College: Engineering and Information Technology
Department: Information Technology
Program: Data Analytics

Programming for Data Analytics II course Project

The Effect of Obesity on the Physical Conditions of Human Beings

Prepared by:
202011025 Maryam Ali
202011108 Rahmah Shabanh

Supervised by:
Dr. Salam Freihat

Academic Year 2021- 2022 – Fall

I. Introduction

"13% of adults in the world are obese" [6]. The high prevalence of obesity around the world has been a major concern globally! It's not just a matter of physical health, it affects individuals mentally. We, in this report will be trying to focus on those people and create a neural network algorithm that performs its best to detect the occurrence of obesity in humans as well as talk about its importance and how we can prevent it from happening with the help of our dataset. In summary, we will answer as many of these questions as possible: Can your physical conditions lead to obesity? The effect of the number and type of meals on obesity? Is Obesity something inherited? Is it important to know the number of calories contained in the meals you eat and the amount of food? Will the model we create detect if you have obesity, or not? Will using the concept of Neural Networking increase the accuracy and precision of the prediction compared to machine learning?

II. Task And Data

This report uses a dataset by the UCI machine learning that is "Estimation of obesity levels based on eating habits and physical condition" [1] from the countries of Mexico, Peru, and Colombia. This dataset has been published publicly at 27/08/2019, so it is just recent and since this data is life-oriented then we don't expect this data to change if people have the same symptoms of those we experienced on previously.

Our task in this project is to predict/detect whether a person is dealing with obesity or not from their health conditions and surrounding. Our report will also include observations and discussions regarding this topic and trying to really understand what results on a person's obesity or what are the factors that you must protect to stay away from obesity. Our data's original features:

Name	D_Type	Unique values	Description
Gender	Object	'Female', 'Male'	What is their gender?
Age	float 64	14 – 61	How old is the patient?
Height	float 64	1.45 - 1.98	What is their height?
Weight	float 64	39 – 173	What is their weight?
Family_history_with_overweight	Object	'yes', 'no'	Did their family members deal with overweight?
Frequent_consumption_of_high_caloric_food	Object	'yes', 'no'	Do they consume high amount of caloric food?
Frequency_of_consu	float 64	1 – 3	What is the rate of their vegetable's consumption?
Number_of_main_meals	float 64	1, 2, 3, 4	How many main meals do they have?
Consumption_of_food_between_meals	Object	'Sometimes', 'Frequently', 'Always', 'no'	Do they consume more food between meals?
Smoke	Object	'yes', 'no'	Do they smoke?
CH20	float 64	1 – 3	How much water do you drink daily?

Calories_consumption_monitoring	Object	‘yes’, ‘no’	Do they monitor their calories consumption?
Physical_activity_frequency	float 64	0 – 3	How frequent are they physically active?
Time_using_technology_devices	float 64	0 – 2	
Consumption_of_alcohol	Object	'no', 'Sometimes', 'Frequently', 'Always'	Do they consume alcohol?
Transportation_used	Object	'Public_Transportation', 'Walking', 'Automobile', 'Motorbike', 'Bike'	What do they use for transportation?
Type_of_obesity	Object	'Normal_Weight', 'Overweight_Level_I', 'Overweight_Level_II', 'Obesity_Type_I', 'Insufficient_Weight', 'Obesity_Type_II', 'Obesity_Type_III'	What type of obesity are they on?
BMI	float 64	12.999 – 50.812	What is they body max index?
Obesity	Object	0, 1	Are they obese?

