Heart Health

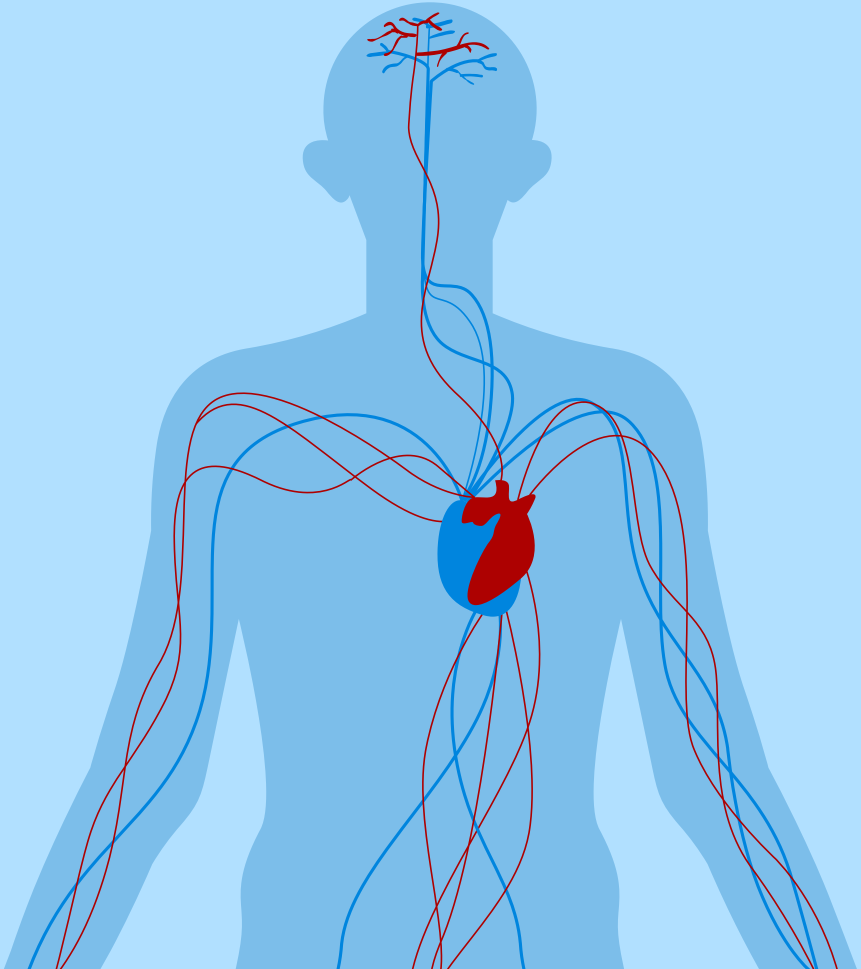
Using Machine Learning to calculate heart attack risk and enhance detection

Parith Avasadanond, Ryoji Ryan Loh, Zhang Zhihe

TABLE OF CONTENTS

01

Introduction



Our Team

This is an ongoing passion project of a group of students from Anglo Chinese School (Independent).

Originally submitted for Singapore Science and Engineering Fair and IDEX 2024, this has evolved into a R&D project aimed at benefitting Singaporeans by making Heart Attack detection more accurate and accessible than ever.

As a team of 3, we are hopeful to work with organizations and governmental bodies, to improve our product by tapping into their experience, to eventually integrate it into the Singapore healthcare system for the good of Singaporeans.

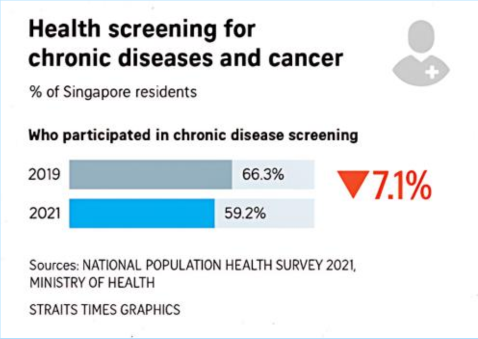
The Problem

Cardiovascular diseases have been described as a "notorious silent killer" by Health Minister Ong Ye Kung.

Singapore has seen a drop in Chronic Disease Screenings (National Population Health Survey 2021), with only 59.2% of residents participating in 2021 compared to 66.3% in 2019.

This means that 1 in 3 Singaporeans may have underlying heart diseases that remain undetected due to concerns over affordability, accessibility and accuracy.

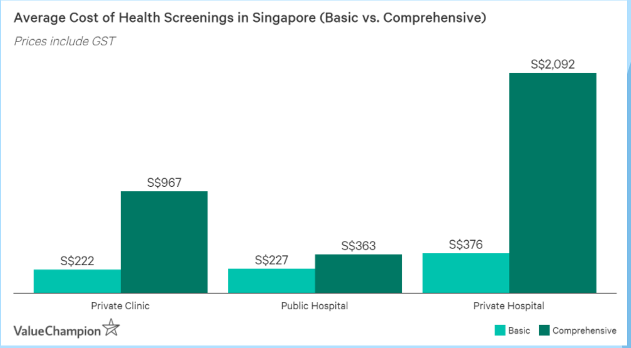
On average, 34 Singaporeans die every day from heart attacks.



