used in this work was obtained by fast pyrolysis of pine wood. The biochar was first boiled in 0.1 M HNO_3 solution overnight to remove any soluble inorganic ions and bio-oil residue, subsequently washed with hot deionized water three times, and dried in an oven at 105°C for 24 h.

Preparation of graphene encapsulated iron nanoparticles (GEINs) in biochar (BC)

GEINs/BC was synthesized through a two-step route including the impregnation of iron on biochar and in situ catalytic graphitization of iron-impregnated biochar. Fourty one grams of iron nitrate hexahydrate were dissolved in 30 mL of deionized water, followed by the impregnation of 50 g biochar in the iron nitrate solution. The iron-impregnated biochar powder underwent catalytic graphitization at 900°C for 60 min under an argon atmosphere with a flow rate of 100 mL/min. After cooling to ambient temperature, few-layered graphene encapsulated iron nanoparticles (GEINs) in biochar (BC) were obtained and labeled as GEINs-BC.

Formation of Fe_3O_4/graphene nanocomposites in biochar-based activated carbon (Fe_3O_4/G-AC)

Fe_3O_4/G-AC was synthesized through steam activation of GEINs-BC at temperatures 500-800°C. Ten grams of GEINs-BC were packed in a tubular reactor and purged for 30 min with argon flowing at 100 mL/min. Then, the argon flow was bubbled through water at 80°C before entering the reactor, and the reactor was heated to the desired temperature (500, 600, 700, and 800°C, respectively) at 5°C/min and held the desired temperature for 60 min. The Fe_3O_4/graphene in biochar-based activated carbon (Fe_3O_4/G-AC) was obtained and designated Fe_3O_4/G-AC-500, Fe_3O_4/G-AC-600, Fe_3O_4/G-AC-700, and Fe_3O_4/G-AC-800, respectively.

Results

Characterization of GEINs-BC and Fe_3O_4/G-AC

The surface areas, total pore volumes, and average pore diameters of the GEINs-BC and Fe_3O_4/G-AC are summarized in Table 1.

Table 1 Characteristics of the GEINs-BC and Fe_3O_4/G-AC samples.

Samples	BET Surface Area (m^2/g)	Total pore volume (cm^3/g)	Average pore diameter (nm)
GEINs-BC	201.5	0.166	2.1
Fe_3O_4/G-AC-500	236.4	0.173	2.2
Fe_3O_4/G-AC-600	305.7	0.215	2.6
Fe_3O_4/G-AC-700	393.2	0.236	2.9
Fe_3O_4/G-AC-800	485.8	0.302	3.3

Adsorption of Pb(II) ions from aqueous solution

The adsorption capacity parameters were investigated using Pb(II) ions as a typical pollutant. The parameters affecting the adsorption, such as temperature, the initial Pb(II) concentration, adsorbent dosage, contact time and pH solution, were investigated. The results demonstrate that the adsorption of Pb(II) is pH-dependent, time-dependent, dosage-dependent, and concentration-dependent.

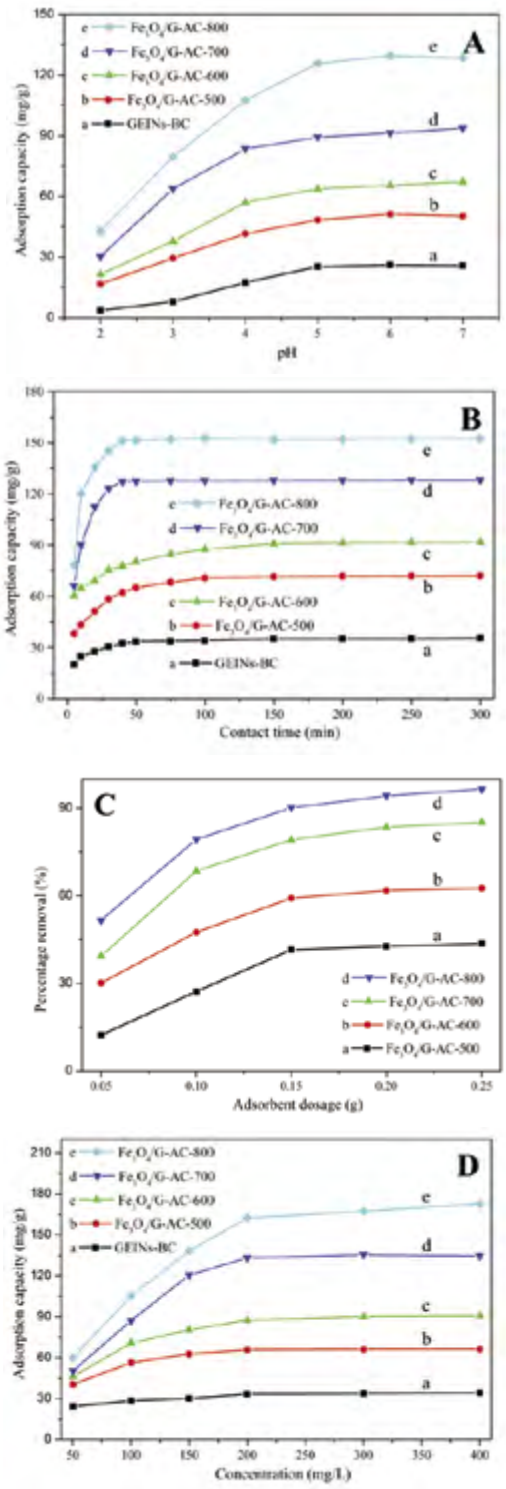


Figure 1 (A) Effect of pH (Conditions: C_0 = 250 mg/L, T = 25°C; adsorbent dose = 0.10 g/100 mL, contact time = 150 min). (B) Effect of contact time (Conditions: C_0 = 250 mg/L, pH = 5, T = 25°C; adsorbent dose = 0.10 g/100 mL). (C) Effect of adsorbent dosage (Conditions: T = 25 °C; C_0 = 200 mg/L, pH = 5, v = 100 mL; contact time = 150 min). (D) Effect of initial concentration (Conditions: C_0 = 50-400 mg/L, pH = 5, T = 25°C; adsorbent dose = 0.10 g/100 mL, contact time = 150 min).

Discussion

This study focuses on the economic feasibility for the synthesis of magnetic graphene adsorbent on an industrial scale. All precursors are cheap and environmentally friendly. A Fe_3O_4-graphene-activated carbon composite prepared through stream activation displayed an increase in cracked graphene nanoplates and oxygen functional groups in activated carbon, which enhanced the adsorption capacity for heavy metals. The adsorption process is pH-dependent, and the maximum adsorption capacity of Pb(II) was achieved at pH 5. The extent of Pb(II) adsorption increases upon increasing initial concentration and contact time. For the industrial-scale separation of heavy metal ions from wastewater, it is a promising effective adsorbent composite for preparation on a large scale, low price, and increased adsorption capability.

Future Implications

Globally, around one in three people face problems in accessing clean, potable water, largely due to heavy metal contamination. Such contamination further presents a risk to humans in the form of health-related diseases while also posing a problem to the sustainability of the environment. Thus, it becomes imperative to meet the Sustainable Development Goals (SDG) established by the United Nations, which include the conservation of clean water (SDG 6) and marine environments (SDG 14). Presenting novel approaches like magnetite/graphene nanocomposites in an activated carbon matrix adsorbent to tackling water contamination can achieve such goals while saving millions of lives, ensuring increased availability to clean water resources, and sustaining the marine ecosystem in the process.

References

Li, L.-L.; Feng, X.-Q.; Han, R.-P.; Zang, S.-Q.; Yang, G., Cr(VI) removal via anion exchange on a silver-triazolate MOF. *Journal of Hazardous Materials* 2017**,** *321*, 622-628.

Son, E.-B.; Poo, K.-M.; Chang, J.-S.; Chae, K.-J., Heavy metal removal from aqueous solutions using engineered magnetic biochars derived from waste marine macro-algal biomass. *Science of The Total Environment* **2018,** *615*, 161-168.

Tokoro, H.; Fujii, S.; Oku, T., Iron nanoparticles coated with graphite nanolayers and carbon nanotubes. *Diamond and Related Materials* 2004**,** *13* (4), 1270-1273.

Xiang, W.; Zhang, X.; Chen, J.; Zou, W.; He, F.; Hu, X.; Tsang, D. C. W.; Ok, Y. S.; Gao, B., Biochar technology in wastewater treatment: A critical review. *Chemosphere* 2020**,** *252*, 126539.

Yan, Q.; Li, J.; Zhang, X.; Zhang, J.; Cai, Z., In situ formation of graphene-encapsulated iron nanoparticles in carbon frames through catalytic graphitization of kraft lignin. **2018,** *8*, 1847980418818955.

Zhang, X.; Navarathna, C. M.; Leng, W.; Karunaratne, T.; Thirumalai, R. V. K. G.; Kim, Y.; Pittman, C. U.; Mlsna, T.; Cai, Z.; Zhang, J., Lignin-based few-layered graphene-encapsulated iron nanoparticles for water remediation. *Chemical Engineering Journal* **2021,** *417*, 129199.

Zhao, G.; Wen, T.; Yang, X.; Yang, S.; Liao, J.; Hu, J.; Shao, D.; Wang, X., Preconcentration of U(vi) ions on few-layered graphene oxide nanosheets from aqueous solutions. *Dalton Transactions* **2012,** *41* (20), 6182-6188.

Assessing Racial Determinants of Wellbeing

Raeed Kabir

Abstract

This paper reveals numerous psychological mechanisms and correlations that are relevant in the sphere of race and wellbeing, and thus relevant to all people because of the inevitable effects that race has on one's life. I show that the valuation of race is asymmetric across racial groups and that this immediately has repercussions in the real world. I propose a theoretical market where race is an item of transaction, under which this asymmetry can be analyzed systematically. Within this framework, I rename this nuanced form of information asymmetry to be experiential asymmetry and discuss a new type of market failure, one where the cost to wellbeing is not immediately financial but psychological. Finally, I note inconsistencies in my dataset with the long-standing idea of loss aversion. I suggest that there exists something inherently unique to being a person of color that correlates strongly with valuing a gain more than an equivalent loss.

I. Introduction

In this paper, I measure the changes in wellbeing an individual would report if their race were to instantaneously change. Using measures like increase and decrease in salary, I price the equivalent value of a race change to some monetary value. This paper finds that Black individuals value the gain of being white more than the loss white people would face if they were to become Black.

I, then, draw an analog between the differences in the valuation of race seen in this dataset and the information asymmetry present in the used car market (Akerlof, 1970). Akerlof presented information asymmetry as a source of market failure: without complete information, a market is unable to maximize the joint utility of the buyer and the seller. I propose a theoretical market where individuals are able to purchase race, and those who have experienced, first-hand, the costs of being a person of color seem to value whiteness substantially more than white individuals seem to value their race. The consequences of this information asymmetry, or rather experiential asymmetry, can be substantial when those in power are the decision-makers for the wellbeing of minority groups.

I. <u>Benchmarking Questions</u>

Not all aspects of the project or survey are elucidated here. The first relevant question asked for the race of the participant.

All individuals who identified as White were asked the following question:

1. Imagine that, at this moment, your race were to change to African American/Black. Consider how your life might change and rate it on a scale from -10 to 10, where 0 is your current comfort, 10 is the most comfortable life you could imagine, and -10 is the most uncomfortable life you could imagine.

Respondents were given a labeled scale at -10 (Most Uncomfortable), 0 (Status Quo), and 10 (Most Comfortable) to record their responses.