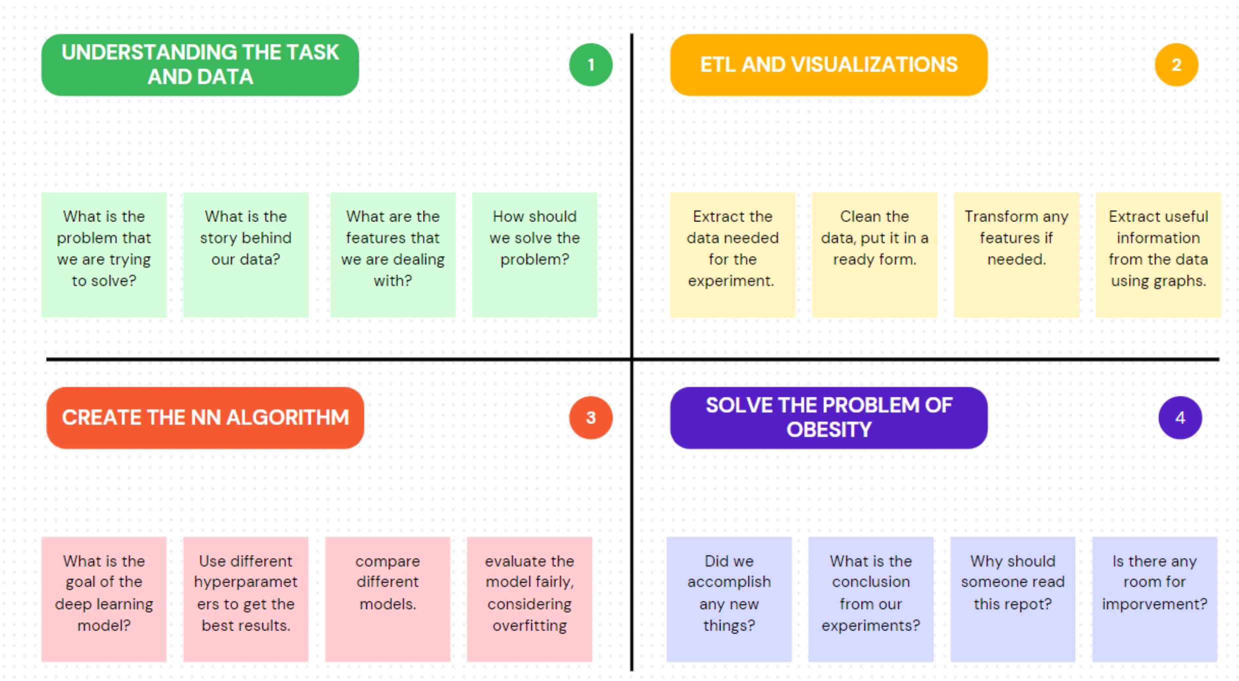
(Table 1: data description)

Note that the attributes’ information in table 1 is taken from the original source [2]

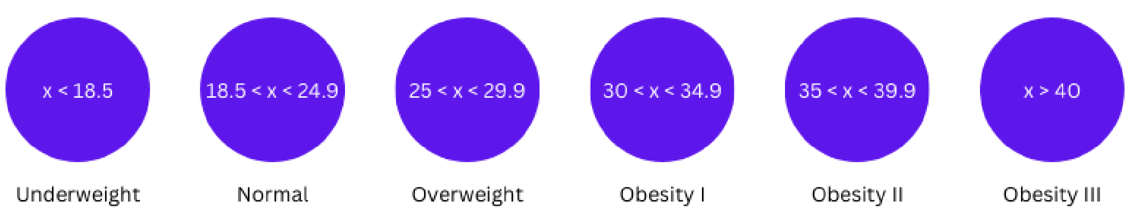
III. Methodology

Starting this project, we have had a vision of what we would like to go through and the steps to accomplishing the goal. Figure 1 shows our pipeline which demonstrating the overall idea clearly.

The methodology that we used to solve this problem goes as follows: first, it is very important to understand what you want to solve, what do you cherish and is curious about? As we concluded that a life-oriented problem is something that everybody would benefit from, and one thing that nowadays people are suffering from is obesity, which increased as technology and video games became more popular. So, we tried to find the data that would answer our questions. After finding the appropriate dataset we made sure that our data is clean performing ETL on the dataset, there are no wrong information, nor any missing or duplicated information that might affect our model’s predictions, decisions and analysis. We created new columns that we come to a decision as relevant to our experiment. Like the “obesity” column just to make things easier, narrowing the experiment to predict whether someone has obesity or not since we aren’t really interested about any of the other categories, and the “BMI” column that we learned is a big factor to determining obesity as shown in Figure 1.1 – knowing that BMI = (weight kg / height m**2) and we already had the height and weight in our dataset. We made visualizations and extracted meaningful decisions about our data. After the process of cleaning, we went to the neural network algorithm. We first encoded all the data to numerical form for the sake of the model and then built the model, trying multiple approaches changing parameters and adding any hyperparameters needed to accomplish the best result, we will be seeing every step of the experiment conducted together, later on in the report. After the process of deep learning model, we will discuss together the importance of this report and how we solved the problem of obesity and answer some questions like "is there any room for improvement?", "what are the things that we can add to get better results in the future?". And comparing our results to other people's work and model's that they created, seeing how we made improvement or not, and whether we can see ourselves making more than what we have done in this project.



(Figure 1: project's pipeline)



(Figure 1.1: BMI indicator)

IV. Motivation

What is the importance behind these experiments? Obesity is a very serious risk to individual's health, including children and adults. It impacts their life mentally like experiencing discriminations, depression, social isolation, decreases productivity, and physically as in disabilities, not being able to engage in physical activities, or could reach to the point of causing death as those people are more likely to get health problems such as type 2 diabetes, cancer of colon, liver, kidney, pancreas, breast…etc. and even severe symptoms of covid-19 [3].

