

Basic Information

Last updated: 08/24/2018

Please complete eligibility and basic demographic information.

Eligibility

Select the option that best describes you:

I am a high school senior living in the United States with an expected graduation date of Winter 2018 or Spring 2019 (eligible to apply)

Basic Student Information

Your Legal Name

First Name on Your Transcript	Neeyanth
Middle Name on Your Transcript (optional)	(No response)
Last Name on Your Transcript	Kopparapu

Alternate/Nickname for Your (No response)

Biography

What do you like to be called?

Gender (optional) Male

Parent/Guardian #1

First Name	Rajani
Last Name	Kopparapu
Relationship to You	Mother

Parent/Guardian #2 (optional)

First Name	Madhu
Last Name	Kopparapu
Relationship to You	Father

Regeneron Science Talent Search Rules Confirmation:

Read the Regeneron Science Talent Search [Rules and Eligibility 2019](#) document.

This is a required step in order for you to understand the eligibility requirements, rules regarding human and animal research and more.

I have read, understand, and agree to abide by the Regeneron STS Rules.

Recommender Requests

Last updated: 08/24/2018

You are responsible for requesting three types of recommendations. The linked instructional documents will help you determine the best person to ask to complete each type. The due date for recommendations is November 7, 2018 (one week prior to your application deadline). Visit our [Application FAQ](#) for tips and to learn more about this process.

Read the instructional documents for each recommendation type below, then click the checkbox to acknowledge that you understand this process. The ability to request recommendations will then appear in your Task List on the internal home page. You may also request a second recommendation for each type (optional); these extra recommendation tasks will appear at the bottom of your Task List.

- [1. Educator Recommendation Instructions](#)**
- [2. Project Recommendation Instructions](#)**
- [3. High School Report Instructions](#)**

Click the box below to acknowledge your understanding of this requirement.

I have read the documents above and understand that it is my responsibility to request three types of recommendations through the online application system. I will ask my recommenders to complete the required documents by the due date, November 7, 2018.

I agree

High School Information

Created: 08/24/2018 • Last updated: 08/29/2018

Tell us about your school.

Current High School Name and College Board Code

Select your school from the list below. You may search by your high school's name, CEEB code, zip code, city or state.

Note: If your school is not listed here, please email sts@societyforscience.org. We are happy to add your school to this list (could take up to 2 business days). Do not select a school you do not actually attend. We typically add 100+ schools to this list each year.

470054,Thomas Jefferson High School for Science and Technology,Alexandria,VA,22312

Where is your high school located?

I attend school in the United States or in a U.S. territory

High School Address

Confirm Name of High School	Thomas Jefferson High School for Science and Technology
High School Address	6560 Braddock Rd
High School Address 2	(No response)
High School City	Alexandria
High School State	VA
High School Zip Code	22312
High School County	Fairfax County

Type of School:

Public

Please select all that apply.

Magnet School

Current Coursework

In what classes are you currently enrolled? Enter up to eight. (Grades are not needed)

1	AP Physics M&EM
2	AP Economics (Micro/Macro)
3	English 11
4	AP United States History
5	Energy Systems
6	Russian 3
7	Complex Variables / Differential Equations
8	(No response)

Activities, Interests and Awards

Created: 08/24/2018 • Last updated: 11/14/2018

Tell us about your activities, interests and awards. Do not feel pressured to meet the word limits for essay questions.

SCIENCE COMPETITIONS AND PROGRAMS

Have you participated in any of the following?

Select all that apply.

Science fair at the local or regional level

Intel International Science and Engineering Fair (Intel ISEF)

Finalist

Other science competitions

Tell us about your Intel ISEF participation:

Year(s) of participation as a finalist	2017
Award(s) won	<p>Samvid Education Foundation: Agni Second Place Award of \$500</p> <p>Sigma Xi, The Scientific Research Honor Society: Second Life Science Award of \$1,000</p> <p>Third Award of \$1,000</p> <p>Association for the Advancement of Artificial Intelligence: Honorable Mention</p>

EXTRACURRICULAR ACTIVITIES, VOLUNTEERISM & EMPLOYMENT

Tell us about your top 10 extracurricular activities, volunteer work, employment, etc. We want to know more about you, your day-to-day life, how you spend your time, and why. Please select an activity type and describe your involvement in the activity. If you do not see an exact match for your activity in the list, select "other." Please do not feel pressured to use all ten activity boxes below.

How many activities would you like to list? 10

Activity #1

Activity Type:	Academic Club
Name of Activity (type of sport, name of club, etc):	Varsity Math Team
Your Role/Leadership Roles:	2018 - Co-Captain 2017 - Finance Officer
How many years have you participated in this activity?	4
Explain the time commitment to this activity:	5-10 hours per week
Awards won/accomplishments related to this activity	<p>Some of our biggest awards from last year include:</p> <p>2018 ARML - 1st Place Overall, 1st Place Team Round, 1st Place Relay Round, 2nd Place Power Round.</p> <p>2018 Harvard MIT Math Tournament - 2nd Place Sweepstakes, 4th Place Aggregate</p> <p>2017 Princeton University Math Contest - 3rd Place Team Round, 3rd Place Power Round, 2nd Place Overall</p>
Anything else you would like to share about this activity? (75 words maximum)	<p>Largest Club at TJHSST - 120 students per week attending club and sending over 100 students to competitions.</p> <p>VMT sponsors TJIMO, a middle-school math competition hosted at TJ for STEM-oriented students. VMT members write problems, proctor contests, and teach students concepts related to competition math.</p> <p>As the youngest officer and one of the captain for two years, I've learned about leadership and organization to successfully help run the largest club at a STEM oriented school.</p>

Activity #2

Activity Type:	STEM-related Club
Name of Activity (type of sport, name of club, etc):	Computer Team
Your Role/Leadership Roles:	2018 - Captain
How many years have you participated in this activity?	3
Explain the time commitment to this activity:	3-5 hours per week
Awards won/accomplishments related to this activity	As a part of the Computer Team, I placed 2nd at the ACM@UVA High School Programming Contest 2018, and 3rd at Virginia-Tech's HSPC 2017 and 2018. As a USACO Platinum Competitor since 2016, I am very knowledgeable in the field of algorithmic computer science and use that knowledge to teach students at my high school.
Anything else you would like to share about this activity? (75 words maximum)	The Computer Team is the largest computer science club at Thomas Jefferson. Made up of the Freshmen, Intermediate, and Senior teams, the team is responsible for helping over 180 students every week with various aspects of computer science. Functional programming, algorithmic development, and theoretical computer science are few among the many skills taught. The teams participate at various HSPC contests, ACSL, and other CS contests. As a captain, I manage competition logistics and lecture weekly.

Activity #3

Activity Type:	Social Justice
Name of Activity (type of sport, name of club, etc):	Centerville Commission for Labor Justice
Your Role/Leadership Roles:	Part of the 5-member team
How many years have you participated in this activity?	1
Explain the time commitment to this activity:	2 hours per week
Awards won/accomplishments related to this activity	N/A
Anything else you would like to share about this activity? (75 words maximum)	The Centerville Commission for Labor Justice (CCLJ) is responsible for employing hundreds of day laborers and matching them with work that needs to be done in the community. It has a heavier burden of ensuring wages are paid to workers. Out of the 141,098 dollars in wage theft claims, the CCLJ has been able to recover \$26,159. We work weekly to resolve hundreds of claims that come every month.

Activity #4

Activity Type:	STEM-related Club
Name of Activity (type of sport, name of club, etc):	Bioinformatics Club
Your Role/Leadership Roles:	BioCode Chair
How many years have you participated in this activity?	3
Explain the time commitment to this activity:	1-2 hours per week
Awards won/accomplishments related to this activity	N/A
Anything else you would like to share about this activity? (75 words maximum)	BioCode is a algorithmic programming contest that the Bioinformatics Club hosts. For the past 2 years, over 250 students attended the contest, and learned about the intersection of biology and computer science through informative lectures, problems, and keynote speakers. As a chair and the sole problem writer, I balance the curriculum with easy and challenging problems, ensuring the right mix is able to challenge students without dissuading them.

Activity #5

Activity Type:	Scouting
Name of Activity (type of sport, name of club, etc):	Boy Scouts
Your Role/Leadership Roles:	Troop Leader 2016
How many years have you participated in this activity?	5
Explain the time commitment to this activity:	1 hour per week
Awards won/accomplishments related to this activity	Life Scout, Order of the Arrow member
Anything else you would like to share about this activity? (75 words maximum)	(No response)

Activity #6

Activity Type:	Foreign Language
Name of Activity (type of sport, name of club, etc):	Russian Honor Society
Your Role/Leadership Roles:	SLAVA Member
How many years have you participated in this activity?	1
Explain the time commitment to this activity:	1-2 hours per week
Awards won/accomplishments related to this activity	2016 Russian Essay Contest Honorable Mention 2017 Russian Essay Contest Gold Medal
Anything else you would like to share about this activity? (75 words maximum)	The Russian Honor Society consists of around 30 students. Together, we work in over 12 activities that benefit the school and the community around us. One of the biggest activities includes the annual Election day bake sale, where we encourage families to come out and vote near Thomas Jefferson by hosting a booth where people can get coffee, hot chocolate and other baked goods from 5:30 AM through 8 PM.

Activity #7

Activity Type:	Research
Name of Activity (type of sport, name of club, etc):	Parkinson's Research
Your Role/Leadership Roles:	Researcher
How many years have you participated in this activity?	1
Explain the time commitment to this activity:	3-5 hours per week
Awards won/accomplishments related to this activity	N/A
Anything else you would like to share about this activity? (75 words maximum)	Since early August, I have been engaged in examining various potential diagnosis methods for Parkinson's Disease.

Activity #8

Activity Type:	Volunteer Work
Name of Activity (type of sport, name of club, etc):	GirlsComputingLeague
Your Role/Leadership Roles:	Cofounder, GirlsComputingLeague
How many years have you participated in this activity?	3
Explain the time commitment to this activity:	3-5 hours per week
Awards won/accomplishments related to this activity	N/A
Anything else you would like to share about this activity? (75 words maximum)	GirlsComputingLeague is a 501(c)3 nonprofit working towards empowering girls and other underrepresented groups in the field of Computer Science with access to a high quality education. We host workshops in partnership with DC Housing Authority and the White House's CS4ALL initiative, and we work with the Tiger Woods Foundation to teach teachers how to integrate computer science in their curriculum. We also host an annual AISummit for 350 students, raising over \$100K in 2 years.

Activity #9

Activity Type:	Volunteer Work
Name of Activity (type of sport, name of club, etc):	TJSTAR
Your Role/Leadership Roles:	Lead, Organizing Committee
How many years have you participated in this activity?	2
Explain the time commitment to this activity:	1 hour/week
Awards won/accomplishments related to this activity	N/A
Anything else you would like to share about this activity? (75 words maximum)	TJSTAR is the annual research symposium for Thomas Jefferson High School. Students present their research, Companies come to talk about research opportunities, and researchers, economists, executives, and engineers come to talk about their work. I work to inviting speakers, raising money from our sponsor organizations, and lay out the plan for the day, including keynote speakers, room assignments, and more.

