Screening for Non-Alcoholic Fatty Liver Disease Using Anthropometric Indices and Machine Learning Techniques

KEYUR PARMAR
60003210219

SARTHAK MAHALE
60003210190

ANUPKUMAR SINGH
60003210202

Under the guidance of

Prof. Richa Sharma
Project Guide

Department of Information Technology
DJSCE
Mumbai University

2022-23

Overview

- Introduction
- Related Works
- Proposed Algorithm
- Conclusion
- References

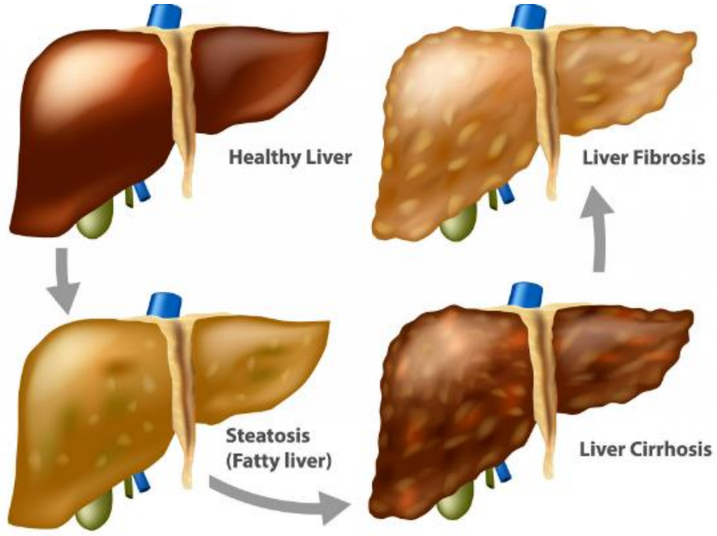
Introduction

- An overview of a study that explores the use of anthropometric indices and machine learning techniques to screen and predict stages of nonalcoholic fatty liver disease.
- Non-Alcoholic Fatty Liver Disease (NAFLD) is a condition characterized by the accumulation of fat in the liver, not caused by alcohol consumption
- Becoming increasingly prevalent worldwide and considered a common chronic liver disease.
- There are 3 Major Phases of NAFLD: Steatosis, Fibrosis and Cirrhosis.

- **Steatosis:** a condition characterized by the accumulation of fat within liver cells.

  Fibrosis: the formation of excess connective tissue in an organ or tissue, often as a result of inflammation or injury.

  Cirrhosis: A late-stage liver disease in which healthy liver tissue is replaced by scar tissue, leading to liver failure.



Problem Statement

- Non-alcoholic fatty liver disease (NAFLD) is a prevalent liver disorder characterized by the accumulation of excess fat in the liver of individuals who consume little to no alcohol. Early detection and prediction of NAFLD are crucial for timely intervention and prevention of disease progression. Anthropometric variables, which are measurements of the human body, have been identified as potential predictors of NAFLD. The problem at hand is to develop a predictive model that utilizes anthropometric variables to accurately predict the presence or likelihood of NAFLD in individuals.

- Objective: The objective is to build a machine learning model that takes into account various anthropometric measurements and predicts the probability or presence of NAFLD in individuals. The model should provide a reliable and efficient means of early detection and risk assessment for NAFLD, facilitating timely intervention and preventive measures.

Literature Review

No.	Research Paper	Year of publication	Author name	Observation and Conclusion
1	Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices.	2023	Farkhondeh Razmpour, Reza Daryabeygi-Khotbehsara , Davood Soleimani , HamzehAsgharnezhad, Afshar Shamsi, Ghasem Sadeghi Bajestani, Mohsen Nematy, Mahdiyeh Razm Pour , Ralph Maddison & Sheikh Mohammed Shariful Islam	The main aim of this study was to apply machine learning (ML) methods to identify signifcant classifers of NAFLD using body composition and anthropometric variables. A cross-sectional study was carried out among 513 individuals aged 13 years old or above in Iran.
2	Learning High-dimensional Associations for Nonalcoholic Fatty Liver Disease Diagnosis Prediction.	2022	Zhijin Wang College of Computer Engineering Jimei University Xiamen, China, Bing Cai College of Computer Engineering	This paper proposed a multi-scaled CNN diagnosis by solely using physical examination data.
3	What's new in non-alcoholic fatty liver disease	2022	Jessica Spiers, James Hallimond Brindley, Wenhao Li, William Alazavi	This paper gave the gist of NAFLD and why is it caused and factors responsible for it.