There is a big difference between overweight and they are not words to use interchangeably. Obesity is "a disease marked by excessive generalized deposition and storage of fat" which cases serious health issues, while overweight is where someone weights more than their gender, age and height should yield [5].

Figure 2 and Figure 3 illustrate the top 20 observations of obesity in men and woman respectfully throughout the years [4].

This Show us how it is a very important topic to discuss, and try to share awareness, detect and make serious notice of obesity around the world. The patient should be notified that it is obesity and not something to ignore.

#	Country	% obesity
1	Nauru	59.85
2	American Samoa	58.75
3	Cook Islands	53.97
4	Palau	53.15
5	Marshall Islands	49.85
6	Tahiti (French Polynesia)	48.89
7	Tuvalu	48.47
8	Niue	46.17
9	Kiribati	42.87
10	Tonga	42.72
11	Federated States of Micronesia	41.48
12	Tokelau	41.40
13	Samoa	41.28
14	United States	36.47
15	Kuwait	34.28
16	Qatar	33.46
17	Saudi Arabia	31.73
18	New Zealand	31.07
19	Australia	30.57
20	Canada	30.47

(Figure 2: prevalence of obesity – men)

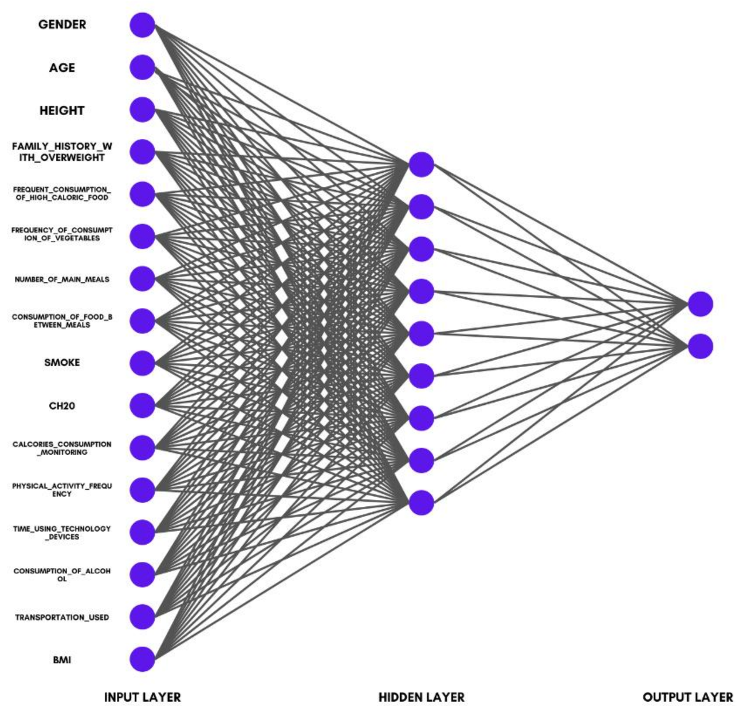
#	Country	% obesity
1	American Samoa	65.32
2	Nauru	64.81
3	Cook Islands	60.85
4	Palau	60.48
5	Marshall Islands	59.02
6	Tuvalu	57.85
7	Tahiti (French Polynesia)	56.90
8	Niue	56.77
9	Samoa	56.62
10	Tonga	56.06
11	Federated States of Micronesia	53.17
12	Tokelau	52.18
13	Kiribati	51.96
14	Kuwait	47.08
15	Qatar	44.60
16	Jordan	44.60
17	Saudi Arabia	43.74
18	Bermuda	43.17
19	Egypt	42.48
20	United Arab Emirates	42.46

(Figure 3: prevalence of obesity – women)

V. Description

As illustrated visually in figure 4, we have created a deep learning model 16 neuron input layer that takes 16 attributes' records as an input. Where each neuron has a weight corresponding to each neuron in the next layer (hidden layer) that has only 9 neurons, therefore we will be having 144 weights between the input layer and the hidden layer. We will discuss later what are the weight that we have initialized, and what are the activation functions that we implemented in the layers. After the hidden goes through the activation function built in it, the output of each neuron will be multiplied by the 2 weights each to result in the output that we will have, illustrated in 2 neurons. Either it is 1 or 0. And then test the accuracy of the model to see how well it performs, with other evaluation metrics that we also implemented.

In this deep learning model experiment we tried to minimize the number of hidden layer and the neurons as much as possible, as that will minimize the number of computations and time! luckily our dataset was performing great with only an input layer with 16 neurons, a single hidden layer with 9 neurons and an output layer with 2 neurons.



