IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 10, Issue 4, Nov 2022

Estimation of obesity levels based on computational intelligence

[1]B. Govardhan [2]Nuthalapati Koteswararao, [3]Harikrishna Chilakala

Article Info

Received: 03-07-2022 Revised: 17 -08-2022 Accepted: 12-09-2022

***Abstract*—**

It's commendable that the study included participants from multiple countries providing a more diverse dataset. However, it would be interesting to explore how cultural and regional differences may influence obesity rates and causes. Focusing on students aged 18 to 25 is strategic, as this age range often represents a critical period where lifestyle habits are formed. Understanding factors leading to obesity in this demographic can contribute significantly to preventive strategies. The inclusion of factors such as caloric intake, physical activity, genetics, socioeconomic status, and mental health is comprehensive. Analyzing the interplay between these factors could yield valuable insights into the complex nature of obesity. The use of Decision Trees, Support Vector Machines, and Simple K-Means for analysis is robust. It would be interesting to delve deeper into how each algorithm performs concerning sensitivity, specificity, and predictive accuracy in identifying obesity levels. The mention of a comparative analysis is intriguing. Highlighting the strengths and weaknesses of each algorithm in the context of obesity detection could provide valuable guidance for future researchor practical applications. While 178 participants is a reasonable sample size, it's worth considering the representativeness of this sample concerning the broader population. Additionally, assessing whether the gender distribution in the study reflects regional demographics could be important. Discussing the practical implications of the findings could be beneficial. For instance, how can the results inform interventions or policies to address obesity in young adults in these regions? Given the sensitivity of health-related data, it's essential to address ethical considerations, such as participant consent, data privacy, and the potential impact of the study on participants.

***Keywords*— obesity,physicalactivity, eating habits, machinelearning, neuralnetwork, Bayesianoptimization**

1. INTRODUCTION

.

The acknowledgment of various factors contributing toobesity, including familial conflicts, depression, and anxiety, adds depth to the understanding of this complex issue. It reflects a holistic approach to the study of obesity beyond just physical factors. The reference to the World Health Organization's findings and the inclusion of studies from multiple countries broaden the scope of the research. Thisglobal perspective allows for a more comprehensive understanding of obesity as a worldwide health concern. The discussion on the challenges of accurately measuring obesity prevalence, particularly in the USA, highlights the importance of reliable data for informed decision-making. It also emphasizes the need for improved surveillance systems to address the health and economic impacts of obesity. The consideration of long-term effects and sustainability in weight loss strategies and medication regimens is crucial. It reflects a realistic approach to health interventions, recognizing that success should be measured not only in short-term outcomes but also in maintaining positive health changes over time. The incorporation of machine learning in predicting individualswho will benefit from dieting adds a contemporary dimension to the study. Machine learning's versatility, as mentioned in various applications, highlights its potential in addressing complex problems, including obesity. The clear delineation of the study structure, with sections dedicated to methodology, results, and conclusions, enhances the readability and organization of the research. It provides a roadmap for readers to follow the progression of the study. While the focus is on detecting obesity, there's potential for the study to inform future interventions and policies. Understanding the predictive capabilities of data mining techniques could pave the way for personalized and effective obesity prevention strategies.

[1] *Assistant Professor,CSE Department,Rise Krishna Sai Prakasam Group Of Institutions , Ongole,* [2] *Associate Professor CSE Department,Rise Krishna Sai Gandhi Group Of Institutions, Ongole,* [3] *Associate Professor CSE Department,Rise Krishna Sai Gandhi Group Of Institutions, Ongole.*

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 10, Issue 4, Nov 2022

2

2. Materials and Methods

2.1 Dataset

The fact that the study received approval from the Inonu University Health Sciences Non-Interventional Clinical Research Ethics Committee is crucial. It ensures that the research was conducted in an ethically sound manner, and participant rights and well-being were protected.The inclusion of 498 participants from three different locations adds geographical diversity to the study. It's essential for understanding how lifestyle factors vary across different regions.Using a web-based platform for survey administration is a modern and convenient approach. However, it's important to consider potential biases introduced by this method, such as excluding individuals without internet access or those who are not comfortable with online surveys. The chosen features related to eating habits, physical condition, and other variables seem comprehensive and relevant to the study's objectives. It allows for a multifaceted analysis