Activity #10

Activity Type:	Work Experience
Name of Activity (type of sport, name of club, etc):	Acefolios
Your Role/Leadership Roles:	Founder, CEO
How many years have you participated in this activity?	3
Explain the time commitment to this activity:	2-5 hours per week
Awards won/accomplishments related to this activity	N/A
Anything else you would like to share about this activity? (75 words maximum)	Acefolios is a consumer portal website for creating and presenting electronic portfolios, which supports all types of multimedia content and easily shares to social and professional network sites. Uses PHP, MySQL, Apache, and other web technologies including AJAX, JavaScript, jQuery, YUI (Yahoo User Interface) library, and Amazon AWS. Currently in talks with Blackboard to make Acefolios a service for their tens of thousands of students.

Summer Activities

What did you do in the last 3 summers? Tell us about any jobs, activities, research, travel, etc. (Maximum 300 words)

Last summer, as part of the Emerging Technologies Student Program, I was a paid Deep Learning intern at the MITRE Corporation working with the DEEPLANG group to decode EEG recordings into human-readable thoughts. I used various deep learning frameworks and data preprocessing libraries, and received IRB approval from the MITRE IRB committee to conduct experimentation in the MITRE compound. After the completion of my internship, I began my research in the automatic diagnosis of Parkinson's Disease. As one of the captains of the Thomas Jefferson Varsity Math Team, I also worked on writing problems and organizing the Thomas Jefferson Intermediate Math Open - a math competition the Math Team hosts for middle school students.

The summer of 2017 I attended MIT Launch, a summer program learning from industry experts on how to launch a successful startup, from idea, to minimum viable product, to fundraising. While at MIT, I was in contact with many professors, research fellows and students at MIT and Harvard, including postdoctoral fellow Adrian Dalca, who later helped me with my current project. I also began my research into screening for Major Depressive Disorder through social media.

The summer before my freshman year, I spent most of my time at school, taking Research Statistics as a summer course. I spent 4 weeks studying the fundamentals of applying statistics to various fields, including finance and research. The course helped me analyze data that I collected in all three of my projects. I also attended the AwesomeMath Summer Program which helped me sharpen my competitive math skills in the fields of Algebra and Geometry. I also began work on the Eyeagnosis project nearing the end of the summer.

AWARDS

List special recognitions, awards, honors and scholarships from both school and community, if any.

Include national or international honors received during your high school career not previously listed in the activities section above. (Maximum 250 words)

Apart from research awards, I have won many honors in the fields of mathematics and computer science. I have qualified for the USA Junior Math Olympiad in 2015 and 2017, as well as for the USA Math Olympiad in 2018. As a part of the TJHSST Varsity Math Team, I have placed as the first team at ARML, CMIMC, and the Duke Math Meet in 2018 and 2017. I also placed in the top 3 at the Harvard-MIT Math Tournament (2018) and the Princeton University Math Contest (2017). As an individual, I placed in the top 10 in 2016 and 2017 CMIMC in the field of Number Theory.

At the end of the 2016 season of the United States Computing Olympiad (USACO), I ended in the Platinum Division. Throughout the 2017 season, I placed in the top 150 internationally out of high school and college students, placing in the top 80 in the January contest. I also placed 3rd at the 2018 Virginia Tech ACM-HSPC contest.

I have won many awards for my work in the fields of medical informatics. For Eyeagnosis, I was a 2016 Siemens Competition Semifinalist, and placed 3rd in category of Translational Medicine Sciences at ISEF, winning awards from the AAAI as well. For my work in TweePression I became an MIT INSPIRE Semifinalist and was invited to speak at O'Reilly's AI Conference in New York and NVIDIA's GTC Silicon Valley in the spring of 2018. I was invited to the 2017 Nobel Lectures in Stockholm.

INTERESTS

Future Plans

What do you plan to study in post-secondary education and what occupation do you plan to pursue? What do you hope to be doing 10 years from now? (Maximum 200 words)

I plan to pursue an education in the crossroads of artificial intelligence and computational biology. I hope to attend a research institution where I would continue applying artificial intelligence to solve public health problems, like my work with my Eyeagnosis and TweePression projects. Ultimately, my goal is to pursue a PhD to continue developing the fundamentals so I am able develop creative solutions to problems people face on a daily basis.

More generally, however, I hope to have a career where not only would I build these models and platforms, but I would see to that they would be used in the clinical settings. I've learned that the best medical devices aren't the ones with the highest accuracy, but the ones that are the most accessible and can make an impact for millions. When developing a lens attachment system for the Eyeagnosis project, I recognized my love for developing these technical applications to augment the software.

As the cofounder of GirlsComputingLeague, I've also recognized the importance of educating the next generation. I hope to continue advocating for a high quality STEM education for underprivileged students by actually making a difference, and participating in activities that would benefit all of society.

What single accomplishment are you most proud of and why?

(Maximum 200 words)

I have always prized the journey of learning as much as the award or honor at the end of the road. When learning about algorithmic computer science while preparing for the United States Computing Olympiad (USACO), I was fascinated by the intersection of mathematics, computer science, and critical thinking; however, I would spend hours every day memorizing concepts thinking rote memorization was the best way to do well in these contests.

After I became a platinum competitor at the end of my freshmen year, I quickly realized that memorization of algorithms wasn't enough. I began to appreciate the value of thinking through creative solutions and, with my new approach, was able to crack the top 100 international competitors. I was most proud of the solution I created, one that combined knowledge of algorithmics with a creative method of optimization. Out of thousands of competitors' answers, my solution was unique in its creative methodology. This competition was the first to teach me the power of creativity in seemingly analytic tasks, and I have used this lesson in my research and other endeavors ever since.

Share a fun fact about yourself that we might otherwise not know about you.

(75 words max)

At 12, I took my first flying lesson in a Cessna 350, initially intrigued by the forces of thrust and drag and the complex airfoils that generate lift. After the first jittery rides, I began to experience the flight beyond the science. After all, traveling 100 miles just to end up in the same exact spot might seem silly, but I love the 120 minutes of distraction-free time to appreciate the bird's-eye view of nature.

Task 5: Test Scores (optional)

Created: 08/29/2018 • Last updated: 11/03/2018

Sharing your standardized test scores is optional. Enter your scores in this task, then upload evidence that supports your scores in the next task.

SAT Scores

	Score	Test Date (MM/DD/YYYY)
Composite	1530	
Section Score: Evidence-based Reading and Writing	730	
Section Score: Math	800	
Cross-Test: Analysis in Science	38	
Cross-Test: Analysis in History/Social Studies	38	
Test Score: Reading	36	
Test Score: Writing and Language	37	
Test Score: Math	40	

SAT II

(Science, Math, Engineering, and Technology tests only) List top 4 only

	Subject	Score	Test Date (MM/DD/YYYY)
SAT II Test #1	Math II	800	05/05/2018
SAT II Test #2	Chemistry	770	05/05/2018
SAT II Test #3			
SAT II Test #4			

ACT Scores

List highest score in each category on any date.

	Score	Test Date (MM/DD/YYYY)
English		
Reading		
Math		
Science		
Composite		

Advanced Placement Course Test Scores

(Science, Math, Engineering, and Technology tests first) List top 8 only.

	Subject	Score	Test Date (MM/DD/YYYY)
AP Test #1	Computer Sciecne	5	
AP Test #2	Calculus BC	5	
AP Test #3	Biology	5	
AP Test #4	Chemistry	4	
AP Test #5			
AP Test #6			
AP Test #7			
AP Test #8			

US Territories: Other Tests

For students attending accredited overseas secondary schools only. If you do not attend school in the US and take different tests than those listed above, please report the test names and your scores here.

Do not write in this box if you live in the US and/or take the tests listed above.

(No response)

You will need to upload proof of the test scores you have mentioned above in Task 6 (next task).

Please merge your various test score documents into one PDF document for the upload. Should you need assistance, email sts@societyforscience.org.

SAT Score Report

Test Date: **Aug. 25, 2018**
 Registration Number: **0065324754**

Sex: **MALE**
 Date of Birth: **Feb. 02, 2002**

Test Center Number: **47775**
 CB Student ID: **89923838**
 High School Code: **470054**

High School Name: **Thomas Jefferson High School for Science and Tech**

Neeyanth Kopparapu
3123 Oxford Forest Dr
Herndon, VA 20171 - 1942

Am I on Track for College?

Look for the green, yellow, or red symbols next to your section scores. They let you know if your scores are at or above the benchmark scores. Benchmarks show college readiness. If you see green, you're on track to be ready for college when you graduate.

If you score below the benchmark, you can use the feedback and tips in your report to get back on track.

Your Total Score

1530 | 400–
1600

99th

Nationally Representative Sample Percentile

99th

SAT User Percentile

Essay Scores

7 | 2 to 8
 Reading

6 | 2 to 8
 Analysis

7 | 2 to 8
 Writing

Section Scores

730 | 200–800
 Your Evidence-Based Reading and Writing Score

99th Nationally Representative Sample Percentile
97th SAT User Percentile



800 | 200–800
 Your Math Score

99th Nationally Representative Sample Percentile
99th SAT User Percentile



Test Scores

36 | 10–40
 Reading

37 | 10–40
 Writing and Language

40.0 | 10–40
 Math

Cross-Test Scores | 10–40

38
 Analysis in History/Social Studies

38
 Analysis in Science

Subscores | 1–15

14
 Command of Evidence

14
 Words in Context

14
 Expression of Ideas

14
 Standard English Conventions

15
 Heart of Algebra

15
 Problem Solving and Data Analysis

15
 Passport to Advanced Math

Get your full report online at sat.org/scorereport

How Do My Scores Compare?

A percentile shows how you scored, compared to other students. It's a number between 1 and 99 and represents the percentage of students whose scores are equal to or below yours.

For example, if your Math percentile is 57, that means 57% of test takers have Math scores equal to or below yours.

The Nationally Representative Sample Percentile compares your score to the scores of typical U.S. students.

SAT® User Percentile compares your score to the scores of students who typically take the test.

How Can I Improve?

To see which skills are your strongest and what you can do to boost your college readiness, go to your full report online and look for Skills Insight™.

What Are Score Ranges?

Test scores are single snapshots in time—if you took the SAT once a week for a month, your scores would vary.

That's why score ranges are better representations of your true ability. They show how much your score can change with repeated testing, even if your skill level remains the same.

Colleges know this, and they get score ranges along with scores so they can consider scores in context.

SAT Summary of Scores

Date	Aug. 25, 2018					
Grade	11					
SAT						
Total	1530					
Evidence-Based Reading and Writing	730					
Math	800					
Essay	Reading	7				
	Analysis	6				
	Writing	7				
Old SAT						
Critical Reading						
Mathematics						
Writing						

SAT Subject Test Scores

Date	May 05, 2018					
Grade	10					
Subject Test						
	Chemistry					
Test Score	770					
Language Subscores	Reading					
	Listening					
	Usage					
Subject Test						
	Math Level 2					
Test Score	800					
Subject Test						
Test Score						

* Scores from the SAT Subject Test in Mathematics aren't comparable to Math section, test, and related subscores on the SAT.

*Not all SAT Subject Tests™ have subscores.

How Do I Send My Scores to Colleges?

This student score report is for your use only.

Most colleges require you to have the College Board send them official score reports. They don't accept copies of student score reports, online score reports, or score report labels on transcripts.

Can I Choose Which Scores to Send?

With Score Choice™, you decide which scores you send to colleges. Choose by test date for the SAT and individual test for SAT Subject Tests™. Just make sure you follow each college's stated score-use policy.

What's Next?