(Figure 4: neural network model illustration)

We will summarize the different approached that we used and compared in table 2. Where it shows the different parameters for each Model I, Model II, Model III and Model IV. In reality we have been trying different epoch numbers, batch size, loss functions, metrics, and optimizers but we haven't shown that for the sake of the length of this report, but we have gathered the closest comparison after choosing the best epochs, batch size, and compilation method from observations and experiments.

However, we explored further with the activation function used in the output layer, we used softmax in ModelII and ModelIV, meaning it will output a probability of this record X to either be of class 0 (no obesity) or class 1 (obesity). ModelI and ModelIII uses sigmoid function which outputs something between 0 and 1. We have decided for the output layer to have two neuron and used sparse categorical cross entropy as it performs better that way and logically it is not wrong since we have two possible outputs – nevertheless, we will be discussing why we have not used 1 neuron.

Notice that in these 4 models we have used early stopping. Early stopping stops the model as soon as it detects a fall, so it takes that highest point before the downfall. Therefore, we end up with the best weights for our architecture. And this approach will also prevent us form entering the overfitting phase (we will be seeing more of that in the results)

<table>
<tr><th rowspan="2">Model ID</th><th rowspan="2">Accuracy</th><th colspan="2">Activation function</th><th rowspan="2">No. epoch, batch Size</th><th colspan="3">Compilation</th><th rowspan="2">Kernel initializer</th><th rowspan="2">Time</th></tr>
<tr><th>Hidden layer</th><th>Output</th><th>optimizer</th><th>loss</th><th>metrics</th></tr>
<tr><td>ModelI</td><td>96.89%</td><td rowspan="4">reLu</td><td>sigmoid</td><td rowspan="4">100 epochs, 30 batch size</td><td rowspan="4">Adam</td><td rowspan="4">Sparse categorical crossentropy</td><td rowspan="4">accuracy</td><td>He normal</td><td>11s</td></tr>
<tr><td>ModelII</td><td>97.61%</td><td>softmax</td><td>Glorot normal</td><td>11s</td></tr>
<tr><td>ModelIII</td><td>97.61%</td><td>Sigmoid</td><td>Glorot normal</td><td>21s</td></tr>
<tr><td>ModelIV</td><td>97.61%</td><td>Softmax</td><td>He normal</td><td>9s</td></tr>
</table>

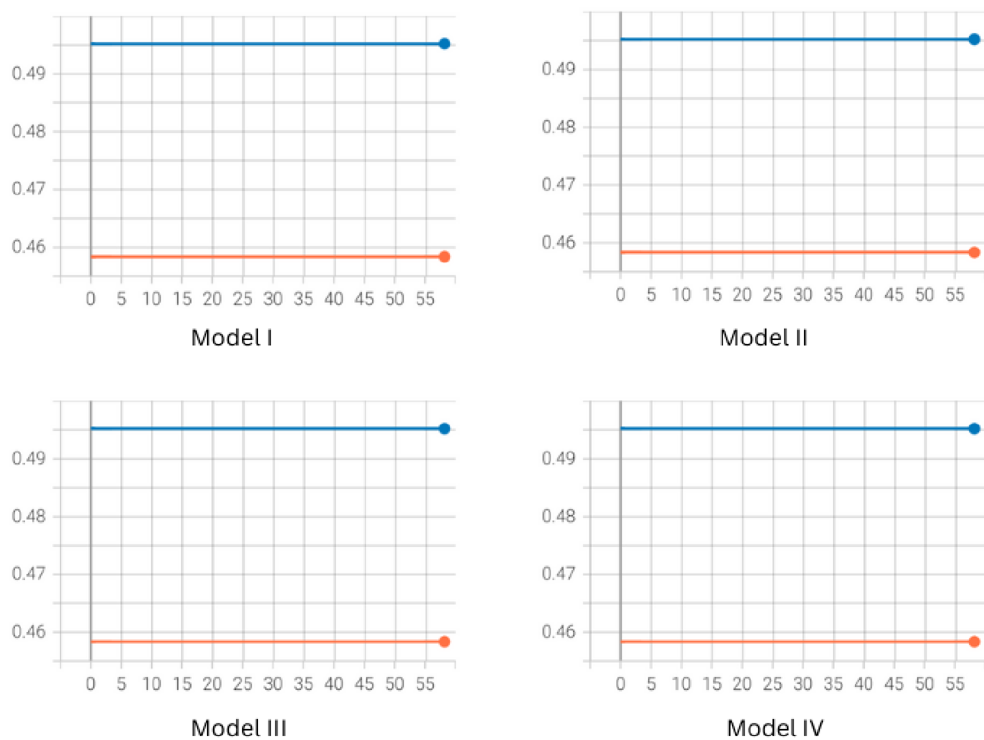
(Table 2: models' comparisons)