The Problem

Unfortunately, there is no intermediary point to assess your heart's health than to see a doctor for ECG, MRI or CT scans to be conducted on you.

The problem is that these tests tends to be costly yet yields normal results for majority of the population.

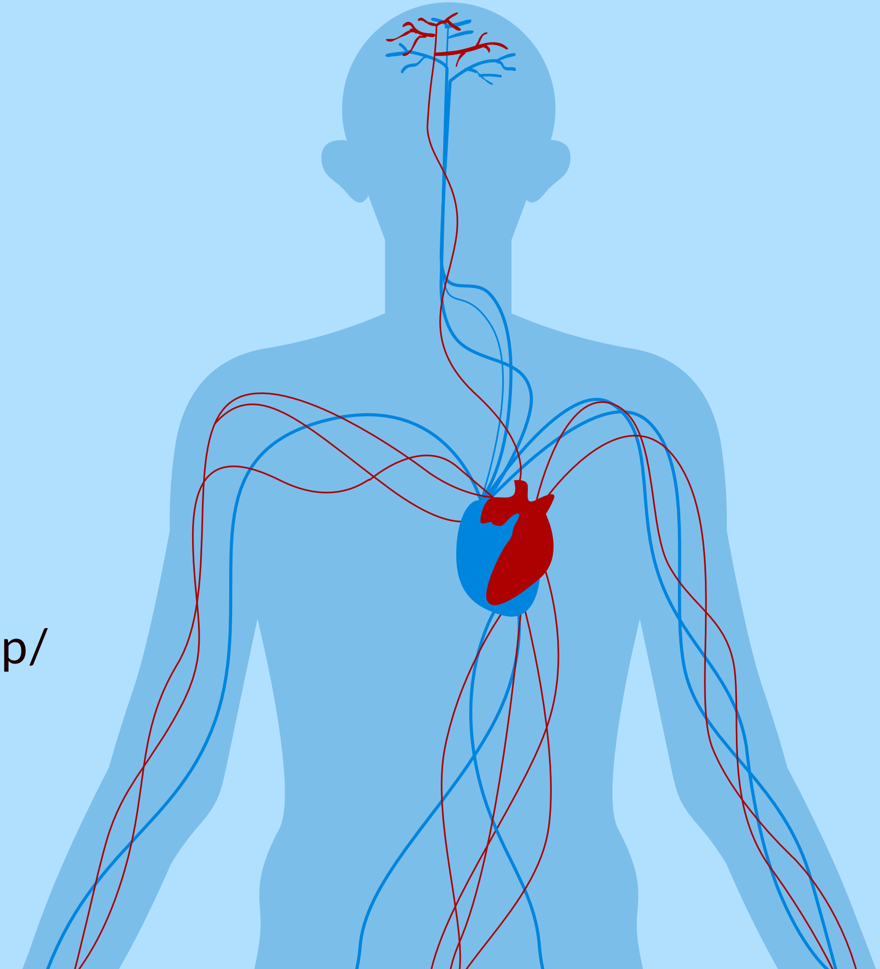


There is hence a need for an intermediary detection and referral system that is **accurate, affordable and accessible.**