Literature Review

No.	Research Paper	Year of publication	Author name	Observation and Conclusion
4	Artificial Intelligence for Detecting and Quantifying Fatty Liver in Ultrasound Images	2019	Fahad Muflih Alshagathrh and Mowafa Said Househ	AI-supported systems show potential performance increases in detecting and quantifying steatosis, NASH, and liver fibrosis in NAFLD patients.
5	The value of combining the simple anthropometric obesity parameters, Body Mass Index (BMI) and a Body Shape Index (ABSI), to assess the risk of non-alcoholic fatty liver disease	2022	Maobin Kuang1,2, Guotai Sheng2, Chong Hu3, Song Lu1,2, Nan Peng1,2 and Yang Zou4.	This paper uses simple parameters like Body mass index and Body shape index predict NAFLD in patients.
6	Anthropometric Predictors and Artificial Neural Networks in the diagnosis of Hypertension	2022	Krzysztof Pytel University of Lodz, Faculty of Physics and Applied Informatics, Lodz, Poland, Wojciech Drygas Tadeusz Nawarycz.	This paper uses anthropome indices to predict whether a patient has hypertension usi artificial tension. The datase used had 2485 real cases of

Conventional and our proposed Method

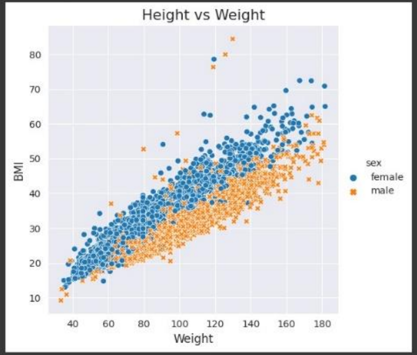
- The conventional method uses of biochemical variables in combination with other factors to predict likelihood of NAFLD like Alanine transaminase (ALT) , Aspartate transaminase (AST) etc.
- However, the conventional method is expensive and not accessible in rural areas.

Anthropometric Indices as a Solution:

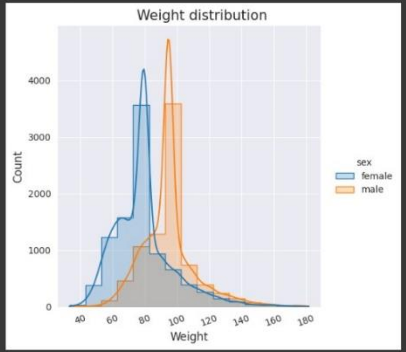
- Anthropometry is a lower cost and more feasible approach for primary screening of NAFLD.
- Anthropometric measurements include weight, height, body composition, circumferences, and subcutaneous fat.
- The stages of NAFLD can be predicted using various machine learning techniques.
- The various ML techniques used are k-Nearest Neighbour (kNN), Support Vector Machine (SVM), Radial Basis Function (RBF) SVM, Gaussian Process (GP), Random Forest (RF), Neural Network (NN), AdaBoost and Naïve Bayes.

Implementation Demo

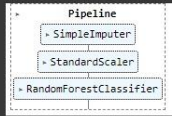
```
[31]
c = sns.relplot(data=df, x="weight_kg", y="bmi", hue = 'sex', style = 'sex')

c.fig.set_dpi(100)
c.fig.suptitle("Height vs Weight", y = 1.02, size = 15)
plt.xlabel('Weight', fontsize = 12)
plt.ylabel('BMI', fontsize = 12)

plt.show()
```




```
[29]
a = sns.displot(df, x="weight_kg", bins=15, kde=True, hue='sex', element="step")

a.fig.set_dpi(100)
a.fig.suptitle("Weight distribution", y = 1.02, size = 15)
plt.xlabel('Weight', fontsize = 12)
plt.ylabel('Count', fontsize = 12)
plt.xticks(rotation=20)

plt.show()
```




```
[ ]
#Train the model on the training data :

pipeline.fit(X_train, y_train)
```




```
[ ]
#Making predictions on the test data :

predictions = pipeline.predict(X_test)
```

```
[ ]
# User input
age = float(input("Enter age: "))
weight = float(input("Enter weight: "))
height = float(input("Enter height: "))
bmi = float(input("Enter BMI: "))
gender = int(input("Enter gender (0 for female, 1 for male): "))
```

```
Enter age: 55
Enter weight: 120
Enter height: 155
Enter BMI: 49.947970863683
Enter gender (0 for female, 1 for male): 1
```

```
[ ]
# Create a DataFrame from user input
user_data = pd.DataFrame({
    'age': [age],
    'weight': [weight],
    'height': [height],
    'bmi': [bmi],
    'male': [gender]
})
```