We have also been trying to initialize the weights using statistical distribution. We decided to choose between 'He normal' and 'Glorot normal' initialization. We might not see much of a difference between the all the models that we tried to compare in the table, but let's view additional data:

```
avg val_accurracy of each model:
Model1| ('sigmoid', 'he_normal')        | 0.9453441305802419
Model2| ('softmax', 'glorot_normal')    | 0.9531932615715525
Model3| ('sigmoid', 'glorot_normal')    | 0.959880388379097
Model4| ('softmax', 'he_normal')        | 0.932731054331127
```

The accuracy itself doesn't show how the model is performing, I added this matrix which take all val_accuracy (indicating the test data accuracy) performed by the model each 100 epoch and saved them, then got the average of its performance evaluation. We can clearly see how models 2 and 3 are performing slightly better than models 1 and 4 – so could that mean that 'Glorot_normal' distribution initializer performs somewhat better than 'He_normal' in our case? Maybe. But when it comes to the time, Model 4 was the fastest and worst performer where Model 3 was the slowest yet best performer (in the range of our experiment). 21 seconds is not that long, but again this architecture and the weights that we have set will take 1-to-2-minute max. so at the end, these models are relevant to use for our study about obesity.

Why haven't we used 1 neuron and binary cross entropy loss function as it's the "ideal concept of binary classification"? from our experiment and observation we can say it is not the best solution, as the test data accuracy of all models have decreased significantly to 50%, and train dataset accuracy is around 45%! That could be because of the number of hidden layers or because of the weights, but if we compare that to the solution that we have by the 2 neurons models, it performs so much better on addition to that it being only 1 hidden layer of 9 neurons! Proofs of the 1 neuron output layer model performance illustrated in figure 5.



(Figure 5: Bad models' accuracy)

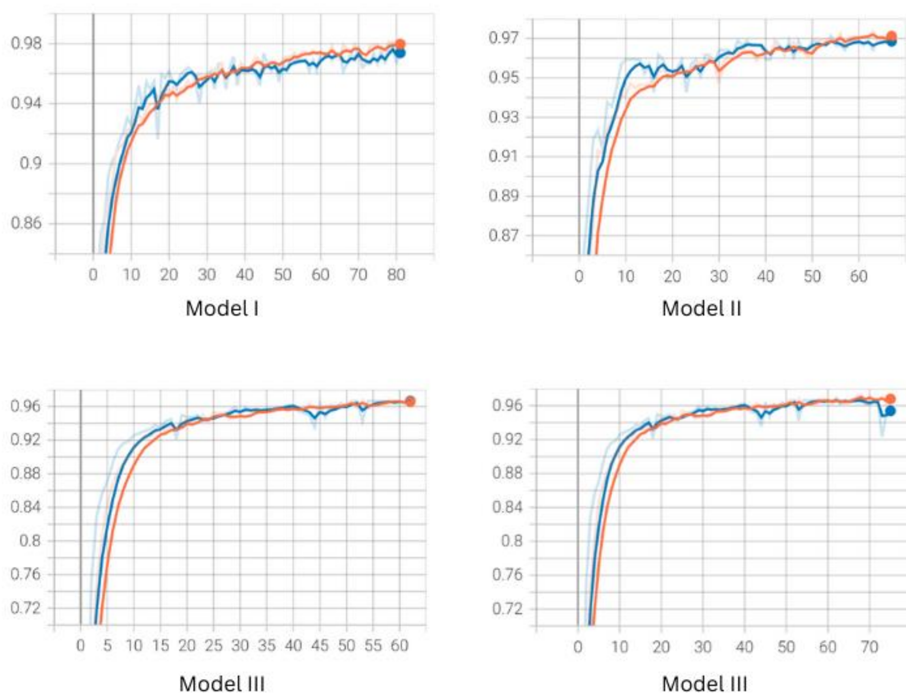
VI. Results

Our model has performed at a great rate, it is not 99% nor 100% but it is good enough to predict if someone has obesity or not. In addition to that we have used early stopping that made the performance even better. Figure 6 shows the accuracy of test dataset (in blue) and the accuracy of the train dataset (in red) and we can see how all the models have a great shape, there is no huge gap between the train dataset and the test dataset, and there is not a big variance in the values. Another good thing as well is that the accuracy of the test dataset is less that the training dataset accuracy, in conclusion we can say that the model is not memorizing, but rather learning from what has been trained with, which means that the model is able to generalize.

