

Prediction of Weight Gain during COVID-19 for Avoiding Complication in Health

¹Prof. Shawni Dutta, ²Prof. Sabyasachi Pramanik and ¹Prof. Samir Kumar Bandyopadhyay

¹Department of Computer Science, The Bhawanipur Education Society College, Kolkata,
India

²Haldia Institute of Technology, West Bengal, India

Address of the Corresponding Author: Prof. Samir Kumar Bandyopadhyay

Email: 1954samir@gmail.com

Abstract

Obesity and overweight is a foremost concern around the globe for each group of age. This can be accelerated by the current imposed lockdown. However, excessive weight gain may result in other chronic diseases. This study has been considering the age group of 25 to 55 years as the sample populations and monitoring them from July, 2020 to November, 2020. The lifestyle of this population, food habit, mental health conditions are explored using deep learning based framework. All these parameters need to be monitored as these have close relation with currently imposed constraints due to COVID-19. A predictive model is constructed using deep learning techniques to predict the risk of gaining weight. The predictive model hybridizes the convolutional layer and gated recurrent neural networks as a unified entity for achieving the objective of early weight gain prediction. The result obtained by this model exhibits an encouraging predictive efficiency of 93.7%.

Keywords: Overweight, obesity, deep learning, Convolutional layer, GRU, COVID-19, lifestyle.

1. Introduction

Obesity is now a major global concern for human being. It affects over 2.1 billion people overweight or obese worldwide. It is almost 30% of the global population. The overweight and obese population is likely to increase to 41% by 2030 if the present situation continues further [1]. Overweight and obesity often contributes to other chronic diseases in human body. In turn it increases the chance of direct or indirect economic impacts, hospital care and pharmaceutical necessity and many other factors. This problem has even interleaving relationship with stress and mental anxiety as well. Both psychological and physical stress

can be often experienced by individual at their daily life. Several studies established that an association exists between stress and the amount of food consumed. People under stress crave more high fat and high sugar foods, since the body under stress requires more energy to function. Therefore, the increment in unstructured time and the psychological impact resulting from the enforced quarantine might induce changes in dietary habits and lifestyle. The concept of stable lifestyle has massively disturbed by the current COVID-19 situation and resultant lockdown event. The increased work-pressure, fear of job loss, and other relevant work load are enforcing people into acute depressed situation. In fact, the psychological distress condition has a clear impact on dietary procedure which in turn is responsible for weight gain event [2]. However, another study [3] has shown that the association between job stress and weight gain is even more severe in case of women.

The main purpose of any weight-loss intervention is to ensure long-term weight loss. The success of this depends on the initial weight loss. There is a necessity to track the progress of the patient in the dietary intervention program. Moreover, there is a necessity to understand and consider the end-goal of patient's adaptability and ability to reach any degrees of weight change. Different approaches may be employed for weight management event. Modifications in lifestyle may ensure weight-loss which may be applied as a treatment for overweight and obese patients. Individuals developing signs of weight gain or obesity are at a risk of developing serious illnesses such as type 2 diabetes, respiratory problems, coronary heart disease and stroke [3]. Physical activity and healthy eating can be a fundamental component to maintain a healthy lifestyle. In fact, another fact is being observed that the weight gain may be even caused by combination therapy of insulin intake and type-2 diabetes [4]. Smoking and tobacco consumption has closely-couple association with obesity problem and type-2 diabetes. A clear indication has been shown in [5] the person who quit smoking has gained weight over a period of time. So the impact of all these factors may have severe effect on obesity problem. Hence, this study has monitored and acquired the data during the lockdown period for analysis.

This study investigates the association among the lifestyles, habits and weight-gain possibilities. COVID-19 and the imposed lockdown have imposed severe mental anxiety, depression on the people. However, the other relevant habits such as smoking, alcohol consumption, diabetes, consumption of snacks may accelerate the obesity problem. The major contribution of this study is to construct an automated tool that will analyse these

interfering factors and predict the weight gain risks in advance. The utility of this tool is to provide more concentration on affected people so that obesity problem may be avoided.