Go to sat.org/scorereport and choose your next steps:



SAT®

SAT Subject Tests

Score Report

Test Date: **May 05, 2018**
Registration Number: **0064163998**

Sex: **MALE**
Date of Birth: **Feb. 02, 2002**
Test Center Number: **47179**
CB Student ID: **89923838**
High School Code: **470054**
High School Name: **Thomas Jefferson High
School for Science and
Tech**

Neeyanth Kopparapu
3123 Oxford Forest Dr
Herndon, VA 20171 - 1942

Math Level 2

800 | 200 to
800

79th Percentile

How Do My Scores Compare?

A percentile shows how you scored, compared to other students. It's a number between 1 and 99 and represents the percentage of students whose scores are below yours.

For example, if your Biology Test score is 500 and the national percentile for 500 is 47, you did better than 47% of all high school students who took this test.

Keep in mind that different groups of students take different SAT Subject Tests™, so you can't compare percentiles of different tests.

What Are Score Ranges?

Test scores are single snapshots in time—if you took the test once a week for a month, your scores would vary.

That's why score ranges are better representations of your true ability. They show how much your score can change with repeated testing, even if your skill level remains the same.

Usually your scores fall in a range of roughly 30–40 points above or below your true ability. There must be a difference of at least 60 points between your score and another student's to say that one of you performed better than the other. Colleges know this, and they get score ranges along with scores so they can consider scores in context.

Your online score report at sat.org/scorereport shows your score range for each test you took.

Should I Take the Test Again?

Each Subject Test measures your knowledge of a particular subject. If you continue to study the subject and take the test again, your new score should reflect your increased knowledge.

You can also register for other Subject Tests, or the SAT, at collegeboard.org/mysat.

Chemistry

770 | 200 to
800

80th Percentile

SAT Summary of Scores

Date						
Grade						
SAT						
Total						
Evidence-Based Reading and Writing						
Math						
Essay	Reading					
	Analysis					
	Writing					
Old SAT						
Critical Reading						
Mathematics						
Writing						

SAT Subject Test Scores

Date	May 05, 2018					
Grade	10					
Subject Test						
	Math Level 2					
Test Score	800					
Language Subscores	Reading					
	Listening					
	Usage					
Subject Test						
	Chemistry					
Test Score	770					
Subject Test						
Test Score						

* Scores from the SAT Subject Test in Mathematics aren't comparable to Math section, test, and related subscores on the SAT.

*Not all SAT Subject Tests have subscores.

How Do I Send My Scores to Colleges?

This student score report is for your use only.

Most colleges require you to have the College Board send them official score reports. They don't accept copies of student score reports, online score reports, or score report labels on transcripts.

Can I Choose Which Scores to Send?

With Score Choice™, you decide which scores you send to colleges. Choose by test date for the SAT and individual test for SAT Subject Tests. Just make sure you follow each college's stated score-use policy.

What's Next?

Go to sat.org/scorereport and choose your next steps:

Student Score Report

AP®

CollegeBoard

Report Date: 11/14/18

Neeyanth Kopparapu
3123 Oxford Forest Dr
Herndon, VA 20171

AP Number: 34482489

School: 470054 - THOMAS JEFFERSON HIGH
SCHOOL OF SCIENCE AND TECHNO

Year taken	Name of exam	Score
2018	Chemistry	4
2017	Biology	5
2017	Calculus BC	5
	- Subscore: Calculus BC: AB Subscore	5
2016	Computer Science A	5

Awards

2018	AP Scholar with Honor
2017	AP Scholar

Essay Questions

Created: 08/24/2018 • Last updated: 11/14/2018

Tell us about your project, inspiration, and impact. Do not feel pressured to meet the word limit for each question.

1. Research Project “Layperson’s Summary” (maximum 200 words)

Summarize your project in layperson’s terms, while maintaining scientific accuracy. Your explanation should be easily understandable and include background, procedures, conclusions and relevance. This summary will aid readers, including evaluators, journalists and the public.

Parkinson’s disease is the second most common neurodegenerative disorder and affects millions of patients across the world. Unfortunately, due to the lack of concrete, objective diagnostic tools, the disease is not diagnosed until its later, irreversible stages. Currently, doctors have to comb through years of medical data looking for symptoms to diagnose Parkinson’s.

My project, PDGAN, presents the first automatic diagnosis system for Parkinson’s disease from an MRI scan. Using a training dataset of patient images, PDGAN reached a final accuracy of 96.6% with a system of neural networks that specialize in image recognition. This final accuracy is 16% higher than current clinical methods of diagnosis.

Although the high accuracy and quick diagnosis is notable, the novelty of the project comes from the system’s ability to synthetically generate 3-dimensional MRI images. These synthetic images increase the size of the training dataset and, therefore, the accuracy of the system.

PDGAN is quick, inexpensive, and accurate -- allowing for efficient and accessible screening for Parkinson’s.

2. Project Inspiration (maximum 200 words)

What inspired you to conduct this research project? (What is the story behind your research topic?)

The first question from a patient when diagnosed with a disease is, "So what is the treatment?" And the worst answer, the answer that is given all too often, is, "At this stage, there is none."

This was my family's reality earlier this year when my grandfather was diagnosed with Parkinson's disease, one of the most common disorders for patients above 65. I quickly realized that our family's story isn't unique, and millions of others face the same situation, diagnosed with a disease too late for effective treatment.

I wanted to address this issue of late diagnosis, so I started reading about Parkinson's and the current challenges. I realized the treatment pipeline is retroactive, as doctors must wait for symptoms to develop to diagnose the disease, which leads to significant progression of the disease without any treatment. Instead, I pursued a proactive approach, one that prioritizes automation and presymptomatic detection.

With my past work and knowledge in the field of computational pathology, I believed that using artificial intelligence to aid in the diagnosis of Parkinson's Disease could be the next step in a diagnosis pipeline that would be cheap, affordable, quick, and, most importantly, accessible to everyone.

3. STEM Interests (maximum 250 words)

After completing your research, how has your interest in science, engineering, and/or math been clarified?

Conducting and completing my research has strengthened my interest in science and technology by showing me the ability of the scientific community to impact me. When I was looking for ways to solve the problems I encountered in my project, I looked to this community to help me with my problems. I conversed with members of the USC regarding the use of their data, as well as with Dr. Adrian Dalca, a postdoctoral fellow at MIT, who underscored the importance of exploring creative solutions to solve underlying problems. The interactions I had reinforced my love for science because it showed me the willingness of the community to help other scientists.

I have also been able to present my findings at research symposiums. Attending NVIDIA's GTC in Silicon Valley for a poster presentation about my research showed me the emerging advances in GPU hardware and its applications in parallel computing and artificial intelligence. Giving a talk at O'Reilly's AI Conference gave me the opportunity to see presentations on the forefront of deep learning. Looking at company booths, attending research presentations, and viewing product demonstrations makes me excited about the future of the field.

Overall, my research showed me the importance of solving problems in different ways. The improved accuracy of diagnosis using synthetically generated images confirmed for me the power of creative thinking in the field of artificial intelligence.

4. Project Benefits and Impact (maximum 300 words)

What benefits do you think your research will bring to the world, and/or to your field? What additional steps, and by whom, might be needed for this benefit to be realized?

A common impedance to the advancement of medical research is the lack of a robust dataset. Although medical devices now collect hundreds of types of information, researchers are only able to harness a small portion of the data collected for useful analysis. With the emerging power of artificial neural networks and their application in the fields of healthcare and pathology, new solutions are limited by the lack of useful information, as deep neural networks with tens of layers can include millions of parameters. With only hundreds of images, neural networks that are capable of solving problems are unable to do so.

My research presents the first step in removing this barrier to further progress. Its use of Generative Adversarial Networks to automatically synthesize medical images addresses the problem by augmenting the dataset. My research showed that increasing the size of the dataset by just 15% can yield boosts in accuracy. The accuracy of Alzheimer's analysis, cancer detection, liver segmentation, and more would be improved by larger datasets, and I have provided a method to do this. With the generation of new images, scientists can now augment their datasets for improved performance.

I hope that clinical testing of my system by radiologists will confirm my studies and provide a better diagnostic system for Parkinson's Disease. Additionally, I plan to determine the relationship between number of images generated versus accuracy gained. Currently, due to the lack of resources and processing power, the ability to generate images has been limited. With these improvements and further testing, my hope is that more Parkinson's patients will be diagnosed while treatment is effective. I want to apply this technique to many diagnostic methods for other diseases to improve accuracy, reduce cost, and increase accessibility.

5. Your Potential as a Scientist, Mathematician or Engineer (maximum 300 words)

Address through specific and concrete examples what characteristics you have that best demonstrate your affinity and aptitude for being a good scientist. What have you done that illustrates scientific attitude, curiosity, inventiveness, initiative? How does your experience suggest future success as a scientist, mathematician or engineer?

My goal has always been to help others. Whether it is through improving medical diagnosis or showing students the power of artificial intelligence, my focus has been on using science to improve the world around me. As cofounder of GirlsComputingLeague, I organize the annual AI Summit, where over 350 students attend a day of learning AI from industry experts, business executives, and researchers, including Peter Norvig of Google and John Whyte of WebMD, and participate in interactive lessons through a partnership with NVIDIA and AI4ALL. I carry the goal of generosity through my research as well, focusing on not only improving accuracy and other statistical metrics, but also on reducing the cost of diagnosis, hoping to directly impact lives with my research. After seeing thousands of patients with Diabetic Retinopathy not able to access doctors or treatment, I created a smartphone attachment for the automatic diagnosis of DR that is now being used to triage DR in India and Mexico in low income neighborhoods.

My scientific curiosity has also been fueled by my passion for the fields of computer science and mathematics. My love of mathematics inspired me to become a captain of the Thomas Jefferson Math Team, where I help hundreds of students by writing contests, lecturing on advanced math topics, and organizing competitions for local middle school students. I carry a similar passion in my research, and it has helped me find creative solutions to problems I have faced. After using standard techniques to improve the accuracy of diagnosis for Parkinson's, I thought of generating more images to mitigate the problem of a limited dataset. I plan to bring generosity and a passion for helping others to all of my future scientific endeavors.

6. Major Scientific Question (maximum 400 words)

What is a major scientific question in your field whose answer you believe will have a significant impact on the world in the next 20 years, and why? Using examples from your own experience or research, explain how you might envision addressing the question over the next 20 years.

The artificial intelligence revolution was fueled by the creation of the Graphics Processing Unit (GPU). Its ability to perform hundreds of billions of operations created a symbiotic relationship with artificial intelligence, which requires billions, if not trillions, of calculations to be done for the systems to learn from complex data. As we have begun to tackle larger and larger problems, researchers have thrown multiple GPUs at the same task in an attempt to combine the power of many GPUs. However, because information doesn't pass through GPUs instantly, one additional GPU only provides a boost of approximately 40% to computation speed. This inefficiency piles up, and the computational power merely doubles when using five GPUs.

In my project, I faced the problem of GPU parallelism. With hundreds of gigabytes of data to transfer, multiple GPUs took hours longer than training on a single GPU. Other scientists have recognized this problem and are in favor of using one strong GPU instead of multiple smaller GPUs. However, as GPU technology can't keep up with society's thirst for scientific advancement, problems have begun to take days, weeks, and even years to solve. Many important advances in the field of artificial intelligence, including explainability, require massive amounts of processing power to examine. To achieve these advances, we must understand and solve GPU parallelism.