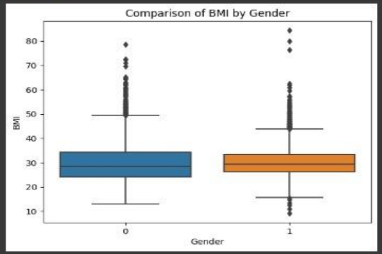
```
if predictions[0] == 1:
    print("The person is predicted to have NAFLD.")
else:
    print("The person is predicted not to have NAFLD.")
```

```
The person is predicted not to have NAFLD.
```

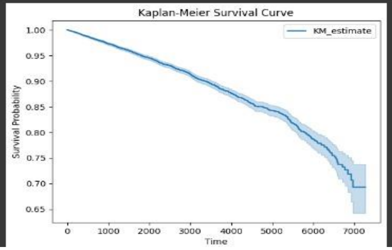
```
liver = pd.read_csv('/content/nafld1.csv')
liver.head()
```

	Unnamed: 0	id	age	male	weight	height	bmi	case.id	futime	status
0	3631	1	57	0	60.0	163.0	22.690939	10630.0	6261	0
1	8458	2	67	0	70.4	168.0	24.884028	14817.0	624	0
2	6298	3	53	1	105.8	186.0	30.453537	3.0	1783	0
3	15398	4	56	1	109.3	170.0	37.830100	6628.0	3143	0
4	13261	5	68	1	NaN	NaN	NaN	1871.0	1836	1

```
sns.boxplot(x='male', y='bmi', data=nafld1_data)
plt.xlabel('Gender')
plt.ylabel('BMI')
plt.title('Comparison of BMI by Gender')
plt.show()
```




```
# Predictive Modeling - Classification
# Use the variables of your choice to create the training and test sets,
# then train and evaluate your classification model.

# Survival analysis

kmf = KaplanMeierFitter()
kmf.fit(nafld1_data['futime'], nafld1_data['status'])
kmf.plot_survival_function()
plt.xlabel('Time')
plt.ylabel('Survival Probability')
plt.title('Kaplan-Meier Survival Curve')
plt.show()
```



Feature Variables

❑ There are various feature variables used in this project. Some of them are Age, Weight, Height, BMI and various other anthropometric measurements(Refer Fig).

❑ The following methods were used to calculate these measurements:

- Sex, age, education, disease history and medications- Questionnaire.
- Weight-Digital Weighing Scale.
- Height-Wall Height Chart.
- Circumferences-Flexible Tape Measure.

MERITS:

❑ Simple and cheap equipment.

❑ Needs less specialty knowledge therefore can be implemented in several health centres and also remote areas.

Variables
Number
Age, years
Weight; kg
Height; cm
BMI; kg m^{-2}
Arm circumference; cm
Neck circumference; cm
Chest circumference; cm
Waist circumference; cm
Abdomen circumference; cm
Hip circumference; cm
Wrist circumference; cm
Subscapular skinfold; mm
Biceps skinfold; mm
Triceps skinfold; mm
Suprailiac skinfold; mm

Previous Algorithm

- The Random Forest yielded the most accurate result for all Three Stages.
- The average accuracy and AUC value was 0.82 and 0.84 for fatty liver, 0.62 and 0.69 for steatosis stages, 0.57 and 0.58 for fibrosis stages.