To facilitate preventive healthcare measures, Machine Learning (ML) in medicinal services can be employed. The reason of employing preventive healthcare services is to forestall the disease occurrence by taking some measures instead of disease treatment. Various degrees of preventive healthcare strategies techniques are available. ML based system can serve as secondary prevention of health care. The system identifies and analyses health conditions prior to the development health status of patients. Promising intervention is necessary to pursue so as to reduce the weight gain risks and related consequences which is a significant concern around the world.

Deep learning (DL) is often considered to be a subfield of machine learning techniques. DL technique is popular because of its' self-adaptive structure that processes data with minimal processing [7]. DL based study is investigated in this research to determine whether a person can gain weights or not. An analytical study has been carried out based on the food habits and lifestyle for weight gain classification task. Use of deep learning is beneficial as it ensures the automatic and efficient feature extraction from the underlying data. Feature extraction process is one of the most important tasks as it enables the model to understand which characteristics in the data can uniquely indicate the prediction label. DL techniques exhibit feature extraction process by a means of multiple neural network based model.

A range of machine learning algorithms have been applied in health domains to predict the presence certain health conditions using a number of characteristic features. Artificial Neural Networks (ANNs) are a commonly used branch of Machine Learning (ML) methods that are used to correlate input parameters to corresponding output data. Numerous examples of application of ANNs in medical applications and engineering systems have been reported with a varying degree of success. A survey of over 300 recent contributions in applying Deep Learning (DL) and Convolutional Neural Networks (CNNs) to predict health risks/disease a set of medical images highlights the scale of ML applications in medical field [8]. This is why construction of predictive model is favoured in this study that employs deep neural networks for weight-gain monitoring event.

The primary workflow of this weight gain classification task can be summarized as follows:

- The age group of 25 to 55 years has been considered as the target population. People residing in rural and urban area within West Bengal, India are provided with some questionnaire where their daily life styles, existing health conditions, food habits, mental condition have been assessed. The data collection period spans from July, 2020 to November, 2020.
- Collected data are inherently noisy, and contains missing values. Hence requires pre-processing step for obtaining cleaned dataset.
- A DL based methodology is applied on the dataset for predicting whether a person can suffer from obesity related problems or not. A hybrid model is constructed for this purpose that automatically captures the associated relationship among the interfering factors leading to weight gain problem. Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) based layers are the major component of the proposed hybrid model.

2. Related Works

The study related to weight gain classification task is quite limited in number. However, existing researches have kept their benchmark outputs in this domain. An artificial neural network based system is developed to assess the impact of medical and sociological data on childhood obesity. Use of radial basis function helped this study to handle this complex problem [9]. Childhood obesity is also predicted by an approach that assembles two heterogeneous approaches such as Naïve Bayes and genetic algorithm with an efficiency of 75% [10]. Another study also focused on childhood obesity prediction by using existing machine learning techniques such as Random Tree, Random Forest, J48, ID3, Naïve Bayes, and Bayes models. Highest accuracy of 85% and sensitivity of 90% is obtained for childhood obesity prediction [11].

Various multivariate regression algorithms and multi-layer perceptron feed forward artificial neural networks have been implemented in order to predict weight gain risks with promising accuracy of 93.4% [12]. Another research work developed a pilot learning with the help of decision tree model for predicting obesity. A study has established the relation between self-quarantine and weight gain based on eating behavior. The study has reported that around 22% of the adults in the considered population have faced the weight gain problems [13].

However, based on our knowledge, very limited studies have assessed the impact of COVID-19 on weight gain risk analysis using machine learning oriented approach. Our study is

different from the other studies as it considers the lifestyle modalities, stress, anxiety, eating behavior which can lead to obesity problem. Apart from this, instead of focusing any of the defined age group (such as young, childhood or adult), this article has enquires the age group from 25 to 55 years. The objective of carrying out this study is to assess the significant impact of lockdown on lifestyle modification which may be directly or indirectly responsible for weight management problem.