02

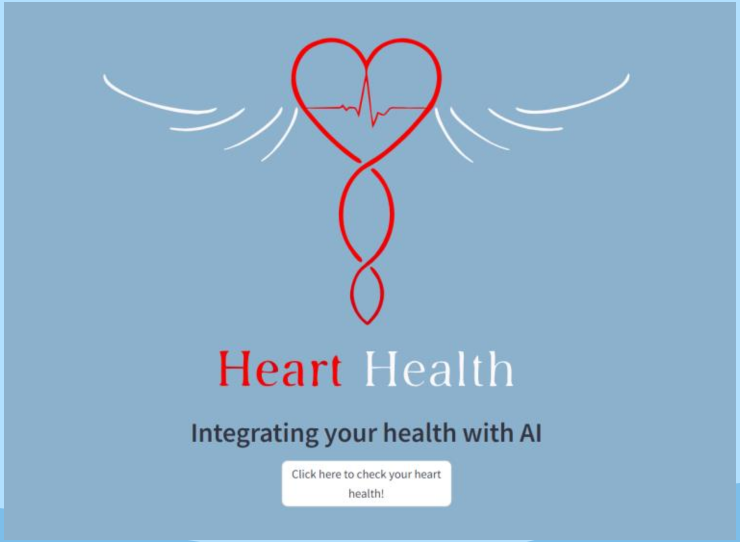
Our Product

https://hearthealth-acsi.streamlit.app/



Our Product

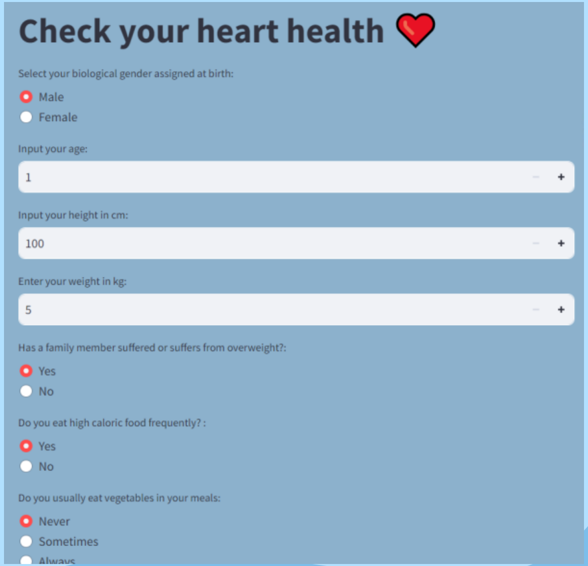
A website with questions about your lifestyle to determine your vulnerability to heart attacks, backed by analysis driven by machine learning on real life data.



Our Product

A website with 16 **questions about your lifestyle** to determine your vulnerability to heart attacks, backed by analysis driven by machine learning on real life data.

1. Select your biological gender assigned at birth
2. Input your age
3. Input your height in cm
4. Enter your weight in kg
5. Has a family member suffered or suffers from overweight?
6. Do you eat high caloric food frequently?
7. Do you usually eat vegetables in your meals
8. How many main meals do you have daily?
9. Do you eat any food between meals?
10. Do you smoke?
11. How much water do you drink daily?
12. Do you monitor the calories you eat daily?
13. How often do you have physical activity in a week?
14. How much time do you use technological devices such as cell phone, videogames, television, computer and others in a day?
15. Do you drink alcohol?
16. Which transportation do you usually use?

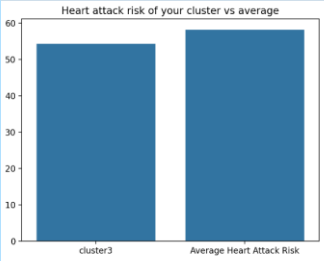


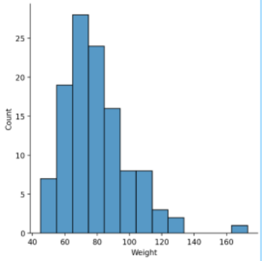
Our Product

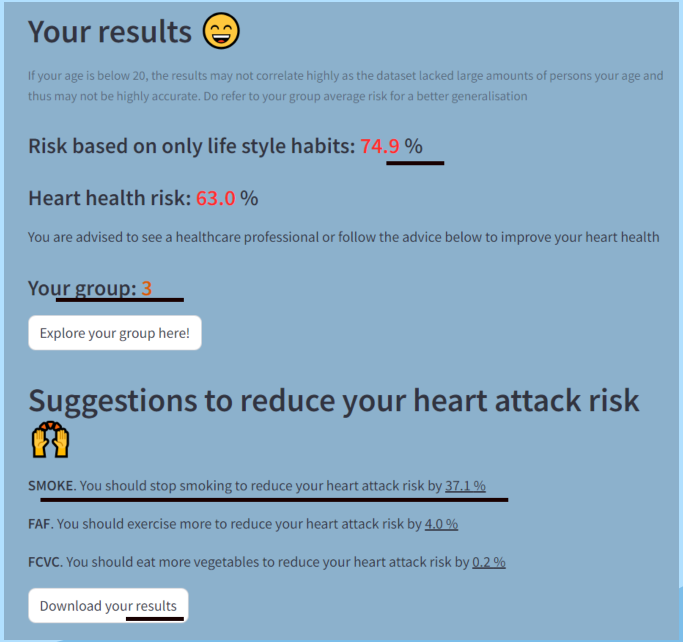
A website with 16 questions about your lifestyle to **determine your vulnerability to heart attacks**, backed by analysis driven by machine learning on real life data.

Results generated from our website:

1. Percentage Risk of Heart Attack
2. Advice according to results (e.g. Stop Smoking, More Exercise, See healthcare providers)
3. "Clusters" of people with similar lifestyles you belong to
4. Statistics on your "cluster" / group