All individuals who identified as a person of color (or were non-white) were asked the following question:

1. Imagine that, at this moment, your race were to change to Caucasian/White. Consider how your life might change and rate it on a scale from -10 to 10, where 0 is your current comfort, 10 is the most comfortable life you could imagine, and -10 is the most uncomfortable life you could imagine.*

All respondents then answered a series of benchmarking questions, of which there were two kinds. These benchmarking questions measured the wellbeing change one would feel for other kinds of life events. One type of question was inherently positive and structured like so:

2. Imagine that [event X were to occur]. Consider how your life might change and rate it on a scale from 0 to 10, where 0 is your current comfort, and 10 is the most comfortable life you could imagine.

In this case, event X = {an increase in salary of $20,000, an increase in salary of $30,000, a doubling of salary, gaining x-ray vision}.

The other kind of question was inherently negative:

2. Imagine that [event Y were to occur]. Consider how your life might change and rate it on a scale from -10 to 0, where -10 is the most uncomfortable life you could imagine, and 0 is your current comfort.*

In this case, event Y = {a decrease in salary of $20,000, a decrease in salary of $30,000, salary falling to $0, losing a leg}.

I. <u>Survey Statistics</u>

This study utilizes an original survey, which contains questions to address all major components of this paper. I collected the responses of 636 U.S. adults, strictly, using the SurveyMonkey Audience feature. The respondents hailed from 50 different U.S. states and territories. The distribution of demographics (age, race, sex, and income) mirrored the proportions seen in census data.

I. Empirical Framework on Pricing Race Change

This paper's first goal is to establish the asymmetry between the valuation of race between White and Black participants.

For all participants, I measured how much a fixed change in salary would affect one's comfort. For white participants, I measured how much their comfort would change if they were to become Black. For nonwhite participants, I measured how their comfort would change if they were to become white. The wellbeing change from race divided by the change from income multiplied by the margin of salary change provides the priced race change.

I. Results

Here are the average values for Priced[π], Priced[ψ], RelativeIncome[π], RelativeIncome[ψ], RelativeHousehold[π], and RelativeHousehold[ψ]. Several variables are not explicitly defined in this paper.

	Absolute	Relative to Personal Income	Relative to Household Income
Black to White (ψ)	$10,400	24.8%	31.4%
White to Black (π)	-$2,500	-13.2%	-11.5%
Ratio of Magnitude:	4.16:1	1.89:1	2.7

I. <u>Constructing a Theoretical Market for Race</u>

I suggest the following theoretical market to demonstrate the consequences of these results:

1. White individuals are able to sell their whiteness for some price to Black people.
2. Black individuals are willing to pay up to $10,400 for the benefit of whiteness; in other words, paying more than $10,400 would outweigh the benefits a Black person would get from being white.
3. White people do not feel that their whiteness has this same value. They would pay up to $2,500 to keep their whiteness. Any dollar amount more than this would not outweigh the perceived benefit of whiteness.
4. Experiential asymmetry exists when buyers and sellers are limited by the experiences they have, and the analog to perfect information, perfect experience, is unattainable.
5. All points above extend to the real-world diversity and inclusion market (DEI). Colored people undergo some psychological costs of assimilation, and DEI initiatives attempt to reduce the cost of color.
6. When white investors sitting on DEI boards undervalue the cost of color, they underestimate and underinvest, failing to maximize the utility of colored people.
7. Hence, a market failure arises in this market of race.

II. Introduction

In the final section of this paper, I note inconsistencies between my measures of benchmarking and the ideas of loss aversion, initially posited by Kahneman and Tversky (1979). Loss aversion suggests that people, regardless of race or income, hold more weight to losses than they do for equivalent gains. This study remains the most cited in economics because it suggested that neoclassical economic models, which assume rational agents, could not explain this disproportionate weight placed on losses. I systematically document a disproportion along the opposite margin; in particular, I find that theoretical gains in salary are valued substantially more than equivalent losses in salary, and that this disproportionate weight on gains is particularly high among people of color.

II. Empirical Framework

Let $\zeta(20{,}000) = \frac{\Delta\lambda}{\tau}$ be defined for each individual, where λ and τ are the changes in wellbeing from a $20,000 increase in salary and the change in wellbeing from a $20,000 decrease in salary, respectively. Intuitively, ζ(20,000) is just the proportion of the gain to an equivalent loss. I will refer to this as the loss aversion scalar for $20,000.

I generalize this derivation for ζ(20,000) to create variables ζ(30,000) and ζ(ALL) for the two other salary change questions. I regress these scalars against a binary for race.

II. Results

	ζ(20,000)	ζ(30,000)	ζ(ALL)
Coefficient for People of Color	0.216** (0.085)	0.433*** (0.108)	0.236*** (0.082)
Mean for White	1.02	1.05	0.964

II. Discussion

I note that the correlations between being a person of color and increased valuation of gains are robust to the inclusion of a host of control variables, even self-perceived valuation of money. There seems to be something inherently unique to being a person of color that explains the tendency to value gains more than equivalent losses.

Implications and Recommendations

The validity of the loss aversion results is not tested by the use of incentivized gains, as this endeavor remains peripheral to the paper. Limitations do exist for all the results above that are common in survey-based studies. The statistical power of these results could be aided by an increased sample size, and the random selection of survey participants was not done out of the entire U.S. population but within the population of SurveyMonkey respondents who engage with the audience feature, albeit a diverse population. Future research should aim to replicate my findings, validate the results on loss aversion using external games, and regress my findings on other margins to mediate the several correlations I observed.

This project addresses sustainability Goal 10 (reduce inequality within and among countries) and Goal 16 (prompts just, peaceful and inclusive societies) because it uses applied microeconomics methods to this sphere of race-induced asymmetries and racial empathy. It is the first in the literature to address and establish these findings, as far as I can tell.

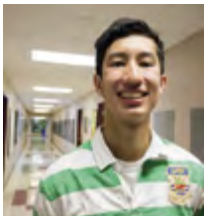
References

Akerlof, G. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Econometrica*, 84(3), 488-500. doi: https://doi.org/10.2307/1879431

Kahneman, D., Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica,* 47(2), 263-292. doi: https://doi.org/10.2307/1914185

Acknowledgements:

Dr. Reshmaan Hussam has my sincerest gratitude for stretching my brain with abstractions and showing me how to systematically study human behavior.



Identifying Essential Tremor Biomarkers

Nicholas Djedjos

Abstract

Nearly seven million individuals in the U.S. have Essential Tremor (ET). Differential Gene Expression (DGE) on ET RNA-seq data identified 86 differentially expressed gene transcripts. The gene transcripts were input into Gene Set Enrichment Analysis (GSEA) where five dysregulated gene pathways were identified. After filtering down to 32 genes with feature optimization, a Random Forest model predicted ET from control 85% of the time. Logistic regression was utilized to analyse the 32 genes individually, and 8 genes had a higher accuracy than 80%. Identifying these biomarkers shows that ET may be related to other neurological diseases more than once thought.

Introduction

Nearly seven million individuals in the U.S. have Essential Tremor (ET), making it one of the most common neurological disorders (Louis & Ottman, 2014). The chronic, progressive disease is characterized by characteristic involuntary and rhythmic shaking of the arm. As prevalent as ET is, there has been a relative lack of progress in understanding its genetic etiology (Clark & Louis, 2019). A major deterrent to understanding the genotypic heterogeneity of ET, until recently, has been the lack of technological advancement in RNA sequencing (RNA-seq) techniques. By examining the expression of genes within the brains of people with ET *post-mortem* and comparing their expressions to controls without the disease, genetic expressions pertaining to each disease could be identified. Predictive models built off genetic data from could elucidate further genetic differences between ET and controls.

Methodology

Raw RNA-Seq data was obtained from the publicly available National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) (Barrett et al., 2012). The contribution was from an essential tremor study by Columbia University (Control, n = 22; ET, n = 33) (Martuscello et al., 2020). Differential Gene Expression (DGE) took in the table of raw RNA-seq counts, normalized them, and removed outliers. Statistical analyses were run to find differentially expressed genes. To study the effect of ET on biological function gene sets, Gene Set Enrichment Analysis (GSEA) was used to discover overregulated or underregulated genetic pathways in ET and controls. GSEA grouped the individually significant genes with relevant gene sets and indicated enriched pathways. Random Forest and Logistic Regression were used to isolate potential target genes for diagnosis.

Results

Out of the 23,802 genes recorded, 86 were differentially expressed gene transcripts and were identified with $p < 0.001$ and false discovery rate (FDR) < 0.25. The FDR value of 0.25 has been used in similar genomic studies (Martuscello et al., 2020; Hawasawi, 2020), and was used in this study to expand the gene pool for future research validation. The Log Fold Change (LFC), an indication if a gene is upregulated or downregulated, for the genes were moderate. Specifically, the top ten genes with the smallest FDR have the smallest probability of being misrepresented.

Gene	False Discovery Rate	Log Fold Changes
NPC2	0.0339	0.724
ZNF26	0.061	0.275
S[illegible]	[illegible]	[illegible]
PTN2A2	0.0955	-0.412
ANKRD17	0.0955	0.156
[illegible]	[illegible]	[illegible]
SNX9	[illegible]	0.210
DPYSL5	0.[illegible]	0.235
PSD1L	0.207	0.200
[illegible]X045	0.207	0.252

Table 1: Indicates the top ten significant genes from DESeq2

To further investigate the potential role of the gene hits, GSEA was used to identify potential genetic pathways. The genes from DGE and their corresponding FDR and LFC were input into the gseares program on RStudio, along with the Hallmark Pathways and KEGG (Kyoto Encyclopedia of Genes and Genomes) gene sets from the Molecular Signatures Database (MSigDB).

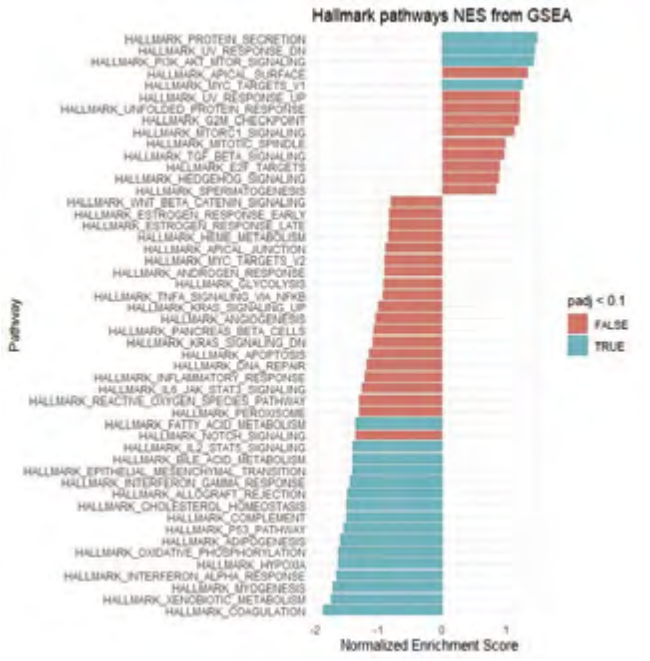


Figure 1: Hallmark Pathways NES from GSEA

Gene Set	Pathway
Hallmark	Cholesterol Homeostasis
Hallmark/KEGG	Fatty Acid Metabolism
KEGG	Ribosome
KEGG	Axon Guidance
KEGG	Parkinson's Disease

Table 2: Top 5 genetic pathways identified

Random Forest

Using the 86 genes from DEGE, a Random Forest classifier model (train-test split, 75%, 25%) with 1000 trees was utilized on the data set. K-Fold Cross Validation, a technique that splits the samples k times, using one split as the validation and k-1 as the test, was utilized. The accuracy of predicting ET or Control was 80% after ten runs. Feature selection, a technique where Random Forest chooses the most weighted genes, was applied to the set to reduce noise and improve accuracy. The gene set transformed 86 genes to 32 genes. The Random Forest model trained with a 0.75, 0.25 train test split along with K-Folding achieved an accuracy of 85% with the refined data set.