3. Data Collections

This study carries out an extensive data collection from July, 2020 to November, 2020. Total of 1270 people were surveyed and provided with a questionnaire for collecting data. Collection of these data is based on existing diseases and life styles of the people. Emphasis on stress, mental anxiety raised due to covid-19 pandemic situation is taken into consideration. The questions asked to the people are summarized in Table 1. For this data collection purpose, six groups each of having 5 people were appointed. Due to imposed lockdown and disruption have given us this opportunity to conduct this survey and enriched our objective. Since, we have considered very short period of time for obesity analysis. This is why, the weight gain of more than or equal to 2 Kgs. has considered to be the targeted population to whom concentration should be provided. The same set of people is monitored for a couple of months ranging from July, 2020 to November, 2020. The weight gain ratio is plotted in Figure1 which indicates that obesity problem increases as the time elapses.

Information	Description of the attributes	Range of Response values
Age	Age of the Respondent (in years)	25-55
Name	Name of the Respondent	Several string values
Gender	Gender of the Respondent	Male, Female
Income	Monthly Income of the Respondent	25000-70000
Baseline Weight	In general weight label of the Respondent	Normal, Overweight

Panic disorder	Whether the Respondent suffers from panic disorder or not	Yes, No
Diabetes	Whether the Respondent is a diabetic patient or not	Yes, No
Depression	How the respondent scale himself in case of having depression	0-10
Happiness	How the respondent scale himself in case of having depression	0-10
Smoking_quit	Whether the patient has quitted his smoking or not	Yes, No
Job status	Current job position during lockdown	Stable, Moderately Stable, Unstable
Consumption of Snacks	Amount of snacks consumption during the specified time	Moderate, Higher, No intake.
Adequate Sleep	Rest taken by the person	<6 hours, >6 hours.
Weight Gain Tendency	Whether the patient has gained weight or not	Yes, No

Table 1: Attribute information for weight gain

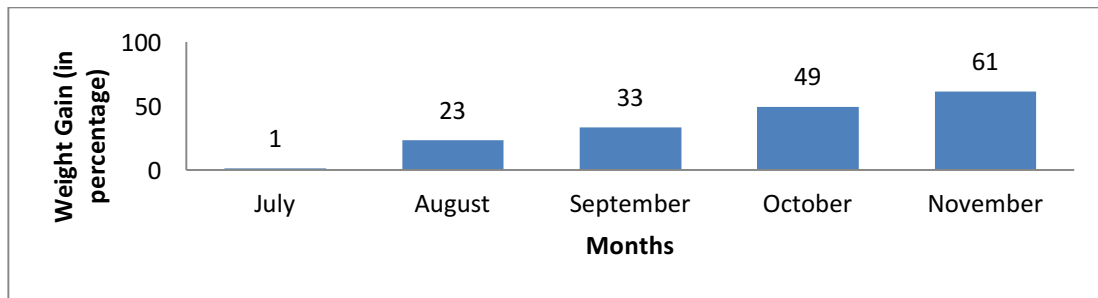


Figure1: Month-wise weight gain percentage analysis

The collected dataset contains some categorical attributes. These attributes need to be converted to numerical attributes which is achieved by assigning numerical values to each of the string values. This step ensures that any quantitative classifier such as neural network based model can be applied successfully. However, this pre-processing step also contains other operations such as missing values elimination and replaced by mean value of the corresponding attribute; irrelevant attribute elimination (such as Name of the respondent). The pre-processing step is followed by separation of training and testing dataset. The training and testing dataset is separated with the ratio of 7:3. The training and testing dataset can be characterized by the presence of target variable 'Weight gain tendency' which is a binary attribute, thus implies a binary classification task.