Attending NVIDIA's GPU Technology Conference gave me a fresh perspective on the problem. I am excited about the recent steps in solving parallelism, including NVLink, a quicker way to communicate between GPUs to increase the information shared. At NVIDIA's GTC, I saw advances in hardware, like the NVIDIA Titan V and DGX-2, both of which have improved information bandwidth. In spite of these advances, I believe that quantum computing is the next step in solving parallelism. My work in artificial intelligence and internship at MITRE exposed me to the power of simple quantum computers and the vast applications they have. Seeing the advances in quantum computing at NVIDIA's GTC has inspired me to work on improving quantum machines at the intersection of physics, mathematics, and computer science. I hope to take the information about hardware I gained from the GTC and the depth of computer science and artificial intelligence knowledge I learned from my research and internship to work on solving this problem. Removing the roadblock of parallelism is the first step to solving major problems in healthcare, socioeconomics, and computer science.

7. "Tweet" about your project! Tell us about your project in 280 characters or less.

The Society might share this response if you are named a scholar or finalist.

PDGAN uses a novel method to create more medical data and diagnoses Parkinson's disease 16% better than current clinical practice. It is cost-effective, quick, and efficient for diagnosing Parkinson's, and paves the way for making early diagnoses available to all patients.

Previous Research

Created: 08/24/2018 • Last updated: 11/15/2018

Tell us about your previous science research projects

1. Individual Science Projects (optional)

List any individual research projects (not class projects) to which you have contributed during high school. A project does not need to have been submitted to a competition to be listed. List projects in chronological order, starting with the most recent project on top.

	Start Date (MM/DD/YYYY)	End Date (MM/DD/YYYY)	Supervising scientist/men- tors (if any)	Project Title	Competition/ Awards (if any)	Check this box if this is the project you are entering to Regeneron STS
1.	08/16/2018	10/28/2018	Mr. Hannum, TJHSST; Dr. Adrian Dalca, MIT;	MRI Image Synthesis for the Diagnosis of Parkinson's Disease using Deep Learning	N/A	<input checked="" type="checkbox"/>
2.						<input type="checkbox"/>
3.						<input type="checkbox"/>
4.						<input type="checkbox"/>
5.						<input type="checkbox"/>

2. Team Science Projects

Yes

Have you ever completed research with
a team member?

2a. Team Research Projects

Research conducted as part of a student team project is not eligible for Regeneron STS. This includes any research or portion of research regardless of whether it has been or will be submitted for competition or not. Even if you were a primary member of a team or conducted one portion of the research, it must still be considered a part of that team project. Please provide information on any team projects to which you have contributed.

	Start Date (MM/DD/YYYY)	End Date (MM/DD/YYYY)	Teammate Name(s)	Mentor Name(s)	Project Title	Competition/ Awards (if

))			any)
1.	06/02/2016	04/18/2017	Kavya Kopparapu, Justin Zhang	Dr. Hejtmancik, National Eye Institute; Dr. Beck, Harvard Medical School; Dr. Fuchs, Memorial Sloan Kettering Cancer	Eyeagnosis: Automatic Diagnosis of Diabetic Retinopathy using Deep Learning 2017 ISEF 3rd Place Grand Prize in Translational Medical Sciences; 2017 Virginia State Science Fair Grand Prize/Governor's Award; 2017 International Conrad Spirit of Innovation Challenge 1st Place; 2016 Siemens Competition Semifinalist
2.	06/15/2017	08/18/2018	Kaien Yang	Mr. White at Thomas Jefferson High School for Science and Technology.	TweePression: A Computational Approach to Detecting Major Depressive Disorder (MDD) Using Twitter Feeds and fMRI Scans MIT INSPIRE Semifinalist
3.	9/20/2014	4/8/2015	Kavya Kopparapu, Rahul Rajan, Alex Peng	Dr. Cobb at Thomas Jefferson High School	StentSense: A Stent-Smartphone System to Monitor Patient Blood Contents using Raman Spectroscopy 2015 Toshiba ExploraVision National Honorable Mention
4.					
5.					

2b. Clarification of Team Research (maximum 400 words)

Explain how the research you are submitting with this application is different from these team projects. Be sure to address each aspect of your research—purpose, procedure, data, and conclusions.

The work I have performed in team projects have related to the intersection of artificial intelligence and healthcare, but are very different from the current work I am pursuing. The purpose is different, as it is targeting a different disease and underlying case. The data was collected from different places, and I drew up conclusions differently as I used an additional procedure of using GANs for image synthesis in the current project. The image analysis procedures may be similar to these projects, but that is because they are very similar to the standard use of Convolutional Neural Networks in image analysis.

3. Conference Presentations, Abstract Publications, Student-Level Publications, Paid Publications, or Future Submission Plans (optional; maximum 400 words)

If any, please list any conferences in which you have presented your work; instances where the abstract of your research has been published; publications in student-level, non-peer reviewed, or pay-for publication journals; and future publication plans (including articles submitted to journals, but not yet published) (If any):

My work in Eyeagnosis has been published in various news outlets, including IEEE xplore's paper, NVIDIA's Deep Learning blog, and more. It was also the subject of a presentation at the O'Reilly AI conference in 2017 and a International Society for Computational Biology (ISCB) Rocky Colorado Conference Poster. It is being reviewed for publication in the Indian Journal of Ophthalmology, a peer-reviewed journal.

My work in TweePression was the subject of a presentation at the O'Reilly AI conference in 2018, a poster at the NVIDIA's 2018 GPU Technology Conference in Silicon Valley, and a presentation at NVIDIA's 2018 GPU Technology Conference in DC.

4. Ph.D.-level Peer-reviewed No

Journal Publications

Are you listed as an author or coauthor on any scientific publications? (you may select more than one option)

Science Research Description

Created: 10/05/2018 • Last updated: 11/15/2018

Tell us the basics about your project

1. Project Category

Computational Biology and Bioinformatics

Select the category that best fits your project. Three PhD-level evaluators with expertise in the area you choose below will review your work.

2. Project Title

Enter your project title. For any symbols, please write the name of the symbol in all capital letters (ALPHA, GAMMA, etc.)

Otherwise, please use normal Title Case.

Full Project Title:	MRI Image Synthesis for the Diagnosis of Parkinson's Disease using Deep Learning
Short Project Title (50 characters or less):	MRI Image Synthesis to Diagnose Parkinson's Disease
If the title of your project requires any special symbols or formatting (such as italics) please explain here:	N/A

3. Where was the experimentation / research conducted?

Home

High School

Select all that apply.

4. Please check all that apply to your research experience:

Participated in research at my high school after school or on weekends

Identified and contacted a scientist independent of any support

Read science journals, books, and/or magazines, and performed the research without mentorship

5. Primary Research Location

Name of Research Location (institution, university, park, school, etc.)	Thomas Jefferson High School for Science and Technology
Describe the Research Location	The Computer Systems Lab and Neuroscience Lab provided hardware resources, including GPUs, and mentorship that was vital to the success of the project.
Was the primary research location in the United States or a US territory?	yes
Research Location City	Alexandria
Research Location State	VA

6. Mentors

Please provide information on the adults, undergraduate student or higher, with whom you met and worked with in any way related to your research. You will be asked to provide their contact information in Task 12.

	Prefix	First Name	Last Name	Institution (if different from above)	Job Title	Time spent with this mentor
Mentor #1	Mr.	Mark	Hannum		Lab Technology Teacher	1 hour per week in September and October
Mentor #2	Dr.	Adrian	Dalca	Massachusetts Institute of Technology	Postdoctoral Fellow	3 hours total
Mentor #3						

7. How did you get the idea for your research? (maximum 200 words)

Explain the development of your research question and/or engineering goals.

My grandfather was recently diagnosed with the late stages of Parkinson's Disease, and it was at a point where much of the treatment was ineffective due to its stage. Upon further reading, I found this occurrence my family encountered was not uncommon. With my background in solving diagnoses problems using Artificial Intelligence, I believed that I could potentially make an impact on families like mine with early diagnosis, so adults that have Parkinson's could be diagnosed quick enough for treatment to be effective.

8. What was the duration of the research? (maximum 150 words)

Explain the amount of time you spent on the research project that you have submitted.

The duration of the research spanned from mid-August to mid-October, from when I thought of the problem idea to when the research was complete. In that time, I performed my literature review, designed and tested protocols, analyzed data, and formed conclusions. Every day, I would spend anywhere from three hours to eight hours between discussing with mentors, and coding. This time does not account for the time it took for the models to train and converge, as those would say on overnight for weeks at a time.

9. If your research was conducted as part of a larger research project or group, explain how your work is independent of this larger project. (maximum 250 words)

If there were other high school students in the group, be specific about how your work was similar to and different from other students vs. your independent work.

This research was not done as a part of a larger group.

10. Please attribute the support you received in each area of the research process and highlight what you claim as your own, original, unique contribution. (maximum 200 words per section)

Provide a description of what you did in sentences, more than simply stating a percentage or writing "all."

a. Developing / Initiating the purpose of the research	As a family member was the primary reason I conducted this research, I developed the initial purpose and goals of the research myself. However, the goals evolved as I began to face problems and found solutions during the research. I consulted with members of the scientific community, including postdoctoral fellow Adrian Dalca, and my mentor Mr. Hannum in looking for solutions to the problem.
b. Designing the procedures	I designed the procedures that I would use to train and evaluate my models. The idea for using Convolutional Neural Networks, however, is common knowledge and something I learned from doing my literature review. After using CNNs, I consulted with scientists to determine methods to improve my accuracy, but after hearing general advice like "solve underlying problems", I thought of using Generative Adversarial Networks to generate more images to augment the training data. I designed the procedures of training and evaluating the GANs as well.
c. Implementing the procedure (including special techniques or the use of special equipment)	From my past research, I am very knowledgeable in the field of diagnosis with artificial intelligence. I implemented the procedures I created myself. However, the hardware I used to train and evaluate my models were not mine. The training of the models was done on an AWS p2.xlarge machine on an NVIDIA K80 GPU. The preprocessing and evaluation of data was done at Thomas Jefferson High School, on the computer systems lab. The TJ CSL has multiple NVIDIA GTX 1080 GPUs. The deep learning frameworks were also not developed by me. I used a combination of Keras and Tensorflow, 2 popular deep learning frameworks.
d. Gathering / Recording data	I did not gather nor record the data. The dataset I used was from the University of Southern California's Laboratory of Neural Imagine Data Archive. In specific, the Parkinson's Progression Markers Initiative (PPMI) was the study that collected the data I used. The PPMI database had thousands of MRI images from various machines, patients, and locations. It also had collected genetic data that I hope to use in the future as an augment to the current project.
e. Analyzing data	The analysis of data was straightforward using Python libraries like matplotlib and scipy. The analysis metrics are also well known and coded into these libraries.
f. Formulating conclusions	I formulated the conclusions and future research on my own.

11. Indicate any other substantive guidance received, as well as any prior research involvement or training that helped you in conducting your own work in this project. (maximum 250 words)

My past research, with Eyeagnosis and Tweepression, has shown me many of the methods of using artificial intelligence in the field of healthcare. Previously, I have done work with image analysis, image segmentation, and text classification. These past projects have helped me develop techniques for analyzing my current project regarding Parkinson's Disease.