Logistic Regression

Using the 86 genes obtained from DGE, a binary logistic classifier to predict Control or ET solely from RNA-seq data was created. The weights were based on how well each gene After ten runs, the accuracy of a train and test split of 75%, 25% was 78%. A new Logistic Regression model was used on each gene from the Feature Selected Random Forest genes, and notably, 8 genes performed better individually than all 86 in combination.

Gene	Accuracy Score
SFTPA2	0.93
NLRP14	0.91
PLCD1	0.86
SCRG1	0.86
ANKZF1	0.83
INPP5D	0.82
EVAIC	0.80
BTN3A1	0.80

Table 3: Top 8 genes from Logistic Regression

Discussion

ET's lack of clinical diagnosis and its resulting genotypic heterogeneity makes it paramount to analyze RNA-seq data from patients. While this is a broader study not specifically testing for possible genetic variants such as LINGO1 and LINGO2, a possible set of related ET genes and genetic pathways were identified.

Genetic Pathways

The downregulation of the cholesterol homeostasis pathway signals that known cholesterol homeostasis genes are not being expressed as much in ET as they are in controls, indicating that there could be cholesterol buildup in the cerebellum. The downregulation of the fatty acid metabolism pathway indicates that there may be fatty acid buildup in the brains of patients with ET, which is similar to Alzheimer's disease. The downregulation of the ribosome pathway shows that an insufficient amount of protein is being produced which could result in the death of Purkinje cells. The upregulation of the axon guidance pathway indicates that there may be an increased number of axonal profiles and branching, which corroborates with other ET studies. The downregulation of the Parkinson's disease genetic pathway reveals that known inhibition genes to Parkinson's are being downregulated, showing that ET may be related to Parkinson's Disease

Predictive Machine Learning Models

Both Logistic Regression and Random Forest models were used to predict gene candidates from the initial DGE analysis. Prior genomic studies with ET have used solely DGE and GSEA to identify genes associated with the disease. However, machine learning models can identify unseen relationships between two classes (ET and control) and are a viable way to generate gene candidates.

The Random Forest Model had an 80% accuracy with the original 86 DGE genes, and the accuracy increased to 85% after feature selection. This model has clinical significance, as it can be used as a tool in ET diagnosis after brain biopsies. The Logistic Regression model had a 78% accuracy with the 86 DGE genes, providing another model to use for clinical diagnosis. Logistic Regression was also used on each individual gene to predict ET or Control, and notably, 8 genes had a higher accuracy than 78%. This indicates that, solely by looking at the RNA-seq expression of the aforementioned genes, the model was able to predict if the patient had ET at least 80% of the time. It is important to analyze these genes further as they could be potential biomarkers for ET.

Implications and Recommendations

The creation of a machine learning model between ET and Control can be used clinically in addition to other diagnoses for other neurological diseases. This study shows that ET could be more related to Parkinson's and Alzheimer's and may serve as a precursor. The study also corroborates with the Axon Guidance pathway seen in previous ET studies, validating past data. Further, the gene identifications align with Sustainable Development Goal 3: Good Health and Well-Being.

Future Research

In the future, as with all machine learning and bioinformatic studies, more data will make the analyses more consistent. Further, since this was a computational study, the genes need to be isolated in the wet lab in an in-vivo study to examine the effect of their inhibition or activation. Machine learning methods such as CNN would have higher statistical power and could be of use as well.

References

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks119

Clark, L. N., & Louis, E. D. (2018). Essential tremor. *Neurogenetics, Part I*, 229–239. https://doi.org/10.1016/b978-0-444-63233-3.00015-4

Hawsawi, Y. (2020). Peer review #1 OF "Differentially expressed genes and key molecules of BRCA1/2-mutant breast Cancer: Evidence from Bioinformatics analyses (v0.2)". *PeerJ*. https://doi.org/10.7287/peerj.8403v0.2/reviews/1

Louis, E. D., & Ottman, R. (2014). How Many People in the USA Have Essential Tremor? Deriving a Population Estimate Based on Epidemiological Data. *Tremor and Other Hyperkinetic Movements*, *4*, 259. https://doi.org/10.5334/tohm.198 Martuscello, R. T., Kerridge, C. A., Chatterjee, D., Hartstone, W. G., Kuo, S.-H., Sims, P. A., Louis, E. D., & Faust, P. L. (2020). Gene expression analysis of the cerebellar cortex in essential tremor. *Neuroscience Letters*, *721*, 134540. https://doi.org/10.1016/j.neulet.2019.134540


The Impacts of Cosmic Rays on Planetary Climate Change

Stephanie Ressel & Hailee Sexton

Abstract

Amidst potential climate change conclusions and solutions, there are stellar influences on Earth's climate called cosmic rays, which heavily impact the formation of clouds; therefore, they are directly related to the total energy transferred from the Sun to the Earth's atmosphere. In this experiment, cosmic rays were observed and collected through global cosmic ray detectors, and ones assembled by the investigators to explore their relationship with air ionization, cloud formation, and temperature. Data collected from the cosmic ray detection network, along with that of directly recorded data, will be analyzed in conjunction with global cloud coverage information sourced through satellites and then largely compared with planetary climate trends. The trends, if any, were investigated in this experiment. Through the analysis of local data, a significant trend between the concentration of muons in the atmosphere and cloud coverage was not found.

Introduction

In the ever-present search for climate change solutions, one of the space factors which influence Earth's climate are cosmic rays, which heavily impact the formation of clouds and therefore are directly related to the total energy transferred from the Sun to the Earth's atmosphere. As cosmic rays enter the atmosphere, they cascade ionized particles that cause an increase in air ionization and the production of aerosols in the lower atmosphere (30-45km), promoting the formation of clouds. Periods of high cosmic ray activity are called fluxes. More energetic fluxes aid in the production of seed nuclei for condensation known as cloud condensation nuclei (CCN) (Christienson et al,. 2020). Furthermore, clouds directly influence the atmosphere through cooling and heating by reflecting shortwave solar radiation and trapping longwave radiation through the greenhouse effect (Dorman et al,. 2004). Cosmic rays also are presumed to directly influence temperature change, having likely played a role in a 0.6-degree temperature increase over the past century (Devendraa, 2010). The objective of this experiment is to investigate a trend, or lack thereof, between cosmic ray activity and local cloud coverage. Because the upper atmosphere interactions that occur when cosmic rays initially enter the atmosphere deposit seed nuclei, cloud formation is more likely in the presence of an abundance of seed nuclei. It was hypothesized that there will be a positive trend between cosmic rays and cloud coverage. In other words, the more cosmic ray activity a celestial body experiences, the more ideal conditions for clouds to be formed, therefore the greenhouse effect is heightened during periods of high cosmic ray activity.

Methodology

After the 2-week data collection period, data was uploaded for analysis to the QuarkNet Cosmic Ray Collaboration server. Data analysis was then modeled, and a baseline "normal" muon activity of 1.5 was established. Cloud opacity data collected from public record was also analyzed to develop a baseline standard cloud opacity. Deviations from these standards identify periods of high muon flux or low muon flux. The fluxes are compared to spikes in cloud opacity to determine whether these event trends are correlated.

Results

Figure 1 was developed utilizing the QuarkNet Cosmic Ray data analysis servers.

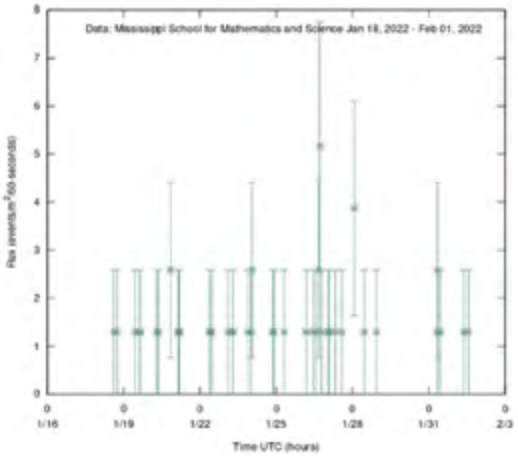


Figure 1. Recorded muon activity/interactions from local detector in Columbus, MS.

According to Figure 1, the baseline for expected cosmic ray flux for the local detector was approximately 2.5 events per area per second.

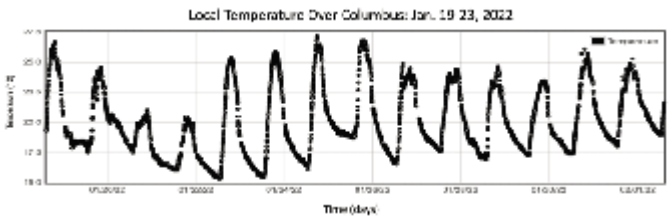


Figure 2. Reflects daily temperature trends. Oscillations can be seen with the rise and fall of the sun.

A spike in flux was recorded between Jan. 26 to Jan. 28, but as observed in both Figure 2 and 3, this spike in cosmic ray flux had no significant influence on cloud opacity or temperature.

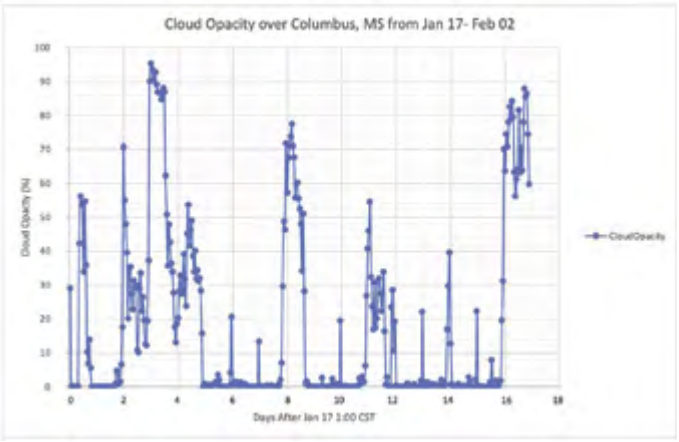


Figure 3. Displays local cloud opacity data recorded over Columbus, MS from January 17, 2022, to February 02, 2022.

However, Figure 2 and 3 suggest an inverse relationship between cloud cover and temperature as data recorded within the period from approximately Jan. 19 to Jan. 23 appear to experience deviations. Figure 2 displays an average temperature to be the lowest during this interval whereas Figure 3 displays a peak in cloud opacity for the same time.

Discussion

Upon comparison of collected muon activity data and that of local cloud coverage, it was concluded that there is no significant trend between the two. However, during the time that low temperatures were recorded, very high cloud opacity was observed. This connection to temperature and cloud coverage validates the direct relationship that these two variables have with one another. According to the small scale, local data samples over the course of a two-weeks, no conclusion could be drawn relating local cloud cover and cosmic ray influence.

Due to the nature of the experiment– data collected from a relatively small geographic area over a short period of time– the data collected represents a niche observation over the city of Columbus, MS. Observations this niche would not present the full picture and would not yield a reliable global generalization.

Implications and Recommendations

Climate change is a growing problem across the globe today, with extensive research on causes and its significance in the area of environmental science. However, cosmic rays are not heavily researched as a phenomenon that correlates with these growing changes in temperature and cloud cover. By better understanding how the upper atmosphere is being affected through external forces in conjunction with internal forces of pollution and carbon emissions, both sides should be considered in providing a solution to the climate crisis. Further research and analysis can lead to new perspectives on indirect causes of climate change, as well as more preventative methods.

Sustainable development goal 13: Take urgent action to combat climate change and its impacts.

Further research into the correlation of concentrations of cosmic rays and climate change can benefit the advancement of this goal, allowing improvement of awareness and impact reduction.

Future Research

In the future, a multitude of advancements will be made to this experiment. Expanding the recorded data to global cosmic ray detectors, as well as global cloud opacity data over longer periods of time, would provide more reliable information to make generalizations toward this celestial phenomenon. There will also be a statistical process added to analyzing data, which will provide more insight into the true significance levels and trends in the data.