4. Background

4.1 Deep Learning and Neural Networks

Deep learning (DL) is essentially a specialized branch of machine learning field that enables the use of human intelligence thereby simulating neural network artificially by incorporating more than two layers. DL techniques are expert in extracting the impactful characteristics or features in the underlying dataset automatically. These techniques are useful in practice as it can minimize the human effort in feature extraction and enhance accuracy of the task. DL techniques essentially exploit the working of neural network based structure. In this structure, many units (or neurons) work in parallel with no centralized control. The weights present between the connections of the units are required for preserving long-term information [7].

4.2 Activation functions and other hyper-parameters

As neural networks are having multiple layers within its structure, the layers are categorized into input layers, hidden layers, and output layers. In order to propagate the output from one

layer to the next layer, activation functions are used. These functions also augment non-linearity property to the network's model capabilities [14]. Sigmoid, ReLu, Tanh are the most popular activation functions those are used in practice.

- Sigmoid activation functions can convert the range of infinite values into simple probabilities between 0 and 1 [15] and can be defined as equation (1)

$$f(x) = 1/(1 + \exp^{-x}) \quad (1)$$

- As compared to sigmoid function, tanh (or alternatively known as hyperbolic trigonometric function) handles negative values as input and normalizes within the predefined range from -1 to +1 [16]. This function can be defined as equation (2).

$$f(x) = (\exp^x - \exp^{-x}) / (\exp^x + \exp^{-x}) \quad (2)$$

- The ReLU activation function does not suffer from vanishing gradient issues if it is compared to the sigmoid and tanh activation functions. DL applying ReLU activation functions is used by researchers to train well without using pre-training techniques [17]. This function can be defined as equation (3).

$$f(x) = \max(0, x) = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ 0, & \text{if } x_i < 0 \end{cases} \quad (3)$$

The neural network based models require training which can be carried out efficiently by a means of appropriate hyper-parameter selection. Correct hyper-parameter selection process assists model in avoiding under-fitting and/or over-fitting the underlying data during training. It is necessary to integrate multiple neural layers into a single framework. Hence, use of an optimizer is necessary. One of the popular optimizers is Adam that makes the system efficient in terms of computation with lower memory requirement and also easy to implement. This optimizer is appropriate for first-order gradient-based optimization of stochastic objective functions. It is based on adaptive estimates of lower-order moments. It is well acknowledged due to its applicability on non-stationary objectives and problems with very noisy and/or sparse gradients [18].

All neural models are undergone through training process execution. Training procedure requires epoch and batch size to be considered. The definition of an epoch is one cycle through which training process is executed. Here the dataset is partitioned into smaller

subsections. A repetitive process is executed through a couple of batch size. It considers subsections of training dataset for completing epoch execution [19]. In order to solve binary classification problem, binary cross entropy function is used as a training criterion. It actually measures the distance from the true value (which is either 0 or 1) to the prediction for each of the classes. To find out final loss class-wise errors are averaged [20]. The loss function, binary cross entropy, is used for binary classification tasks. It is very convenient to train a model to solve many classification problems at the same time.

4.3. Types of neural network

Recurrent neural network (RNN) [21] processes both sequential and parallel information. The It differs from traditional neural network since here relationship observed among input and outputs. It also enables to memorise past input data. The variant of RNN is Gated Recurrent Unit (GRU). GRU eliminates gradient vanishing problem which is one of the disadvantage of RNN. GRU can control the flow of information inside the unit without considering separate memory cells. It has a lesser number of gates which are activated using current input as well as previous output. Here the information flow from the previous activation while computing the new, candidate activation, but does not independently control the amount of the candidate activation being added. It is faster due to reduction of parameters. In GRU update gate determines the amount of past information to be passed to future processing. The reset gate memorizes amount of information to be remembered [22].

CNN has improvement over traditional neural network. By discovering the local relationship between nodes it extract underlying hierarchical features. It is operated by each neighbour node in order to capture inherent relationship in adjacent nodes. The convolutional layer of CNN determines the output of neurons. They are connected to local regions of the input through the calculation of the scalar product between their weights and the region connected to the input volume [23]. The use of learnable kernels focuses around layers parameters. Here an input data and a convolution kernel are subjected to particular mathematical operation to generate a transformed feature map. Convolution is often referred as a filter. In the kernel filters the feature maps for information of a certain kind. It is a linear operation that involves the multiplication of a set of weights with the input. For obtaining results an array of input data and a two-dimensional array of weights, called a filter or a kernel, are multiplied. In Convolutional layer ReLu activation function is popularly used [24]. Non-linearity applied after convolution assists in the successful solution simulation.