12. Statement of Independence

- **Frequently Regeneron STS applicants perform research that is similar to that of parents, mentors, relatives, friends and/or other high school students. This is expected, since science is a cumulative process, each finding built on a previous one. The influence and assistance of those around you may be beyond the individual tasks included in the questions above.**
- **In order to recognize the independent research of student investigators, a clear picture of the evolution of your work and the aspects that are of your own design and execution are required. We would like to give you the opportunity to reassure the evaluation committee that while the above-mentioned influences may exist, the work you have submitted is your own and not that of a parent, mentor, relative, friend or any other person.**
- **Failing to disclose similar or related research of which you are aware or failing to mention any person who has either closely or loosely guided you, and their relationship to you and your family, is a violation of our rules and the ethics statement you must agree to in order to submit this application, and is grounds for disqualification from the Regeneron STS.**

Select the option below that best describes your situation:

I certify that there are no additional people who have done research in an area of science close to mine, nor is there any additional person who has closely or loosely advised me, contributed to my research or had any small influence on my work.

13. Intellectual Property and Viewing Your Application

I certify that I have discussed this submission with the scientists with whom I worked and they do not have concerns regarding intellectual property. I give permission to Society staff to show my entire application, including my research report, to any of my mentors or recommenders.

14. Research Report Guidelines

Read the [Research Report Guidelines](#) before proceeding to the next task, where you will upload your Research Report. The Research Report Guidelines document contains important information about the format of your research paper, naming convention, size limits, and more.

I have read the Research Report Guidelines 2019 document

MRI Image Synthesis for the Diagnosis of Parkinson's Disease using Deep Learning

Name: Neeyanth Kopparapu

Category: Computational Biology and Bioinformatics

Abstract

Parkinsons disease (PD) is a neurodegenerative disease that affects an estimated 1% of adults over 65. While the disease itself is not fatal, complications related to PD are rated as the 14th largest cause of death in the United States by the Center of Disease Control and Prevention. In spite of technological advances, Parkinson's diagnosis methods have not changed, and the accuracy of diagnosis has remained at approximately 81% for the past 25 years. These methods include analyzing years of neurological data to determine if the patient has developed the symptoms of Parkinson's, including limb rigidity and tremors, both of which are common side effects of a number of other diseases.

With the rise of automated prediction algorithms paired with the generation of massive amounts of data, the automatic diagnosis of Parkinson's Disease has not caught up to traditional means. This is commonly attributed to the lack of useful data, as most computational systems require a tremendous amount of medical data that isn't readily available as gathering the data can be expensive.

This study presents PDGAN, a tool to aid pathologists and neurologists in the diagnosis of Parkinson's Disease. PDGAN uses a series of neural networks to classify Magnetic Resonance Images (MRI Images). PDGAN employs Generative Adversarial Networks (GANs) to synthetically generate medical images which is used to augment the classification efforts. The pair of Convolutional Neural Networks exhibited an testing accuracy of 91.4% without the augment of new data, and combined the total accuracy was 96.6%, a 16% increase compared to traditional methods at a fraction of the cost and time. PDGAN demonstrates the feasability of utelizing GANs to generate unseen data for the improvement of classification accuracy in the medical setting.

Introduction

Parkinson's Disease (PD) is the second most prevalent neurodegenerative disease, affecting approximately 1% of the population above the age of 65 [31]. Although PD is not a fatal disease, it decreases life expectancy by 16 years [36]. PD affects movement, originally with small tremors, slowness of movement, and shortness of breath, but symptoms worsen as the condition progresses while untreated [5]. Although the exact cause of PD has not been determined, PD is known to kill cells in the substantia nigra, the area of the brain responsible for the production of dopamine, a neurotransmitter responsible for the control of movement. The loss of dopamine, an early sign of Parkinson's, leads to symptoms including bradykinesia, the slowness of movement [35].

Although there is no specific test for the diagnosis of PD, the current procedure is as follows: (1) a physical exam is conducted, (2) further lab tests are conducted (Blood Work, MRI, Brain Ultrasound), (3) manual analysis of past neurological history, (4) determine if symptoms are present [4]. These non-deterministic tests often take weeks and produce unreliable results. It is also very difficult to predict symptoms until they are already present, and the longer PD goes undiagnosed in a patient, the harder it is for treatment to be effective [36].

Therefore, it is important for patient's future for them to have a quick and accurate diagnosis. However, due to the lack of a specific diagnosis pipeline for PD, the accuracy of diagnosis has not increased despite the advancements in the fields of medicine, technology, and bioinformatics. The accuracy of diagnosis reported in 1992 from 100 various studies was 82% [15], and the accuracy in 2014 was 83.9% [26]. Although numerous computational vision attempts at diagnosing PD have emerged, including automatically regressing neurological data and classifying patient micrography, efforts haven't reached classification metrics that compare to manual efforts, with accuracies near 67% [25].

Many previous studies attribute their difficulties to two main problems. First, there is a lack of useable data. Complex regression and deep learning models that include millions of parameters require a hundreds or thousands of data points to learn from, and many available datasets contain anywhere from 50 to 200 images, not nearly enough for capable models to

converge. Second, studies determined the hardest diagnosis class for PD is the patients right after showing signs of symptoms. At this stage, symptoms have not fully present, so it is nearly impossible for neurologists or automatic tools to detect the presence of PD [25].

This study presents an approach of the automatic approach to aid the diagnosis of PD through patient Magnetic Resonance Imaging (MRI) scans. MRI scans have historically been used to determine alterations in the Central Nervous System (CNS), some of which might be indicative of PD. MRI scans have been shown to be more effective than other imaging processes in displaying these changes [12], and most diagnosis efforts from scientists include an MRI scan of the brain. The approach uses a system of Fully Convolutional Networks (FCNs) and Convolutional Neural Networks (CNNs) to perform image analysis tasks. It also utilizes Generative Adversarial Networks (GANs) to augment the current dataset with more artificial scans that were used in the classification training.

Materials and Methods

Dataset Description The data used for the system came from the University of Southern California's Laboratory of Neurological Imaging (LONI). The Parkinson's Progression Markers Initiative under LONI's Image Data Archive contains MRI scans, genetic data, medical history, and motor assessments from over 1400 Parkinson's and Control patients. The Image Archive contains 921 Control MRIs and 2633 PD MRIs viewed in the Axial form[29].

Magnetic Resonance Imaging (MRI) is an imaging procedure that uses magnets and radio waves to capture detailed images of the brain. Although there are many types of MRI scans, including functional MRIs (fMRI), cardiac MRIs, Magnetic resonance angiography (MRA), and Magnetic resonance venography (MRV), the most common MRI is a anatomical scan of the brain. This anatomical scan displays the shape, volume, and developmental changes in the brain as a three-dimensional image.

The anatomical MRIs obtained from the data archive came in 43 different sizes, ranging from $1024 \times 1024 \times 256$ to $32 \times 24 \times 32$. Due to the large variance in sizes, the data actually used was one that had a high volume of images, but also had a representative split compared to the total dataset. The chosen size was $256 \times 240 \times 176$, as it had a total of 612 images, 146 from the Control group, and 466 from the PD group, representative of the approximate 1 : 3

split in data throughout the study. The 86 generated images had 23 MRIs of the control group and 63 MRIs of the PD group, keeping consistent with the same ratio as the previous data as shown in Figure 1.

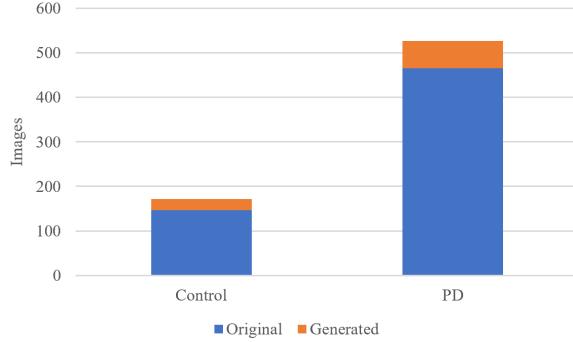


Figure 1: Chart displaying sizes of generated and original dataset belonging to each class.

Software and Hardware The software platform used for the creation of this project was Python. Python was used due to its modularity and compatibility with numerous standard machine learning, plotting, and artificial intelligence frameworks. To perform standard image preprocessing and analysis tasks, the OpenCV and PIL (Python Imaging Library) libraries were used [20].

Both Tensorflow and Keras were used as deep learning frameworks. Keras was used due to its code readability and modularity, and Tensorflow was used for its extensive online documentation, as well as its use for machine learning, mathematical computation, and deep learning [32, 21]. Artificial Neural Networks (ANNs) were used to perform the classification task. ANNs have grown in popularity for outperforming humans in various tasks, including computer analysis tasks [23].

Throughout the study, Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) were utilized. CNNs are a flavor of ANNs that specialize in image processing. Figure 2 shows the basis of what a Convolutional Network does.

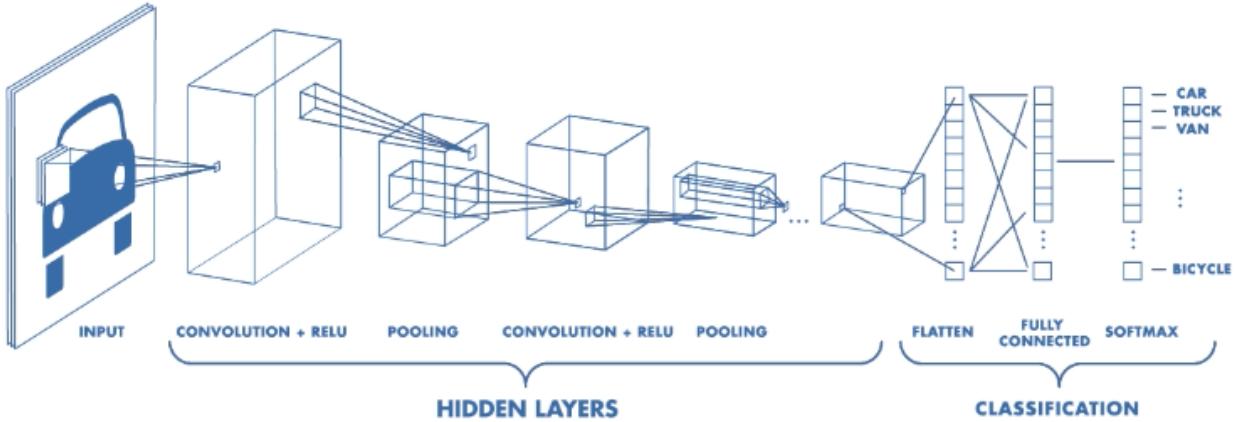


Figure 2: Example Fully Convolutional Network that has Convolution layers [22].

GANs [11], in the class of unsupervised machine learning, are a version of an adversarial network which works to generate images similar to the samples given using a noise vector. As shown in Figure 3, the Generator takes a noise vector which it uses, in addition to its trained weights, to create images. The Discriminator is tasked to determine if the sample given to it is real. The adversarial networks's goals are to make it harder for each other, and in turn improve. GANs are trained over thousands of epochs to converge, but these epochs take less time compared to traditional training epochs [11, 2].

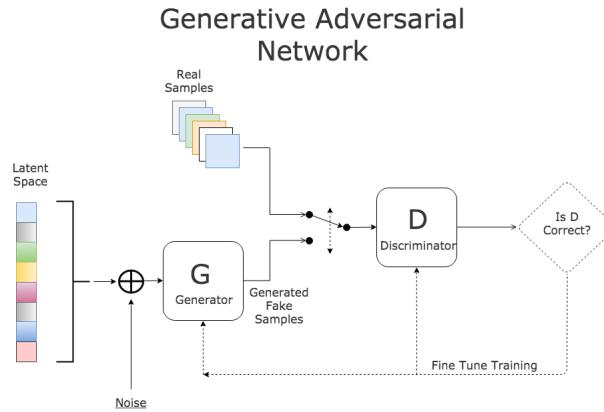


Figure 3: Example GAN data flow (Stack Exchange).