Another thing that we must take into consideration just for a deeper look into the model's fitting performance which is the training is how fast was it able to get that high point, in other words the "perfect" weights. And we can see from Table 3 that Model 1 did stop earlier but the accuracy isn't the best out there and it took 11 seconds. Whereas Model 3 didn't use the early stopping, took the most time (which makes sense) and the model accuracy is just the same as the other two (Model 2 and 4).

ModelID	Accuracy	Time	Early stopping
ModelI	96.89%	11s	52/100
ModelII	97.61%	11s	92/100
ModelIII	97.61%	21s	100/100
ModelIV	97.61%	9s	76/100

(Table 3: Early stopping)



(Figure 6: Models' accuracy)

Other metrices that we have measured to get the performance of the models is illustrated by the code output in figure 7. After this point we are sure that Model 2, 3, 4 have the same accuracy, precision, recall, f1-score, and auc scores at the end of the model's predicting phase after training it. But they are different in terms of speed (could be because of the resource/machine used) and the different accuracy in each epoch, indicating that the model is learning.

```
              precision    recall  f1-score   support

           0       0.95      0.99      0.97       203
           1       0.99      0.95      0.97       215

    accuracy                           0.97       418
   macro avg       0.97      0.97      0.97       418
weighted avg       0.97      0.97      0.97       418

ROC AUC: 0.969355
```

Model I

```
              precision    recall  f1-score   support

           0       0.99      0.96      0.97       203
           1       0.96      0.99      0.98       215

    accuracy                           0.98       418
   macro avg       0.98      0.98      0.98       418
weighted avg       0.98      0.98      0.98       418

ROC AUC: 0.975644
```

Model II

```
              precision    recall  f1-score   support

           0       0.99      0.96      0.97       203
           1       0.96      0.99      0.98       215

    accuracy                           0.98       418
   macro avg       0.98      0.98      0.98       418
weighted avg       0.98      0.98      0.98       418

ROC AUC: 0.975644
```

Model III

```
              precision    recall  f1-score   support

           0       0.99      0.96      0.97       203
           1       0.96      0.99      0.98       215

    accuracy                           0.98       418
   macro avg       0.98      0.98      0.98       418
weighted avg       0.98      0.98      0.98       418

ROC AUC: 0.975644
```

Model IV

(Figure 7: Evaluation metrices)

VII. Interpretation And Discussion

In the beginning of the report, we have been talking about some problems that we want answers for. And by the help of visual analytics, we can get those answers. We found out the following *(visualizations are found in the colab file):*

1. Obesity can occur due to genetic factors and mutations in some obesity-causing genes. Obesity occurs in this case due to an increase in appetite and the amount of food eaten, while the feeling of hunger does not stop or that there is a defect in the genes responsible for controlling the percentage of fat in a specific place in the body. It appears from the graph that there is a close relationship between obesity and hereditary genetics. Families that suffer from obesity genes inherit this disease to their children, so they are more likely to develop obesity than those who do not have genetic obesity genes.
2. People who eat meals with high calories and large quantities have a susceptibility to obesity, and people who do not monitor the calories that enter their body have a greater chance of obesity.

And the ideal person who monitors his calories and controls the amount of food he eats gets a healthy body.

3. People who have a low Physical Activity Frequency have a higher susceptibility to obesity because their metabolism is low, and vice versa.
4. There is a direct relationship between obesity and the transportation that a person uses. Obesity appears in people who use public transportation and automobiles to a large extent, and to a lesser extent for people who walk to work or ride motorcycles and bicycles.
5. Alcohol can cause weight gain in four ways: it stops your body from burning fat, it's high in kilojoules, it can make you feel hungry, and it can lead to poor food choices. For these reasons, drinking alcohol gives a greater chance of gaining weight than people who do not drink alcohol.
6. When you smoke or even inhale cigarette smoke, nicotine has the effect of suppressing your appetite, which is why non-smokers have a greater chance of gaining weight than smokers.
7. when eating large quantities of vegetables, whether they are from starches that increase weight or from non-starches that do not contain large amounts of starch, this leads to the accumulation of calories inside the body and leads to weight gain. Overweight people should eat vegetables in limited quantities.
8. The number of meals may affect the meal quantities, and therefore those who eat many meals with eating between meals in large quantities will become obese. While those who eat many meals with eating in between meals in small quantities lose weight.

How exactly is this related to our motivation and the experiment that we did and explained? We in this report are trying to minimize the cases of obesity around the globe – which isn’t easy since the cases of people and their different conditions have a hand on this manner – nevertheless, Table 4 summarizes the solutions that we came up with from the dataset to answer the question: “How to prevent obesity”.

Factor		**Ideal measure**
BMI	Female	Less than 29.9 kg/m2
	Male	Less than or equal to 30.0 kg/m2
Physical activity		Use bike or try walking as much as possible instead of using public transportation or automobiles.
Food		Consume the amount of caloric food that suits your body by monitoring the calories in your meals.