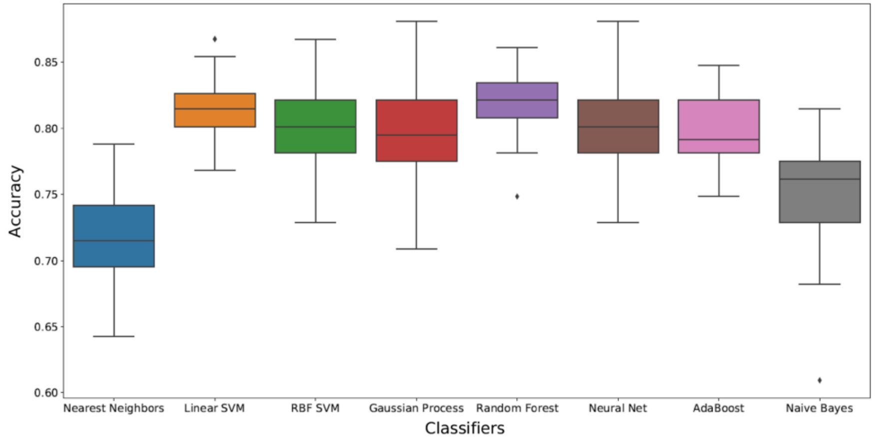


Figure 1. Box plots showing different classification methods applied to the dataset for presence of fatty liver. Box plots are generated by performing 50 individual runs for each classifier. This will assure that the achieved results are reliable.

Fatty Liver Stage

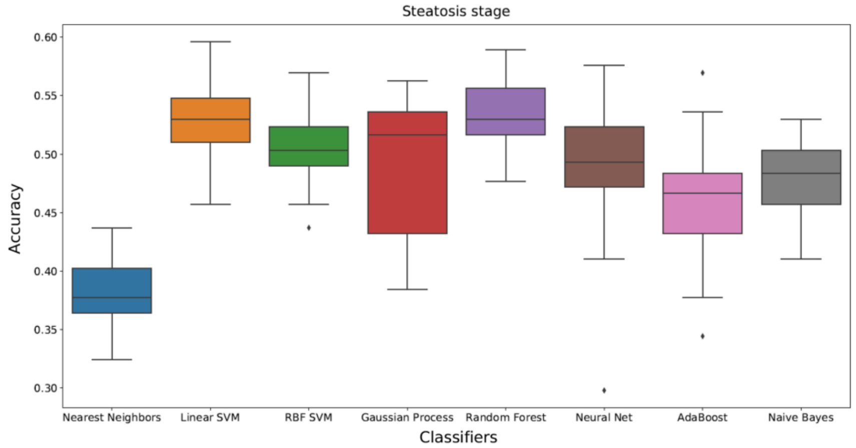


Figure 2. Box plots showing different classification methods applied to the dataset for stages of steatosis. Box plots are generated by performing 50 individual runs for each classifier. This will assure that the achieved results are reliable.