To complete the task, a Tesla K80 GPU was used, which included 11.2 GB of GPU memory, and 4992 CUDA cores. A p2.xlarge AWS EC2 instance was used with an Ubuntu backend and 100 GB extra SWAP space to store data during program calls.

Approach - Preprocessing Minimal preprocessing steps were taken. First, the images were normalized. The images were not subjected to standard augmentation procedures including cyclic pooling and tilting because hemispherical assymtries in MRI images can reveal important features of the brain anatomy [34]. The dimensions of the image array after preprocessing was $256 \times 240 \times 176$.

Approach - First Iteration The first iteration of the system consisted of the preprocessing steps and the image classification neural network, trained only on the images in LONI's PPMI database. Figure 4 shows the flow of data in the system.

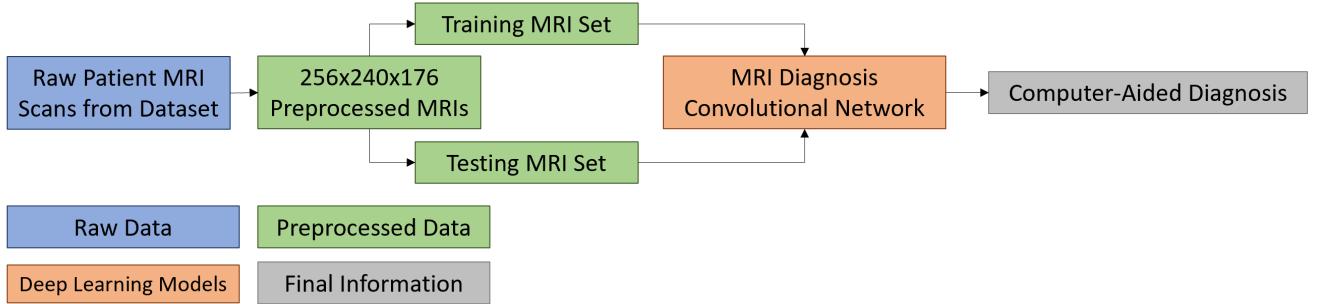


Figure 4: Graphical Description of the flow of data and processing through the system.

There were various models used, in an attempt to maximize accuracy. First, the traditional CNNs, including the VGG-19, GoogLeNet, and Resnet-50, were used due to their high accuracy in standard benchmarks like ImageNet [27, 30, 14]. The VGG model is described in Figure 5, where "E" is the VGG-19 model. Because MRIs are three dimensional, the CNNs were repurposed for 3D classification in Keras.

To further improve the accuracy, the Spearmint [28] package was used. Spearmint, a Bayesian optimization package which changes hyperparameters to minimize loss, was used on the VGG-19 model which had the highest accuracy before using the package. With 5 variables to change the different size of the filter layers, the model's accuracy was increased

slightly to 91.4%. Spearmint was not allowed to fully complete its optimization due to time constraints, as each trial would take upwards of 6 hours with early stopping. Thus, only 18 spearmint programs were run.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 5: VGG System Architecture [27]

To improve the accuracy, some problems were highlighted.

1. The size of the dataset is small - with over 400 million parameters, only a few hundred images is not enough to make the model converge.
2. With deep networks like the vanilla GoogLeNet, the *vanishing gradient problem* only allowed small changes to be made progressively, with large changes not having much of an impact on the early layers.

Many steps were taken to combat these problems, all of which were reflected in the second interaction of the model.

Approach - Motivation Wang et al. described one of the first methods of augmenting a dataset during low-shot learning [33]. The hallucinator, G , described is able to take an input (x, y) , with a noise vector z and output a "hallucinated" input (x', y) that would

then be used to augment the original training data. The hallucinator used was a multilayer perceptron. Similar to [33], the goal of this project was to similarly augment the trainint set with new MRI scans. However, the study used a three layer MLP, which would not be able to make the complex transformations nessecary to generate the 3-dimensional scans.

Kingma and Welling described the first methods of generation using encoders [18], but like Wang et al., the networks described were used for simple image tasks with small data sets and small dimensional images. Goodfellow et al.'s work with GANs and the image generation process created not only a generator (similar to the hallucinator previously), but also a discriminator - serving as the adversarial network in the pair [11]. The GAN would fix the problem of a small dataset.

Meanwhile, like in [30], auxilliary classifiers were used to solve the problem of the vanishing gradient.

Approach - Second Iteration The second iteration of the system consists of 2 network models—one for image classification and the final diagnosis, and one for artificial image synthesis for data augmentation—shown in Figure 6. The images are from a shared dataset, but the images used to train the GAN as well as the ones generated by the GAN were only used in the tranining set for the classifier. This meant that only real images not seen by either the GAN or classifier in training made up the evaluation (test) set.

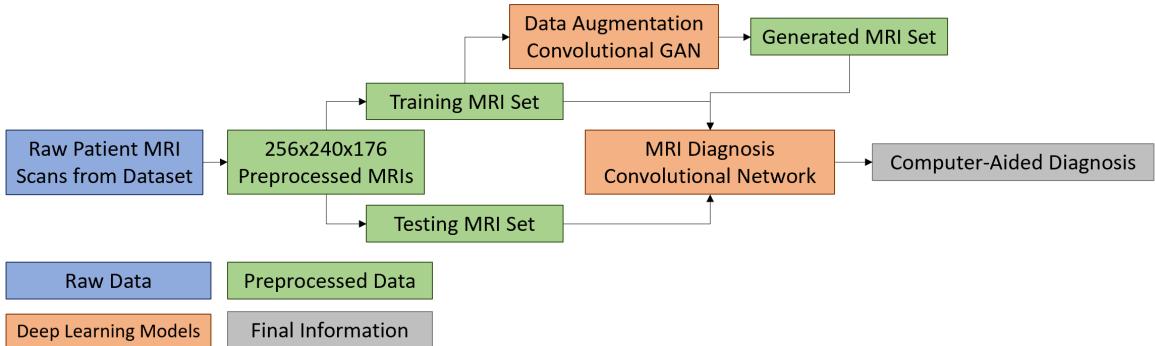


Figure 6: Graphical Description of the flow of data and processing through the system.

Evidently the change made is the GAN and its ability to add to the training data for the diagnosis network. The same architectures from the first iteration (VGG, GoogLeNet, and Resnet) were trained again to determine the effects of the addition to the dataset.

Additionally, one other set of networks was used, including a feature extractor appended to a classification network. Auxilliary classifiers were added throughout the VGG, GoogLeNet, Resnet, and new models. Similarly to [19], the generator network's weights were initialized as block diagonal identity matrices. The GAN model was inspired by [2]. The architecture of the new model is described in Figure 7.

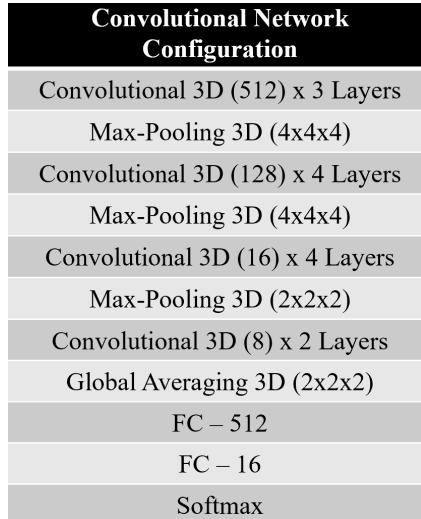


Figure 7: Graphical Description of the flow of data and processing through the system.

The same optimizations made after model creation for the first iteration were applied to the second, including using Spearmint.

Results

First Iteration The following table indicates accuracy, sensitivity, and specificity of the 3 classification algorithms after the 15 training epochs.

Table 1: Table detailing the accuracy, sensitivity, and specificity of the 3 models.

Model	Accuracy	Sensitivity	Specificity
VGG-19	90.20%	93.10%	81.08%
GoogLeNet	84.97%	86.21%	81.08%
Resnet-50	88.89%	92.24%	78.38%

$$Sensitivity = \frac{TP}{TP + FN} \quad Specificity = \frac{TN}{TN + FP}$$

The number of epochs, 15, was chosen as a part of the EarlyStopping callback in Keras [32], which continues the training of the algorithm until the validation loss experiences multiple epochs of no decrease. The graphs of loss are depicted in Figure 8.

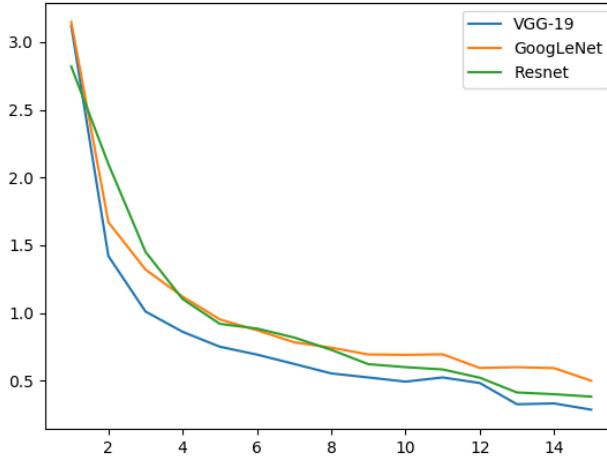


Figure 8: Validation Loss Graphs of the 3 models through the 15 epochs.

The model's weights were fine-tuned to attain improved performance on the validation set, and had minimal overfitting due to many empirically tested methods, including adding dropout, regularization, and patience.

After using the spearmint package, the accuracy of the VGG model was increased to 91.4%. The ROC curve for the finalized model is given in Figure 9.

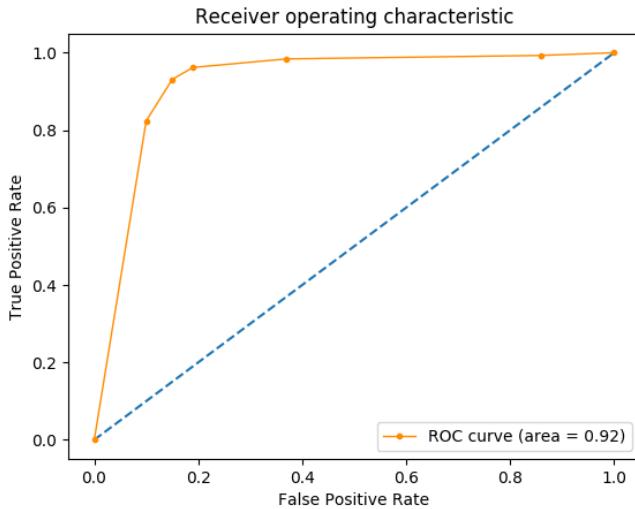


Figure 9: Receiving Operating Characteristic curve for the VGG Model.

Second Iteration With the addition of the GAN, the 3 models performed significantly better. In addition, the new model created surpassed the accuracy of the previous and new models. Table 2 gives the accuracy, sensitivity, and specificity of the new models.

Table 2: Table detailing the accuracy, sensitivity, and specificity of the 4 new models.