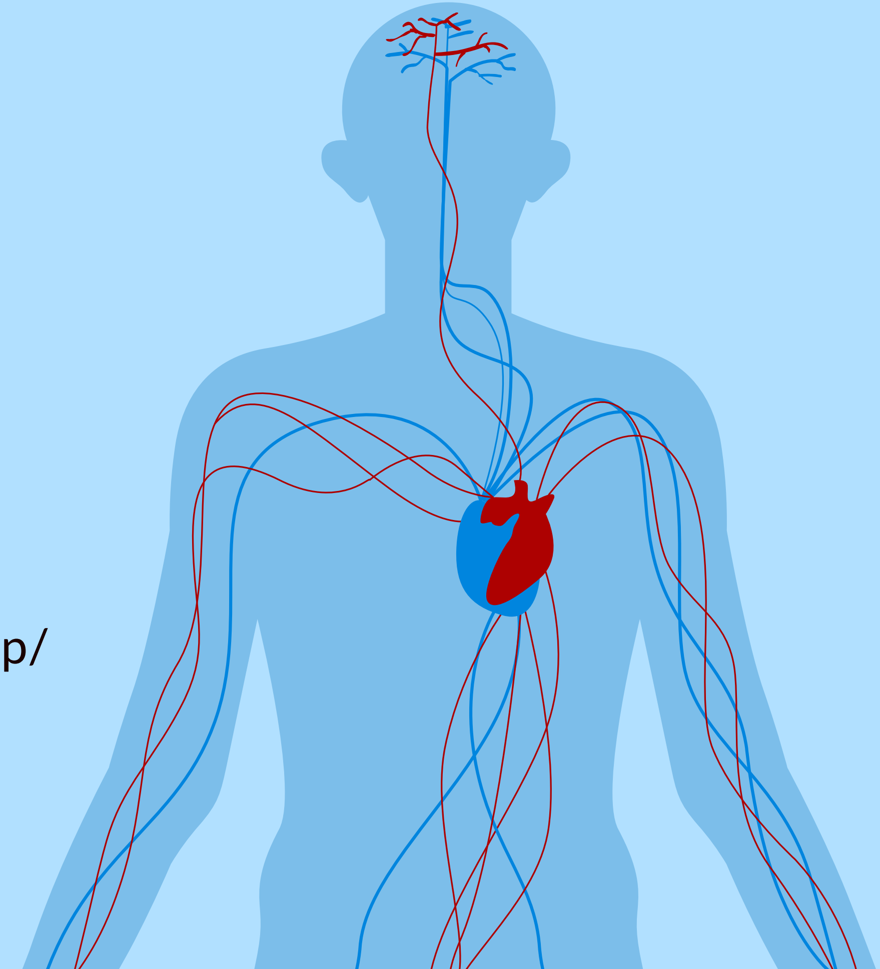




03

Our AI Model

https://hearthealth-acsi.streamlit.app/



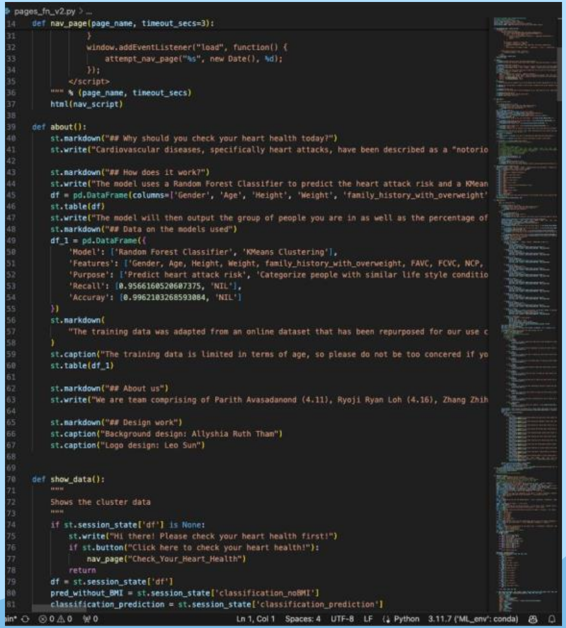
Our AI Model

A website with 16 questions about your lifestyle to determine your vulnerability to heart attacks, <u>**backed by analysis driven by machine learning**</u> on real life data.

There were two forms of machine learning we used - supervised learning and unsupervised learning.

<u>Unsupervised Learning (KMeans algorithm)</u>

Firstly, random features from the data set were selected as the initial centroids (middle of a cluster). The Euclidian distance of each point in the dataset with reference to the cluster centroids was calculated. Each data point was then assigned to the nearest centroid by taking the distance between the centroid and the data point. We recalculated the centroid by taking the average values of the data points. We repeated this process and plotted the distances on a graph, before using the elbow method to identify the optimal total number of clusters.



Our AI Model

A website with 16 questions about your lifestyle to determine your vulnerability to heart attacks, **backed by analysis driven by machine learning** on real life data.

Supervised Learning

We conducted a survey on the following machine learning models for the project. To understand the intuitions behind the models, refer to appendix.

Decision Trees / 2. Random Forest Ensemble / 3. Support Vector Machine / 4. Gradient Boosting / 5. Naïve Bayes / 6. Artificial Neural Network

Firstly, we had to instantiate the model, before running it. We then store the AUC, recall, precision, recall, F1 score and time taken (more details in the subsequent section). We then repeated for a range of hyperparameters, which are parameters that control the learning process and determine the values of model parameters that a learning algorithm ends up learning. (more details can be found in the appendix). We then ranked the data according to the highest values in the metrics as well as the fastest time.