References

Bogdan, Thomas J. “The Chilling Stars: A New Theory of Climate Change.” Bulletin of the American Meteorological Society, vol. 89, no. 3, Mar. 2008, pp. 392–93. EBSCOhost, search-ebscohost-com.libprxy.muw.edu/login.aspx?direct=true&db=a9h&AN=31686546&site=ehost-live&scope=site

Dorman, L. I.: Cosmic rays and space weather: effects on global climate change, Ann. Geophys., 30, 9–19, https://doi.org/10.5194/angeo-30-9-2012, 2012.

Dorman, L. I. “Cosmic Rays and Space Weather: Effects on Global Climate Change.” Annales Geophysicae (09927689), vol. 30, no. 1, Jan. 2012, pp. 9–19. EBSCOhost, doi-org.libprxy.muw.edu/10.5194/angeo-30-9-2012.

Introduction to the QuarkNet Cosmic Ray Detector. QuarkNet Cosmic Ray Detector. (n.d.). Retrieved April 5, 2022, from https://www.i2u2.org/elab/cosmic/teacher/detector.jsp

Sami K Solanki, Solar variability and climate change: is there a link?, Astronomy & Geophysics, Volume 43, Issue 5, October 2002, Pages 5.9–5.13, https://doi.org/10.1046/j.1468-4004.2002.43509.x

“Searching for Data.” Searching for Data | NOAA Climate.gov, https://www.climate.gov/maps-data/climate-data-primer/finding-climate-data/searching-data.

Acknowledgements

We thank Mr. William Funderburk for his contagious passion of physics and inspiring his students to pursue lifelong learning.

Assessing the Effect of Traveling on COVID at the County Level Using Machine Learning

Andrew Yu

Abstract

Population mobility is correlated with Covid-19 transmission, making it an ideal indicator for forecasting future outbreaks. Additionally, given the proliferation of vaccines throughout the US population, vaccination status was also be considered in this study. Using these factors, the goal of this study is to develop an accurate machine learning model to predict future Covid-19 cases in the state of Mississippi, at the county level. Using county-level traveling data from the Bureau of Transportation Statistics and vaccination statistics from the Center for Disease Control, a time series machine learning model to predict changes in these factors was created by leveraging the Python package AutoTS. Then, the newly forecasted traveling and vaccine data were incorporated to predict future Covid-19 cases. Thus, forecasts for Covid-19 at the county level were be created, allowing for comparisons of model accuracy between counties, as well as forecasts of the state as a whole to be made.

Introduction

Because population mobility has been shown to correlate with COVID transmission, it is an ideal indicator for creating accurate predictions (Ribeiro et al., 2020). Furthermore, vaccination status was included in this study as another factor in COVID spread. It is also important for predictions to be made at a local level such that policy implementation can be tailored to small populations. Despite this, little research is conducted at the county level nor is there much research regarding the state of Mississippi. Thus, this research is also important to address the lack of literature in this area.

Methods

Databases: Transportation statistics describing the number of residents staying at home and the number of trips taken by residents were retrieved from the Bureau of Transportation Statistics (University of Maryland, 2022). Detailed vaccination data from the Center for Disease Control was used (IISInfo, 2022). Also from the CDC, the number of new cases per 100,000 people was used as a metric that is comparable among counties of varying populations (CDC, 2022).

Data Imputation: For data present in one dataset but not others, values were deleted. To replace missing data, a K-Nearest Neighbors imputer was chosen to infer missing values based on nearby data points, as it was assumed that data would be more consistent for nearby dates.

Time Series Model Development: Because most machine learning models can only establish correlation between factors that are not temporal, a time series model was chosen to account for the temporal factor in COVID data. Using the Python package AutoTS, a genetic algorithm was implemented to generate the most optimal model for the data and the best set of parameters in that model. In the end, the genetic algorithm chose different models for each county to increase accuracy.

Results

After training the model, the percentage error for each of the counties was extrapolated and displayed below:

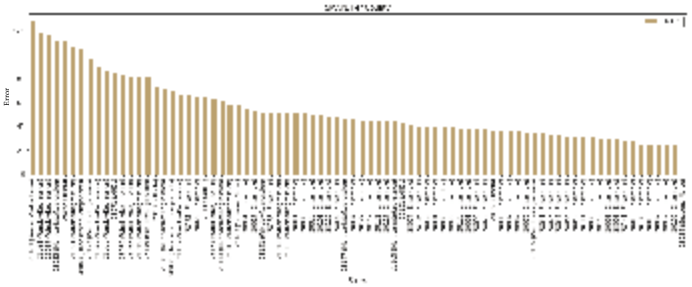


Figure 1: Error in Each County

The error metric was the symmetric mean percentage error (SMAPE), which is an error metric used for time series that can take on values from 0% to 200%. Interestingly, the county with FIPS code 28055, Issaquena county, had an error of 0%, but this is most likely due to lack of data, as Issaquena county has the lowest population in Mississippi. For the next two counties with the lowest error, Webster and Itawamba county, forecasts for the next 100 days are displayed below:

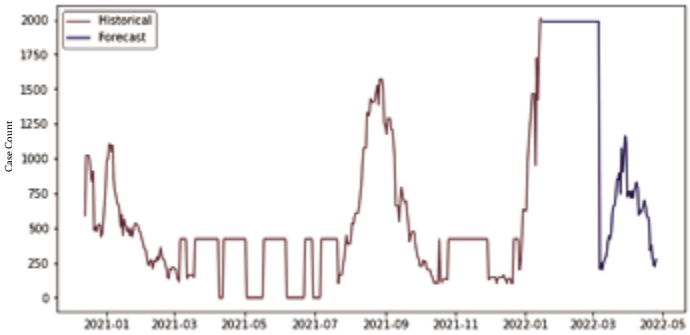


Figure 2: Webster County Forecast

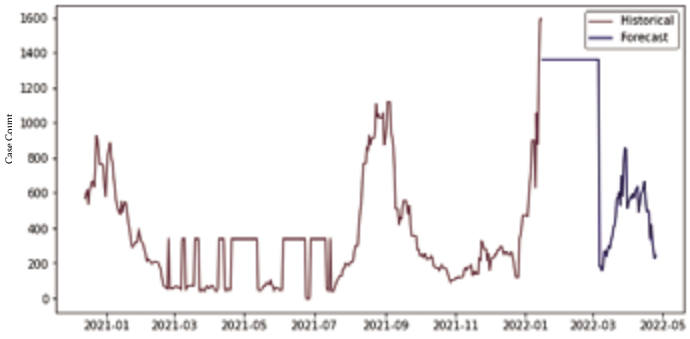
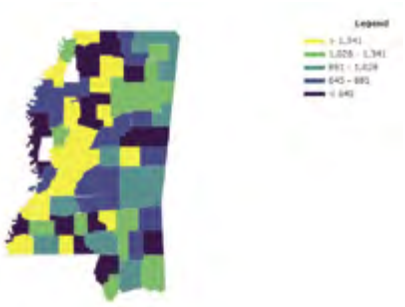


Figure 3: Itawamba County Forecast

To display results from the state of Mississippi as a whole, a choropleth graph was created. The values displayed were calculated by taking the mean of the COVID Cases during the forecasting period. To create a legend of five categories, the data was separated into quintiles. For two counties which are not colored, the data was deemed unsatisfactory for the model and thus were not forecasted. A notable feature seen in this choropleth is the higher values near central Mississippi, where traveling likely occurs more. Furthermore, counties near each other generally have similar values, which shows how traveling occurs across counties.



Conclusion

Through this study, a time series model was developed to predict changes in COVID cases using traveling and vaccination data. In the final model, error rates varied across counties, indicating possible disparities in data collection which may affect model accuracy. In two of the highest accuracy models, the shapes of forecasts were similar, with a relatively stable forecast for cases until March, in which there is both a sharp decline and incline in cases. The rise in cases could possibly be attributed to spring breaks occurring during that time period, in which the number of travels increases. To represent results across the state as a whole, a choropleth was created, in which values were calculated through taking the mean of forecasted data. In the choropleth, travel patterns can be visualized, as greater travel across central Mississippi results in higher predicted COVID count. Moreover, similarities across nearby counties further show that travel patterns do indeed play a role in predictions.

Future Studies

Better optimization may increase the utility of models developed. By default, all models in the library were tested, but it may be beneficial to handpick fewer models which are more suited for the study. Within the chosen models, parameters were mainly tuned automatically by the AutoTS library, but manual hyperparameter tuning may increase accuracy. Creating a better dataset could be possible. New data is continuously being updated, allowing for the model to have a larger training dataset if implemented in the future. Because some counties in Mississippi are very rural, data collection is limited in these areas. Better strategies for data imputation could be considered. For future studies, the methodology used in this research could potentially be extended to a country-wide scale, where every county in the US could be forecasted. These predictions could assist in policy implementation, thus helping accomplish the UN Sustainable Development Goal #3 of Good Health and Well-Being.

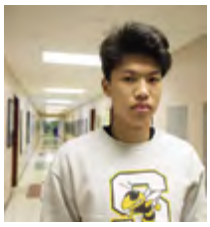
References

CDC. (2022). United States covid-19 county level of community transmission historical changes. Centers for Disease Control and Prevention. Retrieved February 12, 2022, from https://data.cdc.gov/Public-Health-Surveillance/United-States-COVID-19-County-Level-of-Community-T/nra9-vzzn

IISInfo. (2022, February 11). Covid-19 vaccinations in the United States,county. Centers for Disease Control and Prevention. Retrieved February 12, 2022, from https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh

Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland. (2022, February 7). Trips by distance: Open data: Socrata. Bureau of Transportation Statistics. Retrieved February 12, 2022, from https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv

Ribeiro, S. P., Dáttilo, W., Barbosa, D. S., Coura-Vital, W., Chagas, I. A. D., Dias, C. P., Silva, A. V. D. C. E., Morais, M. H. F., Góes-Neto, A., Azevedo, V. A., Fernandes, G. W., & Reis, A. B. (2020, September 16). Worldwide covid-19 spreading explained: Traveling numbers as a primary driver for the pandemic. Anais da Academia Brasileira de Ciências. Retrieved December 13, 2021, from https://www.scielo.br/j/aabc/a/76CfqdL5pPBZLcQy9FdWwxn.



Qiancheng (Sam) Sun

Application of Machine Learning for Energy Materials Prediction

Abstract

Energy storage is key to fully transitioning to renewable energy. The high cost and low durability of batteries make it an unlikely candidate for large-scale energy storage. On the other hand, thermochemical energy storage provides a promising solution with its low-cost and high durability. However, more research and testing are needed to determine the optimal materials for this task. This study will train two machine learning models to predict energy per atom and total dielectric constant. Data was extracted from Materials Project via Matminer. Then it was cleaned so the desired features and target properties could be fed to the Machine Learning Algorithm. XRDPowderPattern describes the crystal structure of materials, and it was used as the feature to train the model. After training and testing with different splits (80-20 and 90-10), we were able to get a training percentage error of 10.95% for energy per atom and a testing percentage error of 34.47%. On the other hand, we got a training percentage error of 17.66% for the dielectric constant and a testing percentage error of 46.19%. From these results, we see that although machine learning can be used to semi-effectively predict materials' energy properties. However, this study also reveals a challenge faced by the implementation of Machine Learning to Materials- few available datasets.

Introduction

In 2020, renewable sources made up of 29% of all global electricity generation, a two percent increase from 2019 (Renewables – Global Energy Review 2021 – Analysis - IEA, 2021). Due to the inherent intermittency of renewable energy, reliable energy storage technologies are required to achieve a fully renewable future. Energy storage is typically associated with batteries, but battery's high cost and low durablity makes it a challenge for large scale energy storage. Over the years, battery prices have gradually fallen with improved manufacturing techniques, dropping to as low as $137/kWh in 2020 with a projected price of $100/kWh in 2023 (Renewable Energy World, 2020). While this is a promising trend, the price is still far too high when compared to an alternative option: thermochemical energy storage (TCES). From initial stages of testing, it is found that the cost of TCES is around $15/kWh, (*Thermal Storage R&D for CSP Systems*, 2021) almost $1/10^{th}$ the cost of current Li-ion batteries.