4.4. Efficiency assessment of predictive model

The efficiency of any machine learning based model can be assessed by using four qualitative terms such as True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). These four terms are summarized in a matrix called as confusion matrix [25]. Accuracy, precision, recall, f1-score are different measures for analysing the model's performance [26]. However, all these measures are calculated by quantifying TP, TN, FP and FN.

Mathematically, accuracy, f1-score can be defined as follows with given True Positive, True Negative, False Positive, False Negative as TP, TN, FP, FN respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F1-Measure or F1-Score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Mean Squared Error (MSE) [27] is another evaluating metric that measures absolute differences between the prediction and actual observation of the test samples.

$$\text{MSE} = \left(\sum_{i=1}^N (X_i - X_i')^2 / N \right) \text{ where } X_i \text{ is the actual value and } X_i' \text{ is the predicted value.}$$

5. Methodology

5.1. Problem definition

The weight gain risk can be defined as a multi-variate function $f(x) = \frac{d}{dx}(W_x)$ where W_x defines the weight of a person x . It is to be noted that the $\frac{d}{dx}(W_x)$ can be denoted as the increment of a person x for a particular period of time. Now, $f(x)$ is dependent on several attributes such as x_1, x_2, \dots, x_n where each x_i 's denotes the lifestyle and other health related parameters. Now, using these attributes it is required to predict $f(x)$. A computer aided classification problem is explored in this study that will identify the strong association among these mentioned attributes and the target function $f(x)$.

5.2. Proposed Model Configuration

This work aims to apply supervised machine learning techniques to classify people having abnormal weight gain problems. Machine learning techniques can assist in unlocking or revealing the underlying hidden patterns within an enormous amount of data. Using

supervised machine learning technique, it is possible to monitor some patients having weight gain abnormalities. Later, this gathered knowledge can be utilized in identifying patients those may have these weight gain difficulties. This study investigates the use of deep learning methodologies which is more efficient and possesses self-adaptive structure than traditional machine learning techniques. A hybrid deep model is suggested as an automated tool in this research that will assimilate 1-dimensional convolutional layers and recurrent neural network based GRU layers as the major components. The hybridization of this model is required because the convolutional layers extract the important features and these features are sent to the recurrent neural layers. Recurrent neural layers analyze the relationship among the features those influence the output.

Stacking of these mentioned layers are followed by a dropout layer for avoiding the over-fitting concept. However, four more fully-connected neural network layers are also stacked into this model. All these layers (except the dropout layers) present in this model are activated using sigmoid, relu or tanh functions. In order to accumulate these layers with proper hyper-parameter identification, adam solver is chosen. Next, this model undergoes through an epoch size of 100 and with a batch size of 64. The detailed specification of these layers, number of nodes present in each layer, activation function used in each layer, dropout rate is summarized in the Table2.

Layer Type	Number of Nodes/ filter size/ Rate	Activation function Used
conv1d_1 (Conv1D)	256	ReLu
dropout_1 (Dropout)	Dropout Rate- 20%	None
conv1d_2 (Conv1D)	128	ReLu
dropout_2 (Dropout)	Dropout Rate- 20%	None

GRU_1 (gru)	64	ReLu
dropout_3 (Dropout)	Dropout Rate- 20%	None
GRU_2 (gru)	32	ReLu
dropout_4 (Dropout)	Dropout Rate- 20%	Tanh
dense_1 (Dense)	16	Tanh
dense_2 (Dense)	8	Tanh
dense_3 (Dense)	4	Tanh
dense_4 (Dense)	2	Sigmoid