Our AI Model

A website with 16 questions about your lifestyle to determine your vulnerability to heart attacks, **backed by analysis driven by machine learning** on real life data.

Results

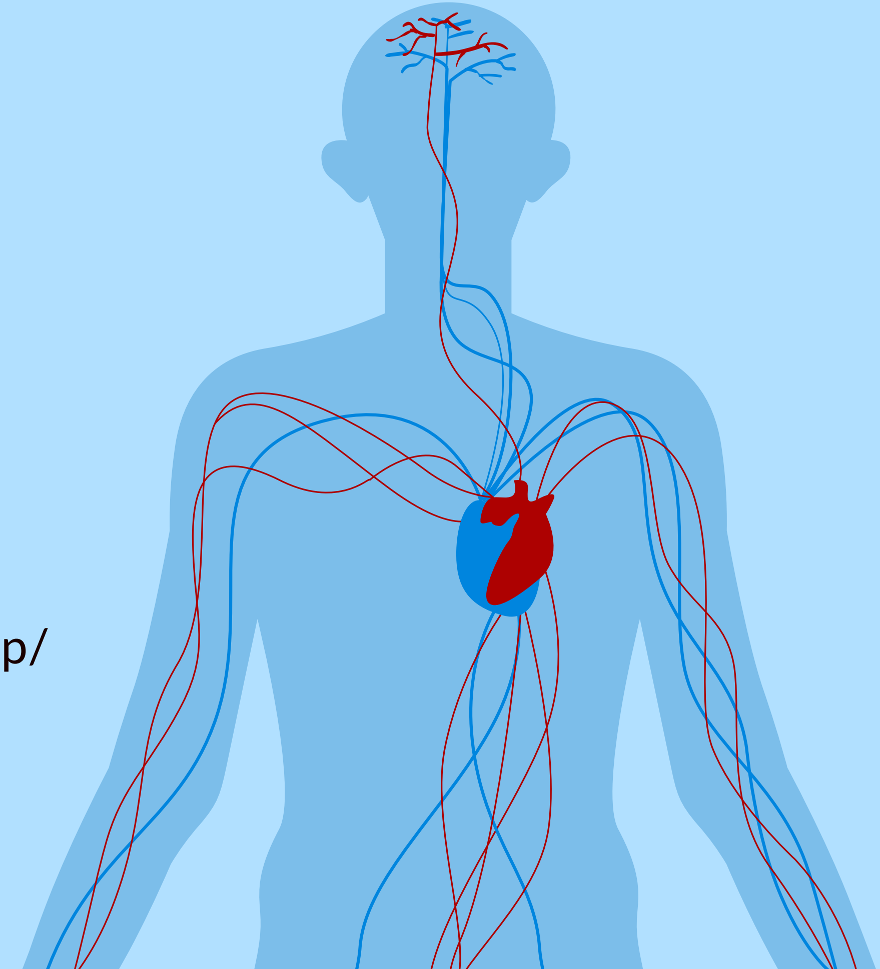
Architecture	AUC (%)	Recall (%)	Precision (%)	Accuracy (%)	F1 score (%)	Time taken for training (s)
Random Forest	98.7	100	98.4	99.0	99.2	0.085
SVM	92.4	95.1	93.5	93.0	94.3	0.0065
Gradient Boosting	50.0	100	61.0	61.0	75.8	0.037
Naive Bayes	97.4	100	96.8	98.0	98.4	0.00096
Neural Network	50.0	100	61.0	61.0	75.8	NA

The Random Forest Architecture yielded the best results in both datasets as it has a high accuracy (99%) as well as high scores in other metrics while sacrificing only slight inference speed.