(Table 4: Prevent Obesity)

Is there any room for improvements? Of course! Our imagination is never limited, and regarding our experiment we will see how we also created a deep learning model that is of the same architecture as our pervious one, changing some parameters, to suite a 7-class target to detect the type of obesity. Now that is not our target in this report as it is beyond our goal. We here are focusing on whether someone has obesity or not and trying to share awareness about how to prevent that from happening. Nevertheless, we will be doing that just for the sake of comparing the results to related works of people who have the same dataset.

Are there any features that we would like to have to help with better results? our data has a perfect balance between the cases in females and males, therefore we weren't able to see how gender would have an effect in obesity, as well as the age feature. And more importantly how exactly does obesity affect the mental aspect of the human! This dataset is superficial to their physical appearance but nothing deep regarding their mental health, are they married? Do they have kids? Do they work? And for how many hours? What is their job? All these questions should be investigated more than just how they weight, or how they transport.

VIII. Related Work

In this section we will be comparing our results, as well the hyperparameters and the type of algorithms used. The person we will be comparing our results with is Vasin Jinopong from 2 years ago using Machine Learning algorithms (Decision Tress and Radom Forest) to solve this problem and their target class is 'Type of Obesity' [7]. So, we will be comparing Model III and TypeofobesityModel to their results. and see in what aspects did we make an improvement. Table 5 shows the difference in results. Table 6 will illustrate the method and hyperparameters used in the Machine Learning algorithms by Jinopong. Table 7 will present the method and hyperparameters used in the Deep Leaning model by us.

What is the difference between deep learning and machine learning? Machine learning is all about enabling computers to perform tasks without being explicitly programmed to do so, but they still think and act like machines. Their ability to perform complex tasks, such as gathering data from images and videos, is still far below that of humans. Whereas Deep learning model introduces a very sophisticated approach to machine learning and is specifically designed to address these challenges as it is modeled on the human brain. complex and layered.

Model	ModelIII	typeofobesityModel	Decision Tree	Random Forest
Type of model	Neural Network	Neural Network	ML	ML
Accuracy score	0.98	0.94	0.94	0.93
Precision score	For 0: 0.99 For 1: 0.96	For 0: 0.95 For 1: 0.93 For 2: 1.00 For 3: 0.91 For 4: 1.00 For 5: 0.87 For 6: 0.86	0.946	0.933
Recall	For 0: 0.96 For 1: 0.99	For 0: 1.00 For 1: 0.86 For 2: 0.93 For 3:1.00 For 4: 1.00 For 5: 0.82 For 6:0.96	0.946	0.930
F1 score	For 0: 0.97 For 1: 0.98	For 0: 0.97 For 1: 0.89 For 2: 0.96 For 3:0.95 For 4: 1.00 For 5: 0.84 For 6: 0.91	0.945	0.931
ROC-AUC	0.97	-	0.97	0.99
Time	21s	42s	-	-

(Table 5: Comparing results)

Machine learning models	Decision Tree	Random Forest
Test size	30%	30%
grid	DecisionTreeClassifier	RandomForestClassifier
Criterion	Entropy	Entropy
Max_depth	9	10
N_estimators		29
splitter	best	

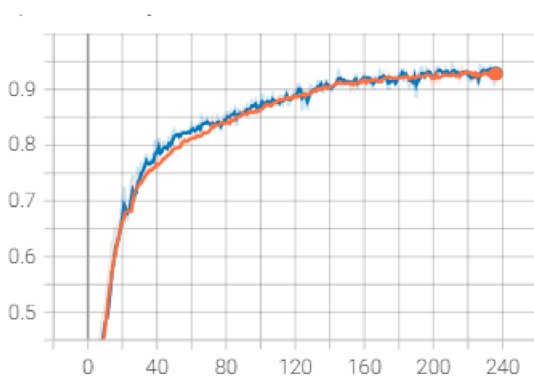
(Table 6: Machin Leaning method used)

Deep Leaning Models	ModelIII	typeofobesityModel
Input layer	16	16
Hidden layer	9	9
Activation function (Hidden Layer)	relu	relu
Output Layer	2	7

Activation function (Output Layer)	sigmoid	softmax
epochs	100	1000
Batch size	30	30
optimizer	adam	adam
loss	Sparse categorical crossentropy	Sparse categorical crossentropy
metrics	accuracy	accuracy
Kernel initializer	Glorot normal	Glorot normal
Prevent overfitting methods	Early stopping	Early stopping

(Table 7: Deep Leaning method used)