Table2: Architecture specification of stacked Convolutional-GRU

6. Experimental Results

The presented deep model i.e., stacked Convolutional-GRU based model is trained by a means of 100 epochs with a batch size of 64. During the training time, the model learns the underlying features of the collected dataset for implementing the weight gain classification task. As the epoch size grows, the accuracy gets increased gradually and loss gets decreased. At the 100th epoch, the proposed model reaches an accuracy of 95.78% and loss of 0.51. The loss and accuracy attained by our proposed model is shown in Figure 2. At the completion of the training procedure, the testing dataset is evaluated by the proposed model and efficiency is measured. The efficiency of the prediction result is outlined in table 3. The impactful parameters which can speed up weight gain risks are ranked based on the strong association. Depending on the ranking order, Stress, depression, smoking, panic disorder, diabetes have the major influence on the weight gain risk prediction. Figure 3 depicts the top five features that accelerate the weight gain difficulties. Based on the analysis, it can be inferred mental health plays significant concern while fulfilling the objective of this study.

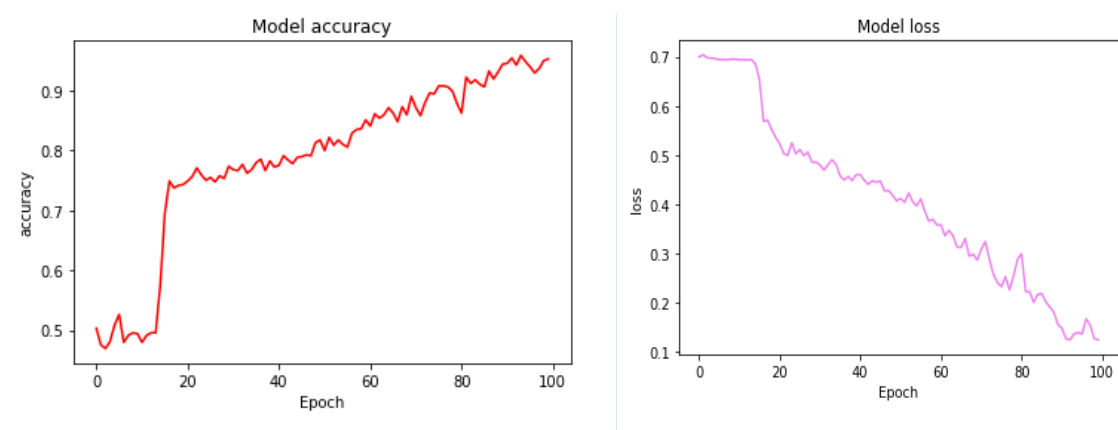


Figure2: Loss and Accuracy acquired by the model during each epoch

Proposed Model	Accuracy	F1-Score	MSE
Stacked Convolutional-GRU	93.7%	0.94	0.063

Table3: Weight gain efficiency Stacked Convolutional-GRU

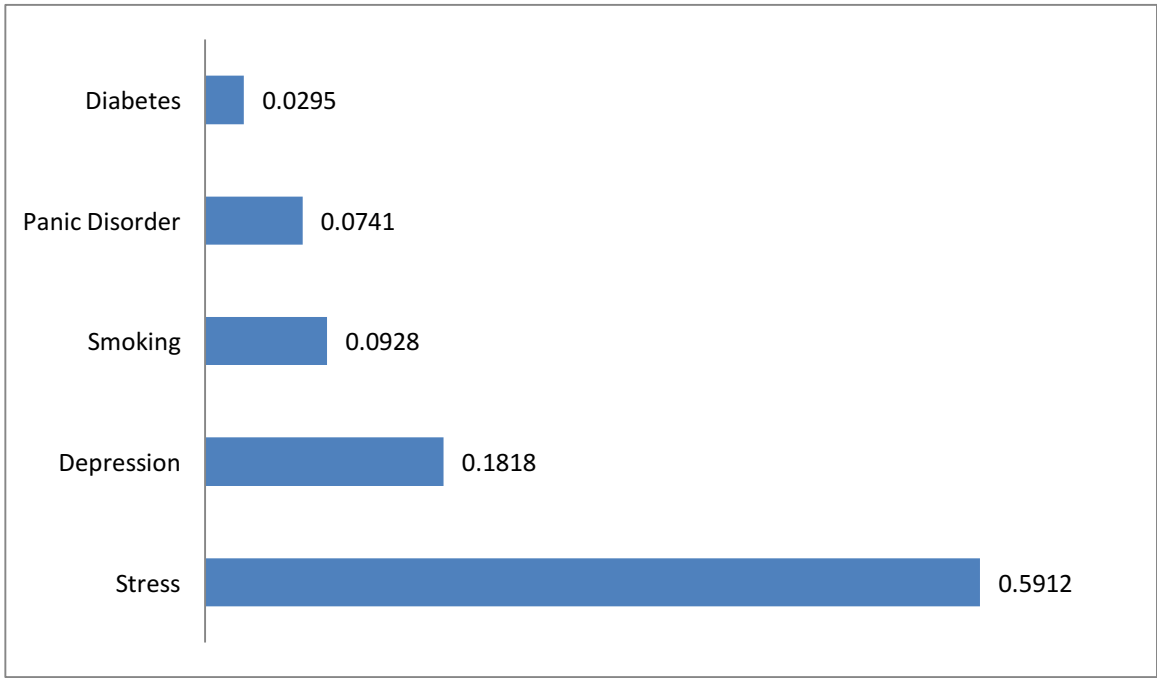


Figure3: Top five parameters that influence obesity

7. Conclusion

The conducted study has demonstrated the collected data during COVID-19 situation. The dietary intervention and other related lifestyle habits are identified from the acquired dataset. These features or characteristics are exploited to predict the short-term obesity. Real-life data are often messy in nature because of the presence of missing and erroneous values. However, application of pre-processing techniques along with the employment of powerful deep neural network based structure have provided with promising accuracy and minimized loss value. This