04

Our Dataset

https://hearthealth-acsi.streamlit.app/



Our Dataset

Data

Gender	Age	score	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	score	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad	BMI	BMI STATUS		MALE	FAVC = yes	FCVC 1/=1 : never	SMOKE = yes	CALC
Female	21	-	1.62	64	yes	no	2	3	Sometimes	no	0	2	no	0	1	no	Public_Transportation	Normal_Weight	24.38652644	normal						
Female	21	-	1.52	56	yes	no	3	3	Sometimes	yes	3	3	yes	3	0	Sometimes	Public_Transportation	Normal_Weight	24.23822715	normal					/	
Male	23	-	1.8	77	yes	no	2	3	Sometimes	no	0	2	no	2	1	Frequently	Public_Transportation	Normal_Weight	23.7654321	normal		/				/
Male	27	-	1.8	87	no	no	3	3	Sometimes	no	0	2	no	2	0	Frequently	Walking	Overweight_Level_I	26.85185185	overweight		/				/
Male	22	-	1.78	89.8	no	no	2	1	Sometimes	no	0	2	no	0	0	Sometimes	Public_Transportation	Overweight_Level_II	28.34238101	overweight		/				
Male	29	-	1.62	53	no	yes	2	3	Sometimes	no	0	2	no	0	0	Sometimes	Automobile	Normal_Weight	20.19509221	normal		/	/			
Female	23	-	1.5	55	yes	yes	3	3	Sometimes	no	0	2	no	1	0	Sometimes	Motorbike	Normal_Weight	24.44444444	normal			/			
Male	22	-	1.64	53	no	no	2	3	Sometimes	no	0	2	no	3	0	Sometimes	Public_Transportation	Normal_Weight	19.70553242	normal		/				
Male	24	-	1.78	64	yes	yes	3	3	Sometimes	no	0	2	no	1	1	Frequently	Public_Transportation	Normal_Weight	20.19946976	normal		/	/			/
Male	22	-	1.72	68	yes	yes	2	3	Sometimes	no	0	2	no	1	1	no	Public_Transportation	Normal_Weight	22.98539751	normal		/	/			
Male	26	-	1.85	105	yes	yes	3	3	Frequently	no	0	3	no	2	2	Sometimes	Public_Transportation	Obesity_Type_I	30.67932798	obese class I		/	/			
Female	21	-	1.72	80	yes	yes	2	3	Frequently	no	0	2	yes	2	1	Sometimes	Public_Transportation	Overweight_Level_II	27.04164413	overweight			/			
Male	22	-	1.65	56	no	no	3	3	Sometimes	no	0	3	no	2	0	Sometimes	Public_Transportation	Normal_Weight	20.56932966	normal		/				
Male	41 [5]	5	1.8	99	no	yes	2	3	Sometimes	no	0	2	no	2	1	Frequently	Automobile	Obesity_Type_I	30.55555556	obese class I		/	/			/
Male	23	-	1.77	60	yes	yes	3	1	Sometimes	no	0	1	no	1	1	Sometimes	Public_Transportation	Normal_Weight	19.15158479	normal		/	/			
Female	22	-	1.7	66	yes	no	3	3	Always	no	0	2	yes	2	1	Sometimes	Public_Transportation	Normal_Weight	22.83737024	normal						
Male	27	-	1.93	102	yes	yes	2	1	Sometimes	no	0	1	no	1	0	Sometimes	Public_Transportation	Overweight_Level_II	27.38328546	overweight		/	/			
Female	29	-	1.53	78	no	yes	2	1	Sometimes	no	0	2	no	0	0	no	Automobile	Obesity_Type_I	33.32051775	obese class I			/			
Female	30 [0]	0	1.71	82	yes	yes	3	4	Frequently	yes	3	1	no	0	0	no	Automobile	Overweight_Level_II	28.04281659	overweight			/		/	
Female	23	-	1.65	70	yes	no	2	1	Sometimes	no	0	2	no	0	0	Sometimes	Public_Transportation	Overweight_Level_I	25.71166208	overweight						
Male	22	-	1.65	80	yes	no	2	3	Sometimes	no	0	2	no	3	2	no	Walking	Overweight_Level_II	29.38475666	overweight		/				
Female	52 [7]	9	1.69	87	yes	yes	3	1	Sometimes	yes	3	2	no	0	0	no	Automobile	Obesity_Type_I	30.46111831	obese class I			/		/	
Female	22	-	1.65	60	yes	yes	3	3	Sometimes	no	0	2	no	1	0	Sometimes	Automobile	Normal_Weight	22.03856749	normal			/			
Female	22	-	1.6	82	yes	yes	1	1	Sometimes	no	0	2	no	0	2	Sometimes	Public_Transportation	Obesity_Type_I	32.03125	obese class I			/	/		
Male	21	-	1.85	68	yes	yes	2	3	Sometimes	no	0	2	no	0	1	Sometimes	Public_Transportation	Normal_Weight	19.86851717	normal		/	/			
Male	20	-	1.6	50	yes	no	2	4	Frequently	yes	4	2	no	3	2	no	Public_Transportation	Normal_Weight	19.53125	normal		/			/	
Male	21	-	1.7	65	yes	yes	2	1	Frequently	no	0	2	no	1	2	Always	Walking	Normal_Weight	22.49134948	normal		/	/			/
Female	23	-	1.6	52	no	yes	2	4	Frequently	no	0	2	no	2	1	Sometimes	Automobile	Normal_Weight	20.3125	normal			/			
Male	19	-	1.75	76	yes	yes	3	3	Sometimes	no	0	2	yes	3	1	Sometimes	Public_Transportation	Normal_Weight	24.81632653	normal		/	/			
Male	23	-	1.68	70	no	yes	2	3	Sometimes	no	0	2	no	2	2	Frequently	Walking	Normal_Weight	24.8015873	normal		/	/			/
Male	29	-	1.77	83	no	yes	1	4	Frequently	no	0	3	no	0	1	no	Motorbike	Overweight_Level_I	26.49302563	overweight		/	/	/		
Female	31 [0]	0	1.58	68	yes	no	2	1	Sometimes	no	0	1	no	1	0	Sometimes	Public_Transportation	Overweight_Level_II	27.23922448	overweight						
Female	24	-	1.77	76	no	no	2	3	Sometimes	no	0	3	no	1	1	Sometimes	Walking	Normal_Weight	24.25867407	normal						
Male	39 [2]	2	1.79	90	no	no	2	1	Sometimes	no	0	2	no	0	0	Sometimes	Public_Transportation	Overweight_Level_II	28.08901095	overweight		/				
Male	22	-	1.65	62	no	yes	2	4	Frequently	no	0	2	no	2	0	Sometimes	Public_Transportation	Normal_Weight	22.77318641	normal		/	/			