On the other hand, Machine Learning has, in recent years, been used to effectively predict material properties such as crystal structures and density. The objective of this study is to develop machine learning models that will be able to predict energy properties of materials.

Methodology

This study trains two different Machine Learning models to predict energy per atom and total calculated dielectric constant (eps_total). The same ML algorithm and cross-validation techniques were used for both model. The Machine learning algorithm for this study was trained and tested on Python. Specifically, Random Forest Regressor was imported from the Scikit-learn library. By using the Kfold and cross_val_score functions from Scikit-learn, the data is randomly split into 80% for training and 20% for testing.

According to osti.gov the most promising metal-oxides for thermochemical energy storage are Magnesium Oxide, Manganese Oxide, Iron Oxide, Cobalt Oxide, Lead Oxide, and Copper Oxide. Data for 64 of materials containing above elements are retrieved from the Materials Project via Matminer by using MPRester. XRDPowderPattern describes the crystal structure of materials, and it was used as the features to train the model. The target property or the property that the Machine Learning model is trying to predict is Energy per atom. This property is chosen because the energy per atom is closely related to how much energy a material can store.

For the Dielectric constant model, of 1296 materials was extracted from the phonon_dielectric_mp dataset on Matminer. XRDPowderPattern and electric contribution are used as features to train the ML model. The target property or the property that the Machine Learning model is trying to predict is the total calculated dielectric constant. This property is chosen because, by definition, it mean the ability of a substance to store electrical energy in an electric field.

Results

The Machine Learning algorithm for energy per atom performed very well in the training stages, producing an average percent error of 10.95% which is widely considered as an excellent prediction accuracy.

However, it did not perform as well during testing, yielding a relatively high average percent error of 34.57%. This is acceptable but not ideal.

Two factors could have hindered the model's accuracy:

1. Machine Learning requires large amounts of training data, but there are limited free publicly available data,

especially when considering only those than can be used for Thermochemical energy storage.

2. The feature used to train this model is XRDPowderPattern and it is possible that there is not a strong correlation between a material's XRDPowderPattern and its energy per atom

The Machine Learning algorithm for Dielectric constant performed well in training stages, producing an average percent error of 17.66%.

However, its performance dropped drastically in testing stages, producing an average percent error of 46.19%. Note, the performance of the model is better than suggested by the percent error due to outliers.

Discussion

From the results of the two Machine Learning Models, it can be concluded that XRDPowderPattern is quite effective at predicting energy per atom but less effective at predicting dielectric constant.

This study reveals a challenge faced by the implementation of Machine Learning to Materials- few available dataset. Traditional applications of ML such as image processing and voice recognition has millions of readily available data point, but there are only a handful of material datasets. As a result, current machine learning models for materials are proven to be useful but tends to be less accurate than other implementations.

Future Research

Other Machine Learning algorithms such as K-nearest neighbor, Gaussian Naïve Bayes Classifier, and Support Vector Regression will be tested to see their effectiveness on predicting materials' energy properties.

Similar ML models should be trained on energy density, redox temperature, and reaction enthalpy so the results will have a more direct correlation with Thermochemical Energy Storage.

With access to more datasets and properties mentioned above, this procedure can be used to accurately predict the properties of unknown materials. This will accelerate the development of thermochemical energy storage and create a more sustainable future by meeting sustainable development goal seven: ensure access to affordable, reliable, sustainable, and modern energy for all.

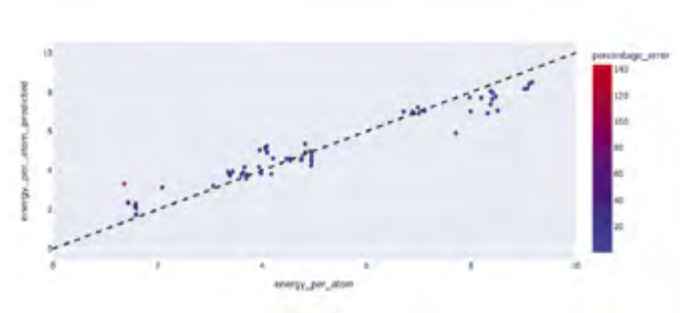


Figure 1: Actual energy per atom vs predicted energy per atom with percentage error (training)

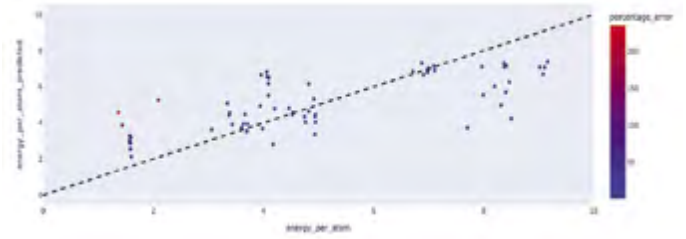


Figure 2: Actual energy per atom vs predicted energy per atom with percentage error (testing)

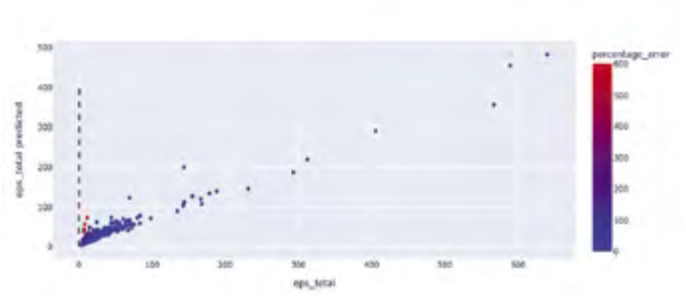


Figure 3: Actual dielectric constant vs predicted dielectric constant with percentage error (training)

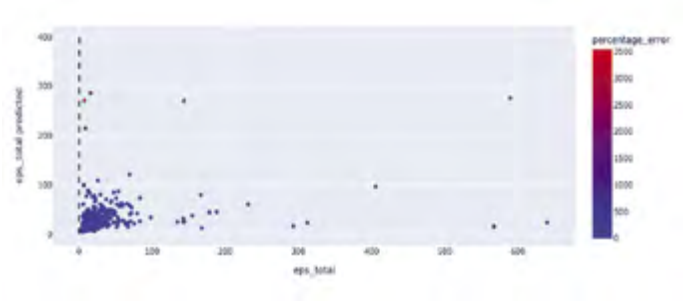


Figure 4: Actual dielectric constant vs predicted dielectric constant with percentage error (testing)

References

Environmental and Energy Study Institute (EESI. (2019). *Fact Sheet | Energy Storage (2019) | White Papers | EESI*. Eesi.org. https://www.eesi.org/papers/view/energy-storage-2019

Matminer (Materials Data Mining) — matminer 0.7.6 documentation. (2015). Lbl.gov. https://hackingmaterials.lbl.gov/matminer/

Renewables – Global Energy Review 2021 – Analysis - IEA. (2021). *Renewables – Global Energy Review 2021 – Analysis - IEA*. IEA. https://www.iea.org/reports/global-energy-review-2021/renewables

Renewable Energy World. (2020, December 18). *Annual survey finds battery prices dropped 13% in 2020 - Renewable Energy World*. Renewable Energy World. https://www.renewableenergyworld.com/storage/annual-survey-finds-battery-prices-dropped-13-in-2020/

Thermal Storage R&D for CSP Systems. (2021). Energy.gov. https://www.energy.gov/eere/solar/thermal-storage-rd-csp-systems

Hayden Anderson

Analyzing the Cause of Triple Negative Breast Cancer (TNBC) using Machine Learning Methods

Abstract

Cancer, one of the leading causes of death around the world, has evolved throughout time and resisted the different drugs that scientists have made throughout the years. Furthermore, the different types of cancer have made the analysis of the disease even harder than before. One type of cancer that has made this analysis hard is Triple Negative Breast Cancer (TNBC), a type of cancer that is negative for estrogenic receptors, progesterone receptors, and human epidermal growth factor. Though researchers have been researching TNBC for years, they still know little about the cause of TNBC due to the lack of these receptors and human epidermal growth factor. However, it is suspected that a mutation within the Breast Cancer Type 1 Gene (BRCA1) is what causes TNBC. Therefore, it is important to be able to use different methods to find out how this type of cancer has been caused. This study uses datasets from the Cancer Genome Atlas (TCGA) and cBioPortal to connect the gene mutation that causes this cancer using trends and models. To do this, machine learning can be used to find these trends and create these models. This process involves the dataset being pre-processed and trained to use different machine learning methods to find a correlation between the BRCA1 mutation and TNBC. After using the methods such as K-Nearest-Neighbors and Support Vector Machine (SVM), it was shown that TNBC most closely resembled the BRCA1 gene which reflected the potential causation of TNBC by BRCA1.

Introduction

Triple Negative Breast Cancer (TNBC) is a form of breast cancer that tests negative on all three components of a breast cell: estrogen receptors, progesterone receptors, and the human epidermal growth factor. Though not severely uncommon, TNBC is a cancer that doctors know little about due to the lack of the components of the breast cell ("Triple Negative Breast Cancer," n.d.). Furthermore, since analyzing the cause of breast cancer has been based on the expression of estrogen receptor α, a receptor not found in TNBC, the causation of TNBC is not known. However, it is expected to be due to the Breast Cancer Type 1 Gene (BRCA1) mutating since BRCA1 is supposed to prevent breast cancer (Kothari et. al, 2020) and mutation will causes the gene to lose its primary purpose. To help provide evidence that the mutation causes TNBC, machine learning, a method that allows for a computer to easily build models for data, can be used to find the trends within people that have TNBC (Chen et. al, 2021). Through the training of datasets to find patterns within certain datasets, this unknown causation of TNBC can be found.

Hypotheses

There were two hypotheses within this project. The first hypothesis was that Machine Learning could be used to determine the gene that causes TNBC. The second one was that if it is available, then that a BRCA1 mutation is the reason for TNBC.

Methodology

Datasets

For this project, two datasets were used. The first one was The Metastatic Breast Cancer Project Dataset which had samples of metastatic breast cancer (MBCA), BRCA1 genes, and the characteristics of these. This was used for personal knowledge rather than for the models. The next dataset was the Cancer Genome Atlas which gave me samples of metastatic breast cancer and the BRCA1 genes within those cells. The data in these datasets were used to test the presence of the human epidermal growth factor in mutated BRCA1 cells.

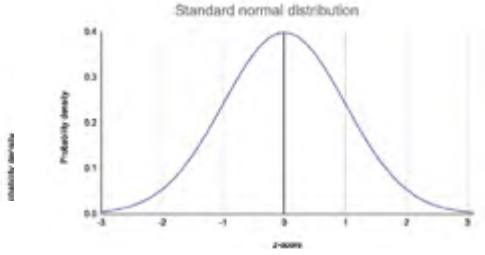
Normalization of Data

The next task was to pre-process the data. This happened for two types of data: categorial and numerical. For categorical data, a method known as one-hot encoding was used to transform the categorical data to numerical data. This type of encoding turns the data into a series of ones and zeroes so that the models were able to read the data easily. One example of this is shown below in the first five rows of the variable known as "her2_status_by_ihc," a variable that shows the status of a human epidermal growth factor.