Model	Accuracy	Sensitivity	Specificity
VGG-19	94.12%	94.83%	91.89%
GoogLeNet	91.50%	92.24%	89.19%
Resnet-50	89.54%	87.93%	94.59%
PDGAN	96.62%	97.41%	94.59%

To determine the effects of only the addition of the new data, the same number of epochs was used for each system, namely 15. However, it was evident the model was still able to learn more, as the loss formed a monotonic decreasing sequence. Looking at the ROC curve in Figure 10, as it is a common evaluation metric among medical devices [7], it is evident that PDGAN's performance is better than all the first iteration models due to the increase in dataset.

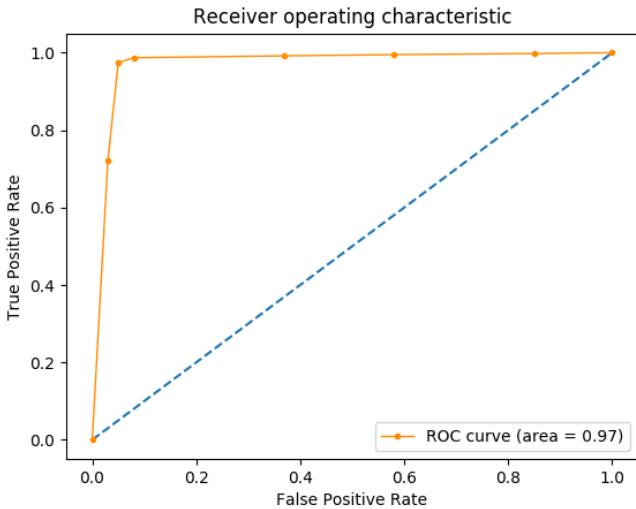


Figure 10: Receiving Operating Characteristic curve for the new PDGAN Model.

Discussion

The primary objective of the work was to improve on the current diagnosis system for PD and increase the chance for early diagnosis among patients. As many patients get MRI scans done as a preliminary test during PD screening, analyzing MRI scans for early signs of Parkinson’s Disease was ideal because of the MRI’s ability to analyze the anatomy of the brain with high precision [12]. Computationally generating an alternative diagnosis gives neurologists the power to access second opinions in a time and money efficient manner. This is crucial for the wellbeing of patients, as treatment plans become less accessible for patients with Parkinson’s in later stages [6].

With around 12 thousand MRI machines in the US, the applicability of the system is widespread. Still, in over 20 years, a firm technological solution has not emerged for the diagnosis of PD in spite of the MRI process being standard procedure during the screening of MRIs. Multiple reports have cited an technological solution as a next step in the diagnostic efforts[15, 26, 16].

A search of scientific literature reveals that the PDGAN is unique in its ability to diagnose Parkinson’s Disease with an augmented, generated dataset. Table 3 illustrates several related

studies, and their difference to the PDGAN’s methodology.

Table 3: Table detailing information about several related studies and the difference between the study and the current study [3, 24, 9, 10, 1].

Study	Description	Input Data	Methodology	Accuracy	Difference between study
Chen, 2013	FKNN – based Diagnosis	Voice Measurements	Fuzzy K-Nearest Neighbors	96.07%	Had thousands of sample data
Frid-Adar et al.	Liver Lesion Classification	Liver Lesion Images	GANs, CNNs	88.4%	Used GANs but with a different classifier – low accuracy
Gil et al.	MLP – based Diagnosis	Voice Measurements	MLP and SVM	92.31%	Had thousands of sample data
Adams, 2017	Typing based Diagnosis	Typing Movements	Various Machine Learning Models	96.1%	Had thousands of sample data
Pereisa et al.	Writing and Medical Exam Diagnosis	Handwriting, Medical Exam Information	Computer Vision Processing, CNNs, MLP	67%	Low accuracy, used a combination of tests.

Although this method of improving accuracy hasn’t been used in the context of Parkinson’s Disease, there is still room for improvement. Evidently, the accuracy could be increased with a larger starting dataset. Additionally, due to hardware constraints, the spearmint code was stopped prematurely, only making 2 optimizations. With more time and better hardware, the optimization would have improved.

Conclusion and Future Work

This study presented the first data augmented approach at classifying Parkinson’s Disease. It is able to classify MRI scans with a high accuracy by augmenting the originally small dataset with more images to learn upon. To obtain this information, the study used a combination of state of the art CNNs and GAN models, as well as a self-created CNN able to outperform the defined ones using the spearmint package.

Several performance metrics for all 7 systems are summarized in Table 4, and demonstrate the positive effect of adding generated images to the dataset.

Table 4: Table detailing the accuracy, sensitivity, and specificity of the 7 models. The bolded models indicate ones that had access to the augmented data.

Model	Accuracy	Sensitivity	Specificity
PDGAN	96.62%	97.41%	94.59%
VGG-19	94.12%	94.83%	91.89%
VGG-19	94.12%	94.83%	91.89%
GoogLeNet	84.97%	86.21%	81.08%
GoogLeNet	91.50%	92.24%	89.19%
Resnet-50	88.89%	92.24%	78.38%
Resnet-50	89.54%	87.93%	94.59%

A future step for this study would be to broaden the scope of analysis of patients. Another method of analysis apart from looking at MRI scans would be to use genetic markers. Cited as another method of early detection [6, 37], looking at methylation profiles as a way to detect early signs of Parkinson’s Disease would be a clear next step [8]. Already being used as a diagnosis tool for cancer [13] and muscular dystrophy [17], the PPMI database already has genetic markers and methylation profiles for patients with Parkinson’s Disease.

Overall, it is clear with this research and the PDGAN model that the addition of a Generative Adversarial Network for the purpose of data augmentation is beneficial to the classification accuracy where there is not an large, accessible dataset.

References

- [1] Warwick R. Adams. “High-accuracy detection of early Parkinson’s Disease using multiple characteristics of finger movement while typing”. In: *PLOS ONE* 12.11 (Nov. 2017), pp. 1–20. DOI: 10.1371/journal.pone.0188226. URL: <https://doi.org/10.1371/journal.pone.0188226>.
- [2] Antreas Antoniou, Amos Storkey, and Harrison Edwards. “Data Augmentation Generative Adversarial Networks”. In: abs/1803.01229 (2018). URL: <https://arxiv.org/abs/1803.01229>.
- [3] Huiling Chen. “An efficient diagnosis system for detection of Parkinson’s disease using fuzzy k-nearest neighbor approach”. In: *Expert Systems with Applications* (Jan. 2013).
- [4] Mayo Clinic. *Parkinson’s Disease Diagnosis and Treatment*. URL: <https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/diagnosis-treatment/drc-20376062>.
- [5] Mayo Clinic. *Parkinson’s Disease Symptoms and Causes*. URL: <https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055>.
- [6] Emily Downward. *Diagnosis & Early Symptoms & Early Diagnosis*. 2017. URL: <https://parkinsonsdisease.net/diagnosis/early-symptoms-signs/>.
- [7] Tom Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters* (2006).
- [8] Suhua Feng et al. “Determining DNA methylation profiles using sequencing.” In: *Methods Molecular Biology* 733 (2011), pp. 223–238. DOI: 10.1007/978-1-61779-089-8_16. URL: <https://link.springer.com/protocol/10.1007>.
- [9] Maayan Frid-Adar et al. “GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification”. In: *CoRR* abs/1803.01229 (2018). arXiv: 1803.01229. URL: <http://arxiv.org/abs/1803.01229>.

- [10] David Gil and Magnus Johnsson. “Diagnosing Parkinson by using artificial neural networks and support vector machines”. eng. In: Global Journal of Computer Science and Technology 9.4 (2009), pp. 63–71. ISSN: 0975-4172.
- [11] Ian J. Goodfellow et al. “Generative Adversarial Networks”. In: abs/1406.2661 (2014). URL: <https://arxiv.org/abs/1406.2661>.
- [12] B Griewing, H Hielscher, and A Lutcke. “The importance of MRI (magnetic resonance imaging) for the diagnosis of brainstem infarction.” In: *Bildgebung* 59.2 (1992), pp. 94, 97.
- [13] Xiaoke Hao et al. “DNA methylation markers for diagnosis and prognosis of common cancers”. In: *Proceedings of the National Academy of Sciences* 114.28 (2017), pp. 7414–7419. ISSN: 0027-8424. DOI: 10.1073/pnas.1703577114. eprint: <http://www.pnas.org/content/114/28/7414.full.pdf>. URL: <http://www.pnas.org/content/114/28/7414>.
- [14] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *CoRR* (2015). eprint: 1512.03385. URL: <http://arxiv.org/abs/1512.03385>.
- [15] A J Hughes et al. “Accuracy of clinical diagnosis of idiopathic Parkinson’s disease: a clinico-pathological study of 100 cases.” In: *Journal of Neurology, Neurosurgery & Psychiatry* 55.3 (1992), pp. 181–184. DOI: <http://doi.org/10.1136/jnnp.55.3.181>.
- [16] Andrew J. Hughes et al. “What features improve the accuracy of clinical diagnosis in Parkinson’s disease”. In: *Neurology* 42.6 (1992), pp. 1142–1142. ISSN: 0028-3878. DOI: 10.1212/WNL.42.6.1142. eprint: <http://n.neurology.org/content/42/6/1142.full.pdf>. URL: <http://n.neurology.org/content/42/6/1142>.
- [17] Takako Jones et al. “Identifying diagnostic DNA methylation profiles for facioscapulohumeral muscular dystrophy in blood and saliva using bisulfite sequencing”. In: *Clin Epigenetics* 6.1 (2014), p. 23. DOI: 10.1186/1868-7083-6-23. URL: <https://clincalepigeneticsjournal.biomedcentral.com/articles/10.1186/1868-7083-6-23>.

- [18] Diederik Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: abs/1312.6114 (2014). URL: <https://arxiv.org/abs/1312.6114>.
- [19] Quoc V. Le, Navdeep Jaitly, and Geoffrey E. Hinton. “A Simple Way to Initialize Recurrent Networks of Rectified Linear Units”. In: *CoRR* abs/1504.00941 (2015). arXiv: 1504.00941. URL: <http://arxiv.org/abs/1504.00941>.
- [20] Fredrik Lundh. *Python Imaging Library (PIL)*. 2018. URL: <http://www.pythonware.com/products/pil/>.
- [21] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <http://tensorflow.org/>.
- [22] MathWorks. *Introduction to Deep Learning - what are Convolutional Neural Networks*. 2017. URL: <https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html>.
- [23] Dimitrios Siganos Christos Stergiou. *Neural Networks*. Imperial College London, 1996. URL: https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html.
- [24] Clayton R. Pereira et al. “A Step Towards the Automated Diagnosis of Parkinson’s Disease: Analyzing Handwriting Movements”. In: *2015 IEEE 28th International Symposium on Computer-Based Medical Systems* (2015), pp. 171–176.
- [25] CR Pereira et al. “A new computer vision-based approach to aid the diagnosis of Parkinson’s disease.” In: *Comput Methods Programs Biomed* 136 (2016), pp. 79–88. DOI: <http://doi.org/10.1016/j.cmpb.2016.08.005>.
- [26] G Rizzo et al. “Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis”. In: *Neurology* 86.6 (2016), pp. 566–576. DOI: <http://doi.org/10.1212/WNL.0000000000002350>.
- [27] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *CoRR* abs/1409.1556 (2014). arXiv: 1409.1556. URL: <http://arxiv.org/abs/1409.1556>.