Our Dataset

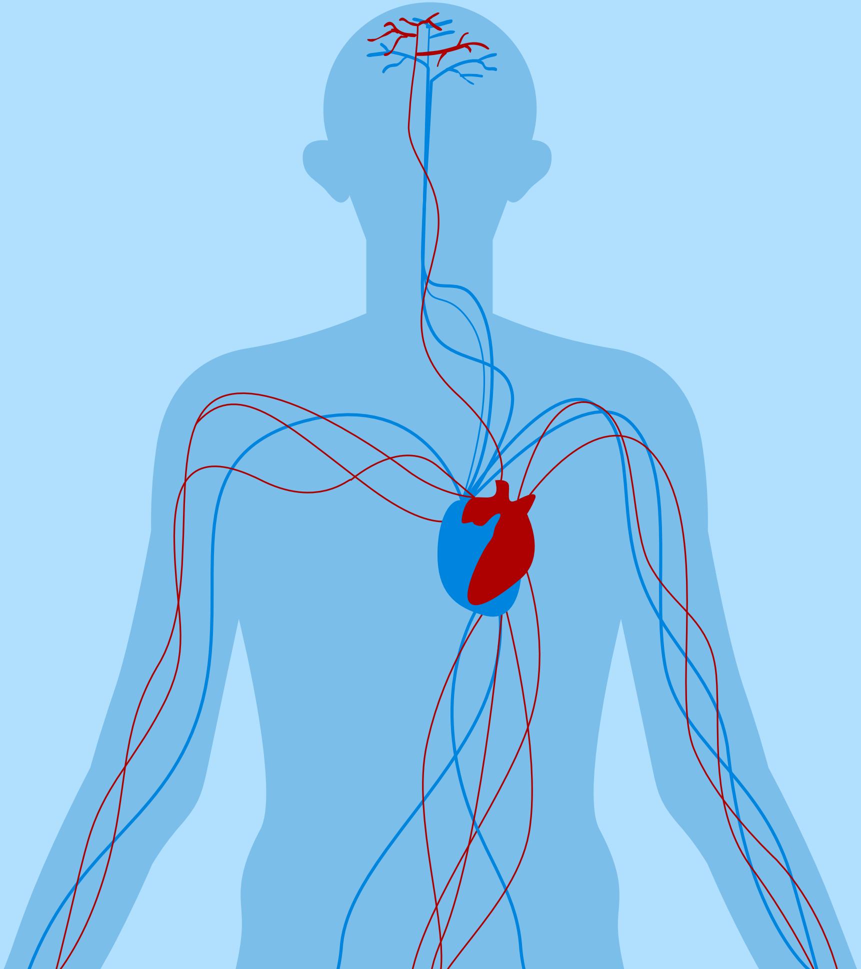
For our niche use case, there has not been much data collected regarding heart attack risk correlation with lifestyle habits.
Hence, we repurposed a popular obesity dataset from Mexico, Peru and Colombia.

This was done with the help of medical students which identified persons at risk. After that, we used different medical students to cross validate the data, on top of that we used the Cohen Kappa score, which is a score used to quantify the level of agreement between judges, which resulted in an accuracy score of 95%, thereby ensuring the validity and accuracy of our dataset. To be on the safe side, recall is at 100%, which means that there is no case of false negatives (Model predicting low risk even though you have high risk) occurring.

This is something achievable within Singapore as there are already ongoing efforts such as Singapore's project RESET by NUS of screening 10,000 Singaporeans.

05

Impact



Impact

01

LOW-RISK

≤ 30%

For low-risk users, our machine learning model can provide them with confidence that their current lifestyle habits are effective in mitigating heart attack risk and help further optimize their lifestyle.

02

MEDIUM-RISK

31%-60%

For medium-risk users, our machine learning model can provide them with effective solution and lifestyle adjustments they can make to mitigate their heart attack risk.

03

HIGH-RISK

61%-99%

For high-risk users, our machine learning model can identify patients who otherwise may not have noticed underlying heart issues and recommend them to see healthcare professionals.

Impact in Singapore

We aim for this to be the intermediary first detection and referral point in Singapore's healthcare system.