Let's now dive a little deeper to what we did, we have seen this table before through this report, but let's talk about these parameters affected the results, now early stopping did prevent overfitting, and from what Jinopong did we aren't sure whether overfitting was dealt with or not, could the results of the Machine Learning algorithm include overfitting? Now in the matter of Model III, it is only using two neurons, we are not detecting the same things as Jinopong therefore the comparison doesn't really stand well. Nevertheless, we have created another model of the same architecture, just increasing the number of neurons in the output layer to 7 from 2 to include all 'Underweight', 'Normal', 'Overweight Type 1', 'Overweight Type 2', 'Obesity Level 1', 'Obesity Level 2', and 'Obesity Level 3'. And increased the number of epochs to 1000 – yet the model stopped earlier because of the early stopping mechanism.



as seen in the picture on the left, the model stopped around 240/1000 epochs, and it took it 42 seconds. The model did perform weaker than Model III, but it is not weaker than using the Machine Learning algorithms! It performed slightly better than Random Forest and at the same rate as the Decision tree.

we found out from our experiments that Glorot_normal distribution suits our model and data, as well as the Adam optimizer and the sparse categorical cross entropy since we are dealing with categorical classes. Using the accuracy metrics so that we can compare our model with Jinopong's machine learning algorithms.

Now let's discuss the Decision Trees. We use it because it is a form of supervised machine learning used to make classifications or predictions based on the answers to a previous set of questions. Can be used for classification and regression. The tree consists of nodes, which are selected by searching for the best split of the features. Splitting the process splits one node into many nodes that must be properly chosen for maximum accuracy. Parameter "criteria", This parameter is a feature used to measure the quality of the split. Entropy is a measure of information that describes the disorder of the target feature. max_depth is the number of levels. Splitting is the process of splitting one node into multiple nodes. The higher the value of max_depth, the more complex the tree will be. The accuracy will be very low when the test data is displayed even if you increase the max_depth value.

Random Forest is a supervised machine learning algorithm composed of decision trees. Random forests are used for both classification and regression. for example, classifying if a person has obesity or not. A random forest has multiple trees, and you can set the number of trees you want in your random forest. criterion is a function that measures departmental quality. You can use max_depth to determine how deep the tree can grow. This can be viewed as one of the stopping criteria that limits tree growth. This is done with a dataset the hyperparameter n_estimators were 29. For better accuracy, the number should be larger.

The size of the input dataset that is provided to the machines has a significant impact on the accuracy of the models. Small datasets are better suited for ML models. Similar to this, deep learning models perform better when the dataset is large, so as part of the improvements for the future, we can increase the number of records in this dataset to suit Deep learning models, as they might in reality be better.

IX. Conclusions

In conclusion, we saw how human's physical condition and activities do severely affect them having obesity. And it is not something simple as the mass body index increases significantly above its normal rate that is corresponding to the person's gender, age, height, and weight. Nevertheless, other things do affect those things in return yielding those results, so preventing things like overeating, not being physically active, drinking alcohol, etc. will impact the weight gain procedure. But other things like inheriting obesity is something that should be treated medically. We made sure to talk about how it is very important to share awareness about this topic since it is the beginning to very serious problems like diabetes and cancer. The process of detecting obesity using deep learning mechanisms has yielded great results when narrowing the target class. We compared our work to other peoples in the same field, making sure that we have at least made improvements and suggesting how to make better analysis in the near future!

X. References

[1] *"UCI Machine Learning Repository: Estimation of obesity levels based on eating habits and physical condition Data Set," archive.ics.uci.edu.* https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition+#

[2] F. M. Palechor and A. de la H. Manotas, "Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico," Data in Brief, vol. 25, p. 104344, Aug. 2019, doi: 10.1016/j.dib.2019.104344.

[3] Mayo Clinic, "Obesity - symptoms and causes," Mayo Clinic, 2021. https://www.mayoclinic.org/diseases-conditions/obesity/symptoms-causes/syc-20375742

[4] "Ranking (% obesity by country), 2016," World Obesity Federation Global Obesity Observatory. https://data.worldobesity.org/rankings/?age=a&sex=f

[5] "Overweight vs Obese - Difference and Comparison | Diffen," Diffen.com, 2019. https://www.diffen.com/difference/Obesity_vs_Overweight

[6] H. Ritchie and M. Roser, "Obesity," Our World in Data, Aug. 2017, [Online]. Available: https://ourworldindata.org/obesity#:~:text=13%25%20of%20adults%20in%20the

[7] kaitod14 (2021) *[starter] estimate obseity level classification, Kaggle*. Kaggle. Available at: https://www.kaggle.com/code/kaitod14/starter-estimate-obseity-level-classification