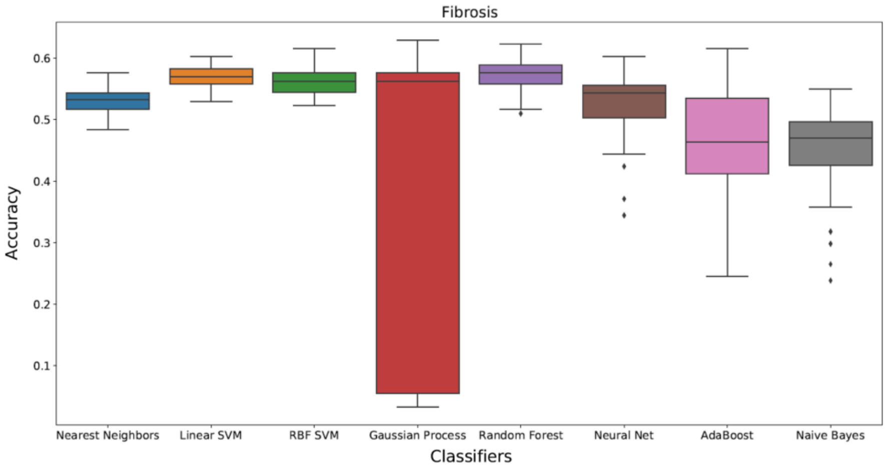


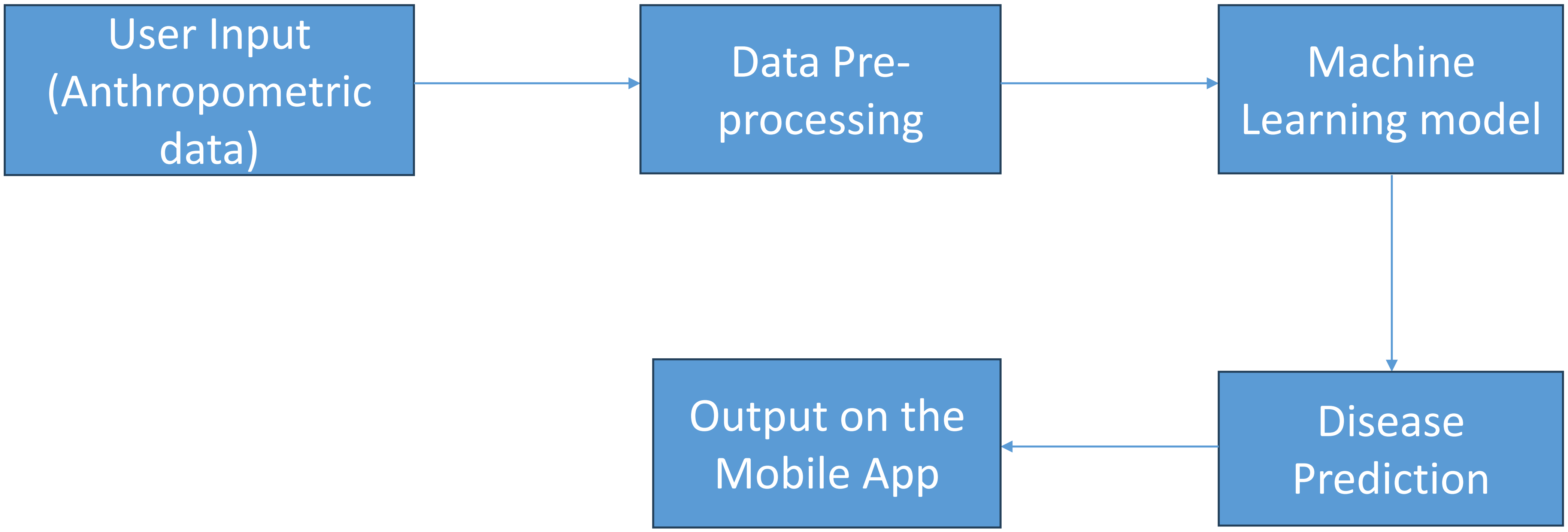
Figure 3. Box plots showing different classification methods applied to the dataset for stages of fibrosis. Box plots are generated by performing 50 individual runs for each classifier. This will assure that the achieved results are reliable.

Data Set Asked from Cooper Hospital

- We met Dr.Prasad Pandit and Dr.Chetan Phirke of the department of Pharmocology.
- They examined our problem statement and agreed to be our co-ordinators for this project.
- Parameters Requested:

Gender	Age	Diabetes History	Height	Weight
BMI(kg/m2)	Waist Circumference	Abdominal circumference(cm)	Trunk Fat(kg)	Chest Circumference(cm)
Left arm fat (kg)	Right arm fat (kg)	Total fat (%)	Neck circumference (cm)	Muscle mass (kg)
Right leg fat (kg)	Right leg muscle (kg)	Left leg fat (kg)	Left leg muscle (kg)	Right arm muscle (kg)
Left arm muscle (kg)	Mid upper arm circumference (cm)	Hip circumference (cm)	Fat mass (kg)	Neck fat
Sub-chin fat	Subscapular skinfold; mm	Biceps skinfold; mm	Suprailiac skinfold; mm	NAFLD Presence (Yes/No)

Project Architecture



Application Design

❑Here are some features that we have considered for adding to an app for predicting non-alcoholic fatty liver disease using machine learning and anthropometric/body composition indices:

- User data input: Allow users to enter their anthropometric and body composition data, including weight, height, waist circumference, BMI, and other relevant measurements.
- Machine learning model integration: Integrate the machine learning model that uses the entered user data to predict the likelihood of non-alcoholic fatty liver disease.
- Visualization of data: Display visualizations of the user data and the model predictions to help users better understand their risk of non-alcoholic fatty liver disease.
- Tracking and monitoring: Allow users to track their data over time and monitor changes in their risk of non-alcoholic fatty liver disease.
- Educational resources: Provide educational resources and information about non-alcoholic fatty liver disease, including risk factors, symptoms, and treatment options.
- Personalized recommendations: Provide personalized recommendations for users based on their risk assessment, such as diet and lifestyle referrals to healthcare professionals.