study has investigated the association between the lockdown period and obesity problem by a means of DL framework. Two heterogeneous neural networks are assembled under a single entity for constructing a predictive model. The prediction of obesity problem has been carried out by this study with a promising accuracy of 93.7%, f1-score of 0.94 and MSE of 0.063. This predictive model automatically finds the underlying relationships among the collected features and makes an informed decision about weight gain difficulties. The analysis carried out by the proposed model identifies the strongly interfering parameters those are accountable for weight gain. Smoking habit, diabetes, stress, mental anxiety, depressions are the foremost parameters for this task. Hence, problem can be handled by providing care to the above mentioned parameters. In this current era of globalization, greater attention needs to be provided to those people suffering from obesity troubles.

References

1. Singh B, Tawfik H. Machine Learning Approach for the Early Prediction of the Risk of Overweight and Obesity in Young People. *Computational Science – ICCS 2020*. 2020;12140:523-535. Published 2020 May 23. doi:10.1007/978-3-030-50423-6_39
2. Kivimäki M, Head J, Ferrie JE, Shipley MJ, Brunner E, Vahtera J, Marmot MG. Work stress, weight gain and weight loss: evidence for bidirectional effects of job strain on body mass index in the Whitehall II study. *Int J Obes (Lond)*. 2006 Jun;30(6):982-7. doi: 10.1038/sj.ijo.0803229. PMID: 16418750.
3. Fujishiro K, Lawson CC, Hibert EL, Chavarro JE, Rich-Edwards JW. Job strain and changes in the body mass index among working women: a prospective study. *Int J Obes (Lond)*. 2015 Sep;39(9):1395-400. doi: 10.1038/ijo.2015.91. Epub 2015 May 19. PMID: 25986779; PMCID: PMC4564350.
4. Al-Hazzaa HM (2018) Lifestyle Behaviors and Obesity: Brief Observations from the Arab Teens Lifestyle Study (ATLS) Findings. *Obes Open Access* 4(1): dx.doi.org/10.16966/2380-5528.136
5. Jacob AN, Salinas K, Adams-Huet B, Raskin P. Weight gain in type 2 diabetes mellitus. *Diabetes Obes Metab*. 2007 May;9(3):386-93. doi: 10.1111/j.1463-1326.2006.00622.x. PMID: 17391167.