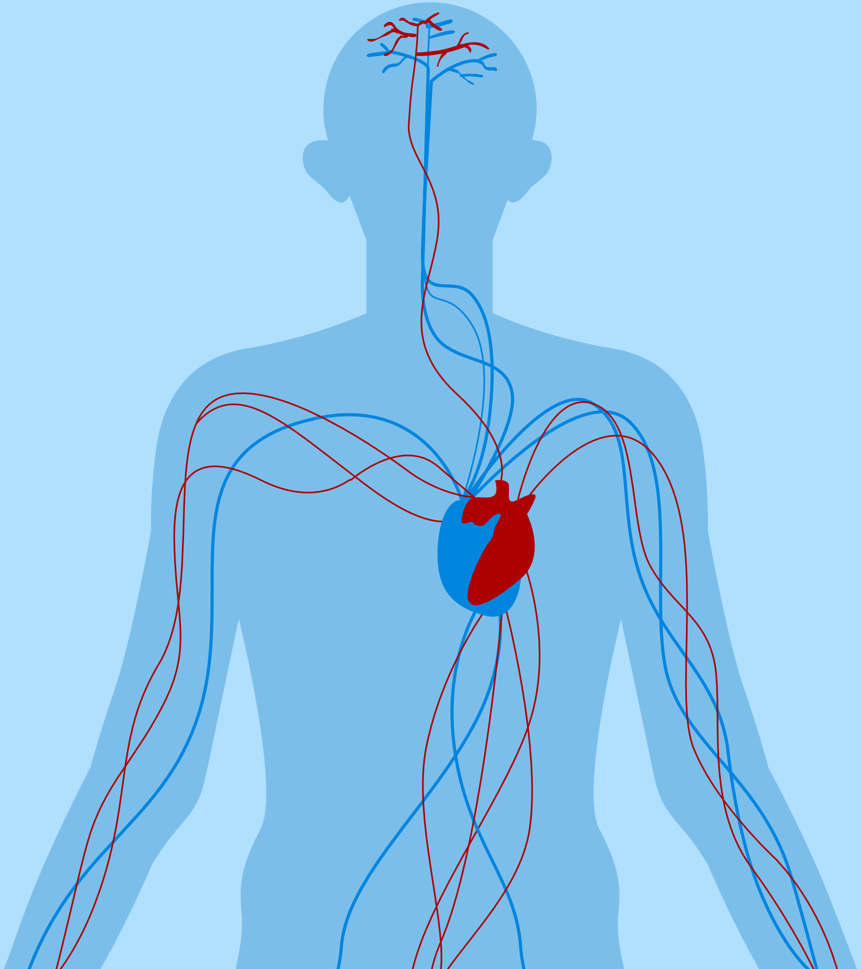
Current Alternatives that are available on the market have two major flaws:

1. Most of them have a high barrier of entry, requiring advanced and costly medical imaging technology such as MRI or CT scans.
2. Many websites raised concerns on personal data privacy. Many required excessive amounts of personal information, creating issues on data protection.

Hence, we chose to take a novel approach to this problem.
Our machine learning model was trained on lifestyle questions, creating a lower barrier of entry for our users. Additionally, we integrated our model into a frictionless website not requiring any personal information or details from the user, enabling us to dissociate the information provided from the user. Data is also cleared each time the page is reloaded, providing an additional layer of security. Our application then provides users with their heart attack risk, possible lifestyle changes to minimize heart attack risk.

06

Extensions



Accessibility

LIMITATIONS

1. Poor integration in the user’s life, thus making it hard for users to constantly track their heart health
2. Given the quiz-based nature of our website, some users may forget to recheck their heart health
3. Questions may be vague leading to inaccurate answers – Some of the answers to the quiz are answered relative to what the user thinks it means, e.g. the definition of “sometimes”

SOLUTION

1. We could integrate with current apps such as ActiveSG or Doctor Anywhere
2. We are also currently developing an IOS version of our website

- Key features to improve accessibility are currently being implemented
- Being able to track the user's sleep and steps through synchronisation with IOS's health app
- Taking pictures of food and water and automatically determining the amount of food and water taken per day
- More personalised features such as widgets to remind the user about their step goal and more
- Multi language processing

Data & Integration

LIMITATIONS

1. Our dataset is not specialized for heart conditions
   - Due to lack of datasets, we are using a retro-fitted dataset that has not been validated using real world data
   - Reliant on educated guesses from medical students which may not be completely accurate
2. Small sample size with little variation
   - Our current dataset comes from South America, which will have different lifestyle conditions than Singapore. This would result in it being less accurate for Singaporeans as compared to South Americans.

SOLUTION

1. Collaboration with healthcare professionals – Healthcare professionals in cardiovascular health would enable us to have more effective questions in our questionnaire, improving our accuracy
2. Collaboration with Singaporean health agencies – This would allow us to acquire and process data specifically for Singaporeans, which would allow us to tailor our machine learning model better to the unique Singaporean context, making it more accurate and precise for Singaporeans to use
3. More customized data for healthcare data specific for their healthcare. For example, when providing details in health declaration form, we can use this product to aid in data collection.

THANKS!

For queries, contact:
21parith.avasadanond@acsians.acsi.edu.sg
21ryoji.ryan.loh@acsians.acsi.edu.sg
21zhang.zhihe@acsians.acsi.edu.sg