- [28] Jasper Snoek, Hugo Larochelle, and Ryan Adams. “Practical Bayesian Optimization of Machine Learning Algorithms”. In: *Advances in Neural Information Processing Systems* (2012). eprint: 1206.2944. URL: <https://arxiv.org/abs/1206.2944>.
- [29] University of Southern California. *Parkinson’s Progression Markers Initiative*. URL: <https://ida.loni.usc.edu/login.jsp?project=PPMI>.
- [30] Christian Szegedy et al. “Going Deeper with Convolutions”. In: *CoRR* abs/1409.4842 (2014). arXiv: 1409.4842. URL: <http://arxiv.org/abs/1409.4842>.
- [31] C. Tanner and S. Goldman. “Epidemiology of Parkinsons disease”. In: *Neurologic Clinics* 14 (1996), pp. 317–335.
- [32] Keras Team. *Keras*. 2018. URL: <https://github.com/keras-team/keras>.
- [33] Yu-Xiong Wang et al. “Low-Shot Learning from Imaginary Data”. In: *CoRR* abs/1801.05401 (2018). arXiv: 1801.05401. URL: <http://arxiv.org/abs/1801.05401>.
- [34] K.E. Watkins et al. “Structural Asymmetries in the Human Brain: a Voxel-based Statistical Analysis of 142 MRI Scans”. In: *Cerebral Cortex* 11.9 (2001), pp. 868–877. DOI: 10.1093/cercor/11.9.868. eprint: /oup/backfile/content_public/journal/cercor/11/9/10.1093_cercor_11.9.868/1/1100868.pdf. URL: <http://dx.doi.org/10.1093/cercor/11.9.868>.
- [35] WebMD. *Parkinson’s Causes*. URL: <https://www.webmd.com/parkinsons-disease/guide/parkinsons-causes#1>.
- [36] WebMD. *Parkinson’s Later Diagnosis*. URL: <https://www.webmd.com/parkinsons-disease/news/20101004/parkinsons-later-diagnosis-earlier-death#1>.
- [37] Tao Wu and Mark Hallett. “A functional MRI study of automatic movements in patients with Parkinson’s disease”. In: *Brain* 128.10 (2005), pp. 2250–2259. DOI: 10.1093/brain/awh569. eprint: /oup/backfile/content_public/journal/brain/128/10/10.1093/brain/awh569/2/awh569.pdf. URL: <http://dx.doi.org/10.1093/brain/awh569>.

Rules Wizard/Form Uploads

Last updated: 10/05/2018

See the Regeneron STS Official Rules for help navigating this section.

Human Research

1. Does your project involve the use of human participants, including surveys (written, in-person, or online), and/or did you test a product of your project (your invention, prototype or computer application) using human participants other than yourself?

No

2. Does your project involve the use of human tissue? No

3. Does your project involve the use of data collected from humans (i.e. surveys, test scores, medical records, etc.) that were not collected by you nor your mentor? Yes

3a. Explain how your data was obtained. (maximum 200 words)

I used data collected by the PPMI Organization, which included a collection of de-identified Medical Images, Genetic Information, and other Medical Data concerning patients in numerous ongoing Parkinson's studies. Permission to access the data was required. Currently there are 1450 patients that have a various amount of MRI samples, Genome profiling data, Motor and Non-Motor assessments, as well as Previous Medical History surveyed.

3b. Identifiable information: choose the option below that best describes your data.

I worked with de-identified data. (This means that there were no names with the data when it was given to you. If it did come with names, do not check this box.)

Click "next" to proceed

4. Does your project involve live, non-human vertebrate animals or non-human vertebrate animal tissue?

Select all that apply:

Additional Information and Paperwork

Optional Paperwork Description

If you wish to provide additional paperwork to show approval processes from your local science fair, etc. you may do so here. Please describe what you are uploading and why in the text box, and upload your documentation in the next upload section.

(No response)

Optional Paperwork File Upload

(No response)

Beyond the Project

Created: 10/05/2018 • Last updated: 11/11/2018

Information collected in this section is confidential and will not be shared with evaluators or judges. The Society will use this information for general record keeping, and to contact your parents/guardians, teachers, principals, etc. in the event you are named a Scholar or Finalist.

Student Information

Your Contact Information

Your Cell Phone Number (optional)	5712636832
Your Home Phone Number	7039040024
Your preferred email address for long term contact:	neeyanthkvk@gmail.com

Your Address

Do you live in the United States or in a US territory?	Yes
Street Address	3123 Oxford Forest Dr
Street Address 2 (optional)	(No response)
City	Herndon
State	VA
Zip Code	20171
Name of County	United States

Date of Birth (MM/DD/YYYY) 02/02/2002

Ethnicity (optional) No

Are you Hispanic or Latino?

Race (optional) Asian or Asian American

T-shirt Size M

Explain your ability to speak, read, or write languages other than English (optional)

I am learning to Speak, Read, and Write Russian.

I speak Telegu at home.

Your Name for Public Materials

Please tell us how you would like your name to appear in public materials relating to the Regeneron Science Talent Search (official bio if selected as a finalist, press releases, etc)

First Name for Public Materials	Neeyanth
Middle Name for Public Materials (optional)	(No response)
Last Name for Public Materials	Kopparapu

Family Information

Do you have any siblings? (optional) Yes

Please list the name(s) and age(s) of your sibling(s): Kavya Kopparapu - 18

Are any of your relatives former top 40 finalists or former top 300 semifinalists/scholars/honors group in the Science Talent Search (under Regeneron, Intel or Westinghouse sponsorship?) Yes

Please list the name(s) and year(s) of their participation in the Science Talent Search:

Kavya Kopparapu - 2017-2018

Tell us more about Parent/Guardian #1:

Name	Rajani Kopparapu
Email Address	rajanikm@hotmail.com
Do you share a primary mailing address with this Parent/Guardian?	Yes
Marital Status in Relation to Parent/Guardian #2 named below (optional)	Married

Tell us more about Parent/Guardian #2 (optional):

Name	Madhu Kopparapu
Email Address	mkopparapu@gmail.com
Do you share a primary mailing address with this Parent/Guardian?	Yes
Marital Status in Relation to Parent/Guardian #1 (name is above)	Married

School Information

Please provide additional information about your high school principal, and the teacher most closely associated with your application.

The teacher listed here will receive prizes if you are named a finalist.

Principal First Name	Ann
Principal Last Name	Bonitatibus
Principal Email Address	anbonitatibu@fcps.edu
Teacher (at your HS) First Name	Mark
Teacher (at your HS) Last Name	Hannum
Teacher (at your HS) Email Address	MSHannum@fcps.edu

High School Newspaper

Please complete this section if your school has a newspaper or newsletter. We will notify this source if you are named a scholar or finalist so that your school community can celebrate this achievement.

Name of High School Newspaper	(No response)
Email Contact at High School Newspaper	(No response)

Project Information

Mentor Information

In a previous task, you shared the name(s) of your scientific mentors for your STS project. Please provide their contact information:

**if the names are not appearing correctly, please leave the mentor names blank, reference Task 9 and enter the contact information for the mentor(s) you listed in slots 1, 2, and/or 3.

Mentor/Supervisor Name 1	Mark Hannum
Phone Number of Mentor/Supervising Adult 1	(No response)
Email Address of Mentor/Supervising Adult 1	MSHannum@fcps.edu
Mentor/Supervisor Name 2	Adrian Dalca
Phone Number of Mentor/Supervising Adult 2	(No response)
Email Address of Mentor/Supervising Adult 2	adalca@mit.edu
Mentor/Supervisor Name 3	
Phone Number of Mentor/Supervising Adult 3	(No response)
Email Address of Mentor/Supervising Adult 3	(No response)

Do you have any patents related to your research?

I do not have a patent for my STS research, but I do have a patent or patent pending for a different project.

Select all that apply

Who has been the most positive influence on your scientific endeavors?

We will send a letter to this person to tell them that you chose them and to thank them for encouraging and assisting you.

Most Influential Person Prefix (Dr., Ms., etc.)	Mr.
Most Influential Person First Name	Will
Most Influential Person Last Name	Ramey
Most Influential Person Address	2788 San Tomas Expressway
Most Influential Person Address 2	(No response)
Most Influential Person City	Santa Clara
Most Influential Person Country	United States
Most Influential Person State	CA
Most Influential Person County Name	Santa Clara
Most Influential Person Zip Code	95050
Most Influential Person E-mail Address	WRamey@nvidia.com
How do you know this person? (maximum 75 words)	Mr. William Ramey is one of the most inspirational people I have met, and has helped me in numerous occasions when I have given him a very short time to respond. I am forever indebted to what he has done for me.

**Do you intend to major in a
STEM-related field in college?**

Yes

Task 13: Optional Research Study

Last updated: 11/11/2018

Optional Study

Optional Research Study

Regeneron and the Society for Science & the Public are doing some research on the types of students who participate in the Regeneron STS. Your participation in this research study is voluntary and will not affect how your application is assessed. You may choose not to participate. Your responses will be confidential and kept separate from the application material; they will not be shared with judges at any stage of competition or review. All data are stored in a password protected electronic format. The questions in this part of the survey should take less than five minutes to complete.

If you agree to participate in this research, please check the box below.

Question 1:

Which, if any of the following STEM (science, technology, engineering or mathematics) competitions did you participate in during high school? Please check as many as apply.

Intel International Science and Engineering Fair (ISEF)

Google Science Fair

Science Olympiad

Davidson Fellows Scholarship

Exploravision

A science fair within your own school

A district, regional, or state science fair

Question 2:

How did you learn about the Regeneron Science Talent Search (STS) competition?

From a teacher or school administrator

Question 3:

How much do you agree or disagree with each of the following statements (Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree)

I am interested in research	Strongly Agree
I am a skilled researcher	Somewhat Agree
I intend to study science, technology, engineering or math in college or graduate school	Strongly Agree
I intend to have a career that involves science, technology, engineering or math	Strongly Agree
I intend to have a career that involves research	Strongly Agree
I have participated in STEM activities outside of school	Strongly Agree
I have mentored others in STEM topics	Somewhat Agree

Question 4:

Please rank the three biggest benefits you think you may get from entering or participating in Regeneron STS. Drag the biggest benefit into choice 1, the second into choice 2, and the third into choice 3.

Ability to share research with others	Choice 3
Access to leaders in the science, math, and engineering	(No response)
Access to the competition's alumni network	Choice 1
Connection to a possible internship or job	(No response)
Improved presentation skills	(No response)
Improved research skills	Choice 2
Improved writing skills	(No response)
More competitive college applications with competition experience added	(No response)
More competitive resume with competition experience added	(No response)
Non-monetary prizes	(No response)
Opportunity to travel to Washington DC for Regeneron STS Finalists Week	(No response)
Prestige associated with the competition	(No response)
Prize money	(No response)
Strong friendships with other research students	(No response)
Other	(No response)

Question 5:

To help us understand the types of people that participate in the competition, we would like to associate some of the information you provided in the application itself with your answers to these research questions. Please indicate if you are comfortable having it associated with your answers here. This is optional, it will not affect what the judges see, and your data will remain confidential.

No, do not associate

Question 6:

May we contact you in the future with a follow-up survey to further this research? Note that no matter how you answer, you are not obligated to talk with us in the future and you may withdraw from the research at any time.

Yes, you may contact me