6. Bush T, Lovejoy JC, Deprey M, Carpenter KM. The effect of tobacco cessation on weight gain, obesity, and diabetes risk. *Obesity* (Silver Spring). 2016 Sep;24(9):1834-41. doi: 10.1002/oby.21582. PMID: 27569117; PMCID: PMC5004778.
7. Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, 3, 4.
8. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6), 1236–1246. <https://doi.org/10.1093/bib/bbx044>
9. Novak, B., & Bigec, M. (1995, November). Application of artificial neural networks for childhood obesity prediction. In *Proceedings 1995 Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems* (pp. 377-380). IEEE.
10. Adnan, M. H. B. M., & Husain, W. (2012, June). A hybrid approach using Naïve Bayes and Genetic Algorithm for childhood obesity prediction. In *2012 International Conference on Computer & Information Science (ICCIS)* (Vol. 1, pp. 281-285). IEEE.
11. Dugan TM, Mukhopadhyay S, Carroll A, Downs S. Machine Learning Techniques for Prediction of Early Childhood Obesity. *Appl Clin Inform*. 2015 Aug 12;6(3):506-20. doi: 10.4338/ACI-2015-03-RA-0036. PMID: 26448795; PMCID: PMC4586339.
12. Singh, B., & Tawfik, H. (2019, June). A machine learning approach for predicting weight gain risks in young adults. In *2019 10th International Conference on Dependable Systems, Services and Technologies (DESSERT)* (pp. 231-234). IEEE.
13. Rodríguez-Pardo, C., Segura, A., Zamorano-León, J. J., Martínez-Santos, C., Martínez, D., Collado-Yurrita, L., ... & López-Farre, A. (2019). Decision tree learning to predict overweight/obesity based on body mass index and gene polymorphisms. *Gene*, 699, 88-93.
14. Mhaskar, H. N., & Micchelli, C. A. (1994). How to choose an activation function. In *Advances in Neural Information Processing Systems* (pp. 319-326).
15. Narayan, S. (1997). The generalized sigmoid activation function: Competitive supervised learning. *Information sciences*, 99(1-2), 69-82.
16. Abdelouahab, K., Pelcat, M., & Berry, F. (2017, September). Why TanH is a Hardware Friendly Activation Function for CNNs. In *Proceedings of the 11th International Conference on Distributed Smart Cameras* (pp. 199-201).
17. Ang-bo, J. I. A. N. G., & Wei-wei, W. A. N. G. (2018). Research on optimization of ReLU activation function [J]. *Transducer and Microsystem Technologies*, 2.
18. Kingman, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. Conference paper. In *3rd International Conference for Learning Representations*.
19. Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.
20. Martinez, M., & Stiefelhagen, R. (2018, October). Taming the cross entropy loss. In *German Conference on Pattern Recognition* (pp. 628-637). Springer, Cham.

21. Medsker, L. R., & Jain, L. C. (2001). Recurrent neural networks. *Design and Applications*, 5.
22. Dey, R., & Salem, F. M. (2017, August). Gate-variants of gated recurrent unit (GRU) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)* (pp. 1597-1600). IEEE.
23. Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5, 23.
24. Ide, H., & Kurita, T. (2017, May). Improvement of learning for CNN with ReLU activation by sparse regularization. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 2684-2691). IEEE.
25. Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion Matrix-based Feature Selection. *MAICS*, 710, 120-127.
26. Juba, B., & Le, H. S. (2019, July). Precision-recall versus accuracy and the role of large data sets. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 4039-4048).
27. Singh, S., & Dayan, P. (1998). Analytical mean squared error curves for temporal difference learning. *Machine Learning*, 32(1), 5-40.