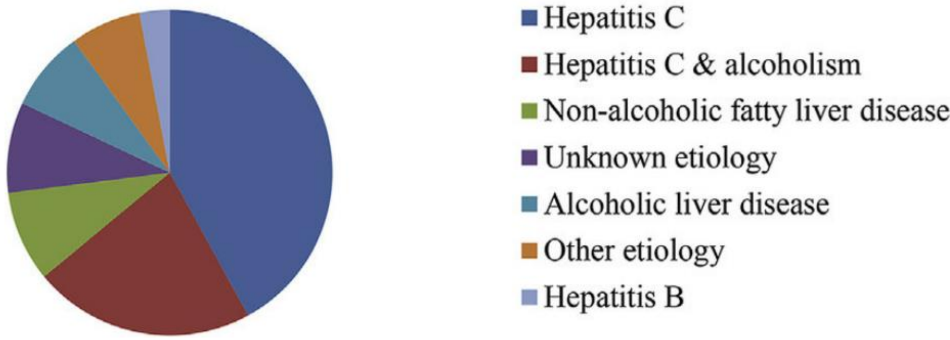
Timeline of Project

- Phase 1: Arrival of Dataset from Cooper Hospital and Data Pre-processing – September – October 2023.
- Phase 2 : Testing various Machine Learning algorithms to find out the best accuracy and fit – October-December 2023
- Phase 3 : Integrating the ML model into the designed Mobile App and Testing – January 2024 onwards.

Future Scope and Datasets

❑Future Scope-

- Expansion of the user base: As non-alcoholic fatty liver disease becomes more prevalent, there will be a greater demand for tools that can help people prevent and manage the condition. The app can be expanded to include additional features and information to appeal to a wider user base.
- Integration with electronic health records (EHRs)
- Advancement in Machine Learning Techniques will cause the accuracy of the app's predictions to improve.
- The anthropometric measurements can be done using image processing.



❑Datasets-

- Kaggle Body Fat Prediction Dataset.
- The above dataset contains age, gender, weight, height, bmi, status at the time of test(alive at last follow-up/dead).

References

❑ Research Papers:

- Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices. Farkhondeh Razmpour, Reza Daryabeygi-Khotbehsara, Davood Soleimani, HamzehAsgharnezhad, Afshar Shamsi, Ghasem Sadeghi Bajestani, Mohsen Nematy, Mahdiyeh Razm Pour , Ralph Maddison & Sheikh Mohammed Shariful Islam.
- Learning High-dimensional Associations for Nonalcoholic Fatty Liver Disease Diagnosis Prediction. Zhijin Wang College of Computer Engineering Jimei University Xiamen, China, Bing Cai College of Computer Engineering Jimei University Xiamen, China, Wen Yang Xiamen Hospital of Traditional Chinese Medicine Fujian University of Traditional Chinese Medicine Fuzhou, China.
- What’s new in non-alcoholic fatty liver disease-Jessica Spiers, James Hallimond Brindley, Wenhao Li, William Alazavi.2022
- Evaluation of automated anthropometrics produced by smartphone-based machine learning: a comparison with traditional anthropometric assessments Austin J. Graybeal1 *, Caleb F. Brandner1 and Grant M. Tinsley.
- Non-Alcoholic Fatty Liver Disease: From Pathogenesis to Clinical Impact Alfredo Caturano 1,† , Carlo Acierno 1,†, Riccardo Nevola 1 , Pia Clara Pafundi 1 , Raffaele Galiero 1 , Luca Rinaldi 1 , Teresa Salvatore 2 , Luigi Elio Adinolfi 1 and Ferdinando Carlo Sasso 1,*
- Artificial Intelligence for Detecting and Quantifying Fatty Liver in Ultrasound Images: A Systematic ReviewFahad Muflih Alshagathrh and Mowafa Said Househ*
- FITME: BODY MEASUREMENT ESTIMATIONS USING MACHINE LEARNING METHOD Sahar Ashmawia, Maram Alharbia, Ameerah Almaghrabia, Areej Alhothali

References

❑ Research Papers:

- The value of combining the simple anthropometric obesity parameters, Body Mass Index (BMI) and a Body Shape Index (ABSI), to assess the risk of non-alcoholic fatty liver disease Maobin Kuang1,2, Guotai Sheng2, Chong Hu3, Song Lu1,2, Nan Peng1,2 and Yang Zou4.
- Anthropometric Predictors and Artificial Neural Networks in the diagnosis of Hypertension. Krzysztof Pytel University of Lodz, Faculty of Physics and Applied Informatics, Lodz, Poland, Wojciech Drygas Tadeusz Nawarycz, Lidia Ostrowska-Nawarycz Medical University of Lodz, Department of Biophysics, Lodz, Poland Medical University of Lodz, Department of Epidemiology, Lodz, Poland.

Thank You!!!