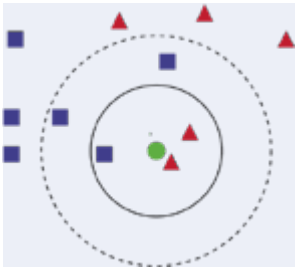
```
0        Positive
1   Indeterminate
2        Positive
3        Equivocal
4        Negative
Name: her2_status_by_ihc, dtype: object
```

	Equivocal	Indeterminate	Negative	Positive
0	0	0	0	1
1	0	1	0	0
2	0	0	0	1
3	1	0	0	0
4	0	0	1	0

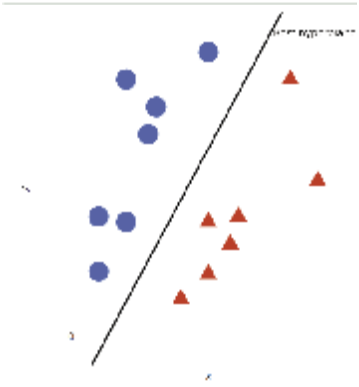
Next, the quantitative data needed to be put inside of a range that was easier for someone to read. For example, it is not easy for a human to analyze a graph that goes from y=1 to y=100,000. Instead of using the raw numerical data, a Z-score was used to put the data inside a range that was easier to read. A Z-score is the number of standard deviations that the datapoint is away from the mean. This usually has a range of seven unless severe outliers are present. This creates a normal graph as shown below.



After the data was normalized, the data was put into two different algorithms. The first one is the K-Nearest-Neighbors (KNN) algorithm, a machine learning algorithm that classifies a datapoint based on the majority K datapoints around it. This classification method helps estimate the likelihood that a certain datapoint is a certain category. To do this, four different variables were created: two training sets and two test sets for both x and y. This allowed for creating a model and then testing it with the testing group. The function was designed to create a graph that showed the best K-value. For this project, a K of five was used because it tested with the best accuracy. In the model below, a K of five means that the model tested for the datapoints within the dotted circle.



The second algorithm was a Support Vector Machine (SVM). For an SVM, different datapoints are put on graph and then a vector is drawn on the graph to model the data. This vector must be as close and equidistant as possible to the other groups for there to be an accurate algorithm. The data was inserted into a graph and code was used to construct a vector that best fits the datapoints. Shown below is an example of how SVMs work.



Results

For this data, again, two different algorithms were used. When tested with the data, both of the algorithms showed a negative correlation for the presence of the human epidermal growth factor within the mutated BRCA1 genes. For the first algorithm, KNN, there was a 67% chance of it predicting whether the human growth factor was negative within the mutated BRCA1 gene or not. The SVM proved a 53% chance of prediciton of the human growth factor being negative within the cnacerous cell. To prove this, the algorithm made the following confusion matrices in order to find the accuracy of the model. In Figure 1, the confusion matrix for the KNN is shown.

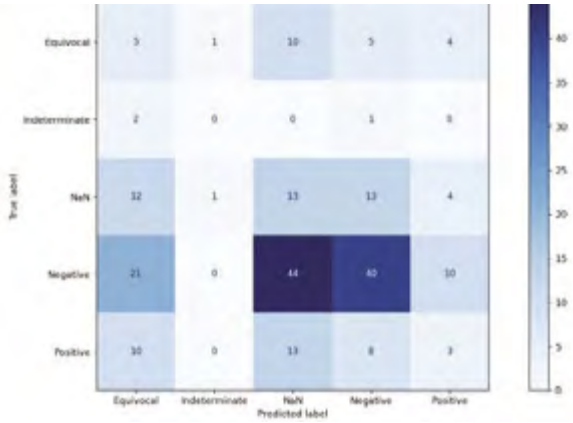


Figure 1: Confusion matrix for the KNN algorithm

Figure 2 shows the confusion matrix for the SVM algorithm

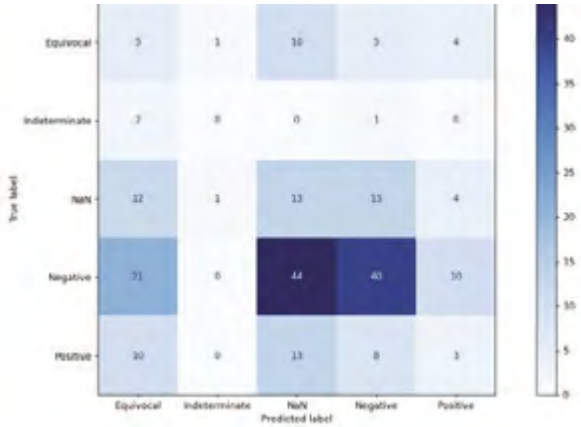


Figure 2: Confusion matrix for the SVM algorithm

Conclusions

The models were constructed using a KNN algorithm and an SVM algorithm to produce the best results possible. The database was used from one of the largest BRCA1 databases that contains mutations and other characteristics about it. For this project, the epidermal growth factor was tested since studies have shown that if one of them tests negative, all the three components will test negative.There was a 67% and 53% chance of success given to the K-Nearest-Neighbors and Support Vector Machine algorithms, respectively. Following the success rates, the model showed confusion matrices which showed how well it calculated different scenarios of different predicted values and various true values. Since the algorithms had relatively high success rate in the models, it can be said that more research needs to be done on BRCA1 mutations and TNBC since there is a probable causation of this mutation. This research helps achieve the goal of Sustainable Development Goal Type 3: Good Health and Well-Being. Since this tries to help cure a type of disease, this research fits this category best.

References

Chen, Z., Wang, M., De Wilde, R. L., Feng, R., Su, M., Torres-de la Roche, L. A., & Shi, W. (2021, September 17). A machine learning model to predict the triple negative breast cancer immune subtype. Frontiers in immunology. Retrieved January 27, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8484710/

Kothari, C., Osseni, M. A., Agbo, L., Ouellette, G., Déraspe, M., Laviolette, F., Corbeil, J., Lambert, J.-P., Diorio, C., & Durocher, F. (2020, June 26). Machine learning analysis identifies genes differentiating triple negative breast cancers. Nature News. Retrieved January 27, 2022, from https://www.nature.com/articles/s41598-020-67525-1

Triple-negative breast cancer: Details, diagnosis, and signs. American Cancer Society. (n.d.). Retrieved January 27, 2022, from https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer/triple-negative.htm

Margaret (Maggie) Buck

Investigating the Effect of *Allium sativum* on *Fragaria* × *ananassa* Propagation Year 2

Abstract

Allium sativum (garlic) has long since been dubbed the king of companion planting, due to garlic's ability to ward off insects and diseases. Garlic as a common companion plant will usually encourage the growth of its intercrops, but there are some disagreements on planting garlic with plants that need plenty of nitrogen due to the competition between them. This study examines the effects of garlic over time on the growth of *Fragaria* × *ananassa* (strawberries). This was done by taking growth measurements every three weeks using two experimental boxes and one control box over the course of five months. The size of the roots were also measured at the end of the growing period. The effects of garlic on the growth of strawberries were compared by the rate of growth between the control and experimental plant boxes. The results showed the difference in intercropping and without garlic was significant initially with the first two measurements of plant height while the later ones were insignificant due to plant death. This shows intercropping garlic in strawberries does lead to poorer strawberry growth.

Introduction

Question: How does *Allium sativum* affect the growth of *Fragaria* × *ananassa*?

H_0: *Allium sativum* does not affect the growth of *Fragaria* × *ananassa.*

H_1: *Allium sativum* decreases the growth of *Fragaria* × *ananassa.*

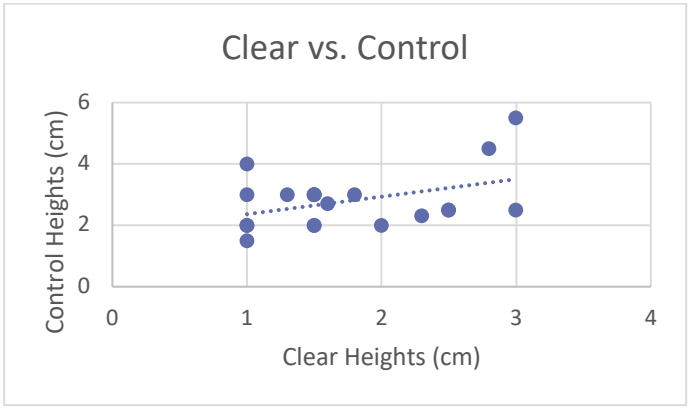
Allium sativum (garlic) has been coined the king of companion plants due to its compatibility with most plants, as well as being a natural insecticide and fungicide. This is due to the garlic chemical composition, which includes many sulfur compounds. Garlic is a relatively common intercrop for insect barriers. One example of this is its use in cabbages where it increased survival rate by 10.2% (Mischeck & Katsaruware, 2014). However it has also been found to increase yield in geranium significantly with a 22.3% increase in essential oil (Singh, 2015). The effect that garlic has on its intercrops is scattered and conflicting. There are many sources that report that garlic increases yield, while others report it decreases yield and the research mostly focuses on cole crops and distillates instead of intercropping. Garlic is also commonly used a pesticide which includes it being used to control *Tetranychus urticae* where 50% of the insects were killed with garlic distillate (Attia, et al., 2011, b). This research even continues into effective concentrations of the distillate which was 7.49% (Attia, et al., 2011, a). This study investigates the effects in *Fragaria* × *ananassa* (strawberries) to expand the scope of garlic intercropping knowledge. The growth of the strawberries and garlic was compared every three weeks once the strawberry plants reached sufficient size for measurement. The root size of the strawberries was also be compared at the end of the growing period.

Methodology

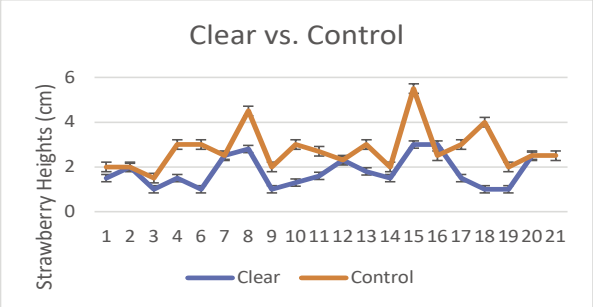
Growing Plants: The garlic and strawberries were planted in October and were watered as needed. They were planted as two intercropped trays of garlic and strawberries and one control tray without garlic. The experiemental group had a clear tray and a tan tray. Four bulbs of garlic were planted in each experimental tray and strawberry seeds were scattered across the soil inside each container. They were grown inside a climatarium with a 12-hour timer for lights. The measurements for plant height began in late November and were taken every three weeks. At the end of the growing period which, was early February the plants were dug up and the roots were measured.

Results

Results for Growth Difference: The difference in growth from the control and the clear containers were significant according to two-tailed t-tests. The p-value was 0.0159 for the difference between the clear and control containers for November was significant, and the p-value for December was also significant at 0.0248. The differences in the last measurements in January were not significant with a p-value of 0.0891. The differences in root sizes were also insignificant at a p-value of 0.518.

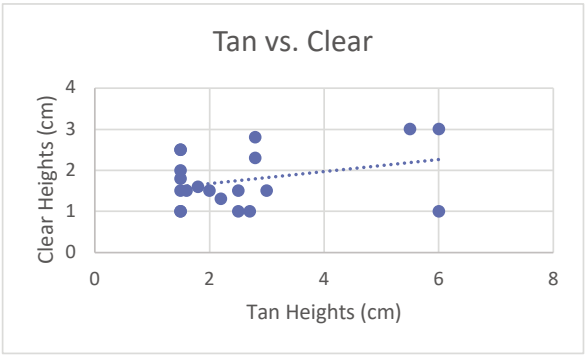


This graph shows a weak correlation with an R of 0.413 between the heights of the strawberries from the control containers and those from the clear experimental containers.

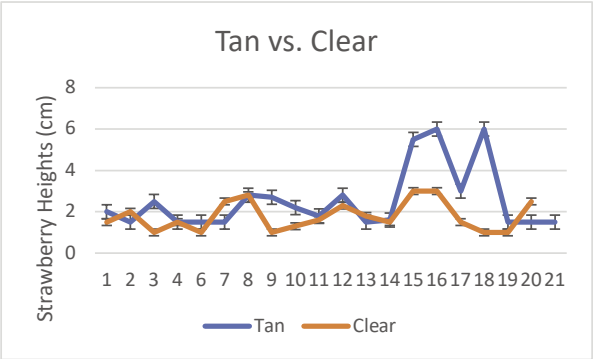


The standard deviation for the clear container plant heights was 0.691 and 0.932 for the control container plant heights.

Results for Difference in Container Materials: There was a noticeable difference between the tan and clear experimental containers. The difference failed to be significant with a p-value of 0.518, but it was enough to keep the tan and clear experimental from being counted together.



This graph shows a weak correlation between the heights of the tan and clear strawberries with an R of 0.317. This shows that while the t-tests found the difference between the two types of plastic containers insignificant, there was some level of a correlation.



The standard deviation for the tan container plant heights was 1.48 and 0.691 for the clear container plant heights.

	Clear vs. Control p-values	Clear vs. Control t-values	Tan vs. Clear p-values	Tan vs. Clear t-values
11/29/2021	0.0159	-2.80	0.311	1.06
12/22/2021	0.0249	-2.56	0.189	1.39
1/12/2022	0.0892	-1.86	0.132	1.63
Root Size	0.518	-0.670	0.622	0.497

The roots were not further analyzed due to the high p-values.

Discussion

The results show that garlic intercropping has a significant effect on the growth of strawberries. This accepts the hypothesis that garlic decreases the growth of strawberries, which goes against garlic being known as the "king of companion plants and rejects the null that garlic does not affect strawberry growth." The last group of data collection likely lost its significance due to the plants beginning to die after December. There was also the accidental discovery of the effects of plastic on the growth of strawberries. The data did turn out to be insignificant, but it still raised questions. This was mostly due to the plants dying in the newer clear plastic containers while living fairly well in the older tan plastic container.

Implications and Recommendations

This research connects to Sustainable Development Goal (SDG) 15 since garlic intercropping is usually used to reduce insecticide use, and this could lead to a deeper investigation on the effects of garlic intercropping. It also goes with SDG 2 since it is an investigation into more efficient farming practices.

Future Research

This research had some fascinating results that could lead to future research. There could be more research into how exactly garlic effects the soil to cause this change in growth. It also raises question for possible research on the long-term effects of growing plants in plastic and whether it was a coincidence or just insignificance due to the shorter growing period. There could also be experiments that are outdoors since most of the benefits are garlic intercropping is protection against insects. There could also be a longer experiment that compares fruit yield too.

References

Attia, S., Grissa, K., Lognay, G., Heuskin, S., Mailleux, A. C., & Hance, T. (2011). Chemical Composition and Acaricidal Properties of Deverra scoparia Essential Oil (Araliales: Apiaceae) and Blends of Its Major Constituents Against Tetranychus urticae (Acari: Tetranychidae). Journal of Economic Entomology.

Attia, S., Grissa, K., Lognay, G., Mailleux, A., Heuskin, S., Mayoufi, S., & Hance, T. (2011). Effective concentrations of garlic distillate (Allium sativum) for the control of Tetranychus urticae (Tetranychidae). Journal of Applied Entomology.

Mischeck, D., & Katsaruware, R. (2014). Onion (Allium cepa) and garlic (Allium sativum) as pest control intercrops in cabbage based intercrop systems in Zimbabwe. IOSR Journal of Agriculture and Veterinary Science.

Singh, S. (2015). Planting row arrangement and nutrient management in geranium (Pelargonium graveolens)-garlic (Allium sativum) intercropping. Indian Journal of Agricultural Research, 49(5), 407-413.

An Analysis on the Impact of Food Insecurity on the Economy

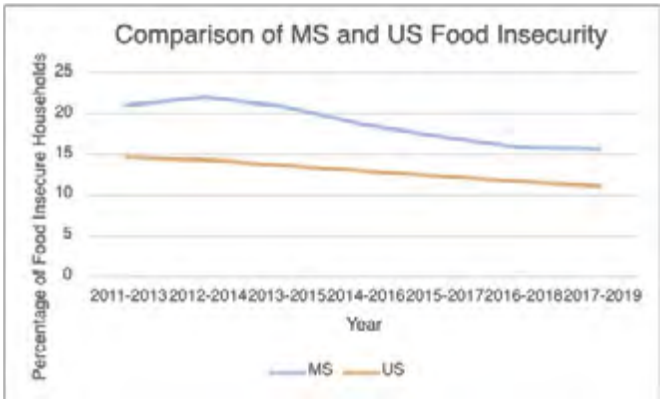
Dia Kher

Abstract

Mississippi is considered to be one of the unhealthiest states in the nation. This health burden and disparity has been found to be associated with food insecurity, the disruption of food intake due to low income. Previous research states that higher nutritional food intake is linked with the Gross Domestic Product (GDP) of a state or country; however, existing research has not assessed the effects or association of nutritional food consumption on GDP over time. Data collected from the U.S. Department of Agriculture from 2016-2021 was utilized to analyze the association between food insecurity and GDP growth. Recent trends show a correlation that suggests high food insecurity is linked to lower GDP of a state. To predict how Mississippi's economy may change as food security and consumption of nutritionally dense food increases, regression models were conducted using yearly household food insecurity, women's education level, excessive drinking among women, and GDP changes over the five years. The results support that there is a predicted inverse relationship between household food insecurity and GDP. As food insecurity decreases, GDP is expected to increase over time. These findings suggest that if food insecurity and hunger-based poverty traps are addressed in Mississippi, individuals will be healthier and more economically productive due to the increased access and consumption of nutritious foods which in turn is likely to produce long terms improvements in Mississippi's economy.

Introduction

The United States is not new to food insecurities (Swann 2017). Faridi and Wadood 2010). Mississippi is known to have the highest food insecurity rate in the United States. Twenty percent of the state's population is considered to be food insecure (about one in four children). The overall food insecurity rate is 22%; 34 out of the 82 counties of Mississippi possess a rate greater than twenty-two percent (Move for Hunger 2018). Studies have been to assess the presence of such food insecurity in Mississippi and its effect on the citizens' health. However, a limited number of studies have been conducted to predict what would occur if Mississippi simply consumed more nutritious food. People do not connect an increase in income with the consumption of nutritious food, keeping people in the cyclic poverty trap (Strauss 1986). The question which then comes to mind: what would happen to Mississippi's economy if Mississippi had a diet involving more nutritional food? The measures of the economy (GDP), excessive drinking, education, and food insecurity will be used to predict in a ten to fifteen year range the change in the economy.



Methodology

The objective of the project is to determine whether there is a relationship between GDP growth and food insecurity.
Question: What would happen to Mississippi's economy if Mississippi had a diet involving more nutritious food?
H_0: As citizens of Mississippi gain the ability to consume more nutritional food in their diets, the Gross Domestic Product (GDP) of Mississippi (MS) will tend to remain unchanged.
H_a: As citizens of Mississippi gain the ability to consume more nutritional food in their diets, the Gross Domestic Product (GDP) of Mississippi (MS) tends to increase.

The data used was collected by a secondary resource. Two data sets used were composed of food insecurity of households in each state and the GDP of each of the 50 states. The data set of food insecurity was collected by the U.S. Department of Agriculture, and the GDP of each state was collected by Our World in Data. The GDP of each state is the dependent variable, and the independent variable is the food insecurity of households in each state, as we are looking at the effect of food insecurity on GDP. No restricting demographics were used to collect the data.

To analyze the data found, we will be using a linear regression to predict the value of GDP based on the value of food insecurity in households. Significance in the linear regression suggests that food insecurity affects the value of GDP. The significant probability indicates that the variability accounted for is not due to random chance. Instead, it can be accounted for by the predicting value.

Results

Correlation Matrix:

Correlation Matrix

		Percentage of Food Insecure Households	Per GDP Growth	Education in Women	Excessive Drinking in Women
Percentage of Food Insecure Households	Pearson's r	-			
	p-value	-			
Per GDP Growth	Pearson's r	-0.291	-		
	p-value	0.039	-		
Education in Women	Pearson's r	-0.230	0.328	-	
	p-value	0.108	0.020	-	
Excessive Drinking in Women	Pearson's r	0.098	-0.195	0.219	-
	p-value	0.499	0.174	0.126	-

The correlation matrix suggests that there is a negative correlation between the average household food insecurity and GDP growth, and a positive correlation between education in women and GDP growth.

Regression Models:

$$Y_i = \alpha + \beta_j X_j + \in_i$$

where β_j is some correlation coefficient for its corresponding variable X_j.

Model 1

X_{ij}	β_j	p-Value
Percentage of Food Insecure Households	-0.234	0.039***

Model 2

X_{ij}	β_j	p-Value
Education in Women	0.590	0.020***

Model 3

X_{ij}	β_j	p-Value
Percentage of Food Insecure Households	-0.156	0.160
Education in Women	0.609	0.018***
Excessive Drinking in Women	-0.442	0.071**

Both Model one and two utilize a univariate regression model against GDP growth with different independent variables. Model three utilizes a multivariate regression with all three independent variables.

Model one illustrates the significance that food insecurity has on GDP growth. We find that the relation between food insecurity and GDP is not due to random chance, rather, as food insecurity decreases, GDP will tend to increase.

Model two illustrates the significance that education has on GDP growth. We find that the relation between education and GDP is not due to random chance, rather as the rates of education increase, GDP will tend to increase as well. From the regression, we see that there is a direct relationship between the two.

Model three illustrates the significance that each variable has on GDP growth. Education and excessive drinking were found to be the strongest predictors out of the three variables; however, this does not discount food insecurity from being a significant predictor.

Discussion

Food insecurity and education become significant facors in GDP growth. The study portrays the findings that education and excessive drinking in women were the greatest predictors of GDP. It is understood that GDP affects the economy with either an increase or decrease in growth. The p-value for education in women was signifcant to the 5 percent level, and excessive drinking was significant at the 10 percent level. Education and excessive drinking tends to directly affect food insecurity, which in turn affects GDP growth.

Implications and Recommendations

As more citizens become aware of the positive effects of consuming more nutritional food, there will be a corresponding reduction of ignorance as more nutritional food is consumed. Utimately, the occurrence of hunger-based poverty traps will decrease in Mississippi and other food-insecure states. This study can aid in reaching the sustainable goals of lower poverty rates and achieving food security. Model one and model two have shown that food insecurity tends to affect GDP growth. With this information, the state of MS can inform their citizens, enhancing the diets of the children.

References

Christopher A. Swann (2017). Household history, SNAP participation, and food insecurity. Elsevier. https://econpapers.repec.org/article/eeejfpoli/v_3a73_3ay_3a2017_3ai_3ac_3ap_3a1-9.htm

John Strauss & Duncan Thomas (1986). Health, Nutrition, and Economic Development. Journal of Economic Literature. https://www.jstor.org/stable/2565122

Unknown (2018). Why is Mississippi the Hungriest State?: Move for Hunger. Move for Hunger. https://moveforhunger.org/why-is-mississippi-the-hungriest-state-in-the-nation

Roser M. and Ortiz-Ospina E. (2019). Global Extreme Poverty. Our World in Data. https://ourworldindata.org/extreme-poverty

Hunger in Mississippi. Feeding America. 2022. https://www.feedingamerica.org/hunger-in-america/mississippi

Health of Women and Children. (2016-2021) Explore Food Insecurity.

Madison Echols

An Analysis of the Effects of Nicotine on Characteristics of *Drosophila melanogaster*

Abstract

Currently, the world suffers from an epidemic: the epidemic of nicotine products. Using the *Drosophila melanogaster* as a model organism, a correlation between nicotine exposure and decreased longevity, impaired locomotor activity, phenotypic changes, and reduced mass was tested. Two *Drosophila* diets were used: a control diet using a standard medium and a nicotine diet. Three trials were tested, with four groups, one for each sex and type combination. This experiment ran for a week and a half. The data was analysed using a one-way ANOVA test. Significant data showed that the survival rate of nicotine-treated flies decreased at a higher rate than that of the control flies. In addition, the mass of the nicotine flies decreased compared to that of the control flies, and locomotor activity in female flies was impaired. Despite the female significance, the male negative geotaxis assay results were not significant, and correlation could not be determined. This study further demonstrates the correlation between behavior, longevity, locomotor ability, and nicotine exposure that is also shown in humans. Further experiments sequencing the flies' DNA and testing varying nicotine concentrations will lend to more clarity regarding the ramifications of nicotine products.

Introduction

Outside of the current COVID-19 pandemic, the world suffers from an epidemic of a different cause: tobacco products. According to the World Health Organization, the tobacco epidemic is one of the biggest public health threats the world has ever faced, killing more than eight million people a year around the world (WHO 2021). Over 80% of the 1.3 billion tobacco users worldwide live in low and middle-income countries; the burden of tobacco-related illness and death is the heaviest there (WHO 2021). Furthermore, the constant tobacco usage in these countries contributes to poverty by redirecting funds for survival and quality of life to tobacco. This tragedy has been exacerbated by tobacco companies, like British American Tobacco, pushing claims that nicotine offers substantial benefits and even protection against COVID-19. Additional research surrounding the health ramifications caused by nicotine exposure can be used to debunk claims driven by the tobacco industry, promote the need for more awareness in less developed countries, and ultimately mitigate the tobacco pandemic. This project aims to provide that research by analysing the effects of nicotine on the longevity, locomotor behavior, phenotype, and mass of *Drosophila melanogaster*. Fruit flies have high similarity to the human genome, with 60% of their genome corresponding (Mirzoyan, 2019). This, compounded with their short generational period and life cycle, low cost to maintain, and availability, make them effective models for this study.

Hypotheses

H_0: There is no correlation between nicotine exposure and decreased longevity, impaired locomotor activity, phenotypic changes, and decreased mass.

H_a: There is a correlation between nicotine exposure and decreased longevity, impaired locomotor activity, phenotypic changes, and decreased mass.

Methodology

Wild-Type *Drosophila melanogaster* were reared in a light-dark controlled Climatarium under standard conditions. The control flies were raised on a rehydrated potato flakes medium, and the nicotine flies were reared on the same ready-made medium. However, during the rehydration process, 300 µL of .1 M nicotine solution was mixed in with the 12 mL water and added to 4.4 g of dry medium.

Using Triethylamine to anesthetize the flies, they were first separated into males and females within one day of eclosion to ensure age synchronicity. Twenty-five flies were placed in each vial, with each trial having a male nicotine vial, male control vial, female nicotine vial, and female control vial. This was repeated three times, making 150 male and 150 female flies in total. Each fly sex remained separate throughout the experiment. Their phenotypes were observed before the experiment to ensure no perceivable mutations.

Figure 1: Negative Geotaxis Test tube made with 2 stacked culture vials.

The negative-geotaxis climbing assay was used to study the difference in locomotor behavior in nicotine-treated flies and control flies. The flies from the different food and sex vials were introduced into their respective negative geotaxis tubes one by one. The flies were observed climbing up the tube, and the number of flies reached the marked orange ring at 10 cm height within 10 sec. The percent climbing activity was calculated using the formula:

$$\%\ Flies\ Climbed = \frac{E}{T} \times 100$$

Where E= Number of flies reaching 10 cm mark in 10 sec, and T= Total number of live flies in the culture tube.

Control and nicotine flies were collected separately, with 20 flies in each group, and weighed on an analytical balance at the end of the experiment to determine changes in adult weight. The value on the balance was then divided by 20 to find the average weight.

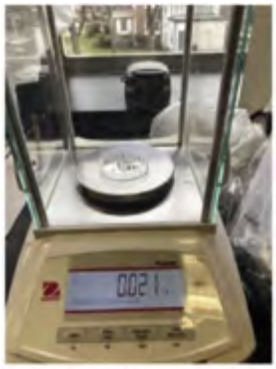


Figure 2: Analytical balance used to find the average weight of each of the fly types.

The number of dead flies was counted every day (not including skipping days from weekends) and used to calculate the percent survival of the flies.

The data was statistically tested using a one-way ANOVA test.

Results

Negative Geotaxis:

The f-ratio value of the percent of female flies that climbed past 10 cm is 6.232. The p-value is .03165. The results show significance at $p < .05$, showing a correlation between exposure to nicotine and an impaired locomotor ability. Regarding the male's Negative Geotaxis Test, though biological significance between impaired climbing ability and nicotine exposure was shown, the f-ratio value is 1.674. The p-value is 0.2100. The results are not significant at $p > .05$, meaning a conclusion of locomotor impairment could not be correlated with nicotine exposure.

Survival Rates:

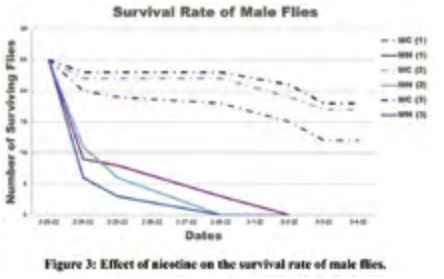


Figure 3: Effect of nicotine on the survival rate of male flies.

The survival rates of male nicotine treated flies were significantly less than that of the control flies. The f-ratio value is 45.25. The p-value is < .00001. The result is significant at $p < .05$. The same can be said about the female flies with the survival rates of nicotine treated flies were significantly less than that of the control flies.

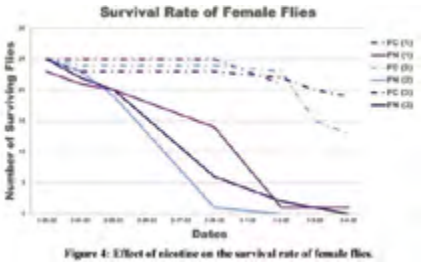


Figure 4: Effect of nicotine on the survival rate of female flies.

Mass and Phenotypic Changes:

Male Nicotine Average Weight	Male Control Average Weight	Female Nicotine Average Weight	Female Control Average Weight
3.50×10^{-4} g	8.50×10^{-4} g	3.50×10^{-4} g	1.15×10^{-3} g

Figure 5: Effect of nicotine on the mass of male and female flies.

The weighing of each fly group showed a significant difference between the control flies and the nicotine flies, with the weights decreasing in the nicotine flies.

The mass changes are further backed up by the phenotypic changes seen in the flies themselves. In both males and females, the nicotine-treated flies shown are smaller in mass than the control fly. The abdomens appear thinner as well as the heads.

Figure 6: Effect of nicotine on the phenotype of male flies.

Figure 7: Effect of nicotine on the phenotype of female flies.

Discussion

Significant data showed that the survival rate of nicotine-treated flies decreased at a higher rate than that of the control flies, the mass of the nicotine flies decreased compared to that of the control flies, and locomotor activity in female flies was impaired. Given the significance of the results, the nulls were rejected and the alternates were accepted for all but the locomotor activity. Despite the female significance, the results of the male negative geotaxis assay were only biologically significant but not statistically significant, and therefore the null altogether was accepted.

This study further demonstrates the correlation between behavior, longevity, locomotor ability, and nicotine exposure that is also shown in humans. Further experiments sequencing the flies' DNA and testing varying nicotine concentrations will lend to more clarity regarding the ramifications of nicotine products.

Sustainable Goal Number Three

The Sustainable Goal of this project is Good Health and Well-Being, which promotes healthy living and well-being for all. By gaining further information on the potential effects of nicotine, this project could turn people away from using tobacco products, mitigating poverty through the reduced usage of nicotine, leading to a healthier world overall.

Future Research and Implications

There are future plans for research regarding the sequencing of DNA of the control flies and nicotine flies to understand better the genetic mutations of using nicotine products such as E-Cigarettes.

Reliable Equipment / More Experience

Due to inexperience, the flies were not accurately sorted, giving rise to pupae before the experiment was over, which led to limitations regarding the longevity assay. In addition, the use of non-standard medium and Triethylamine, which are not as reliable in anesthetization compared to CO_2, could have varied the results in the experiment. Going forward, the usage of materials standard to *D. melanogaster* experimentation will provide more accurate results.

References

Chandran, G. Negative geotaxis or climbing assay for assessment of locomotor function in adult *Drosophila* (2021). Experiments with *Drosophila* for Biology Courses. (Eds: S.C. Lakhotia & H.A. Ranganath). (pp. 471-475).

Mirzoyan, Z., Sollazzo, M., Allocca, M., Valenza, A. M., Grifoni, D., & Bellosta, P. (2019). Drosophila melanogaster: A Model Organism to Study Cancer. Frontiers in genetics, 10, 51. https://doi.org/10.3389/fgene.2019.00051

Velazquez-Ulloa NA. (2017) A *Drosophila* model for developmental nicotine exposure. PLoS ONE 12(5): e0177710. https://doi.org/10.1371/ journal.pone.0177710

World Heath Organization (2021). Tobacco. Tobacco. (2021). Retrieved 12 February 2022, from https://www.who.int/news-room/fact-sheets/detail/tobacco.

NOTES

NOTES

NOTES

SCIENCE JOURNAL ENTRIES

Name/County	Competition and Awards
Hayden Anderson/Lowndes	Region V Best of Fair ISEF Alternate; 2nd Place Biomedical and Health Sciences; United States Air Force for outstanding Science or Engineering project; 3rd Place Biomedical and Health Sciences State
Maggie Buck/Tate	Region VII 1st Place Plant Sciences; 3rd Place Plant Sciences State
Nicholas Djedjos/Rankin	Region II Best of Fair ISEF Recipient; 1st Place Biomedical and Health Sciences; 2nd Place Junior Science/Humanities Symposium $1500; 1st Place Biomedical and Health Sciences State
Madison Echols/Lamar	Region I Best of Fair ISEF Alternate; 1st Place Animal Sciences; 1st Place Animal Sciences State
Raeed Kabir/Harrison	Region VI Best of Fair ISEF Recipient; 1st Place Behaviorial Sciences; 1st Place Junior Science/Humanities Symposium $2000; 1st Place Behavioral and Social Sciences State
Dia Kher/Harrison	Region VI Best of Fair ISEF Recipient State; 2nd Place Behaviorial Science; 1st Place Behavioral and Social Sciences (Regional)
Stephanie Ressel/Jackson	Region VI Best of Fair ISEF Recipient; 1st Place Physics and Astronomy; American Meteorological Society Award; NOAA 2022 Taking the Pulse of the Planet Award; 1st Place Physics and Astronomy State
Hailee Sexton/Harrison	Region VI Best of Fair ISEF Recipient; 1st Place Physics and Astronomy; American Meteorological Society Award; NOAA 2022 Taking the Pulse of the Planet Award; 1st Place Physics and Astronomy State
Qiancheng Sun/Oktibbeha	Region V Best of Fair ISEF Alternate; 1st Place Inorganic Chemistry; Yale Science and Engineering Award; 1st Place Inorganic Chemistry State
Jessica Yan/Oktibbeha	Region V Science Talent Search Top 300 $2000; 1st Place Environmental Sciences, $100 National Geographic Award; NOAA and RICOH Sustainability Awards; United States agency for International Development certificate; Stockholm Junior Water Prize; 3rd Place Junior Science/Humanities Symposium $1000; 1st Place Earth and Environmental Sciences State
Andrew Yu/Oktibbeha	Region V Best of Fair ISEF Recipient; 1st Place Biomedical and Health Sciences; Mu Alpha Theta award; 2nd Place Biomedical and Health Sciences State

MSMS Science Journal
Waves of Science

The MSMS Science Journal is a collection of research projects conducted and defended by the students at The Mississippi School for Mathematics and Science
1100 College Street, MUW-1627
Columbus, Mississippi 39701
The MSMS Science Journal is available to read on our website at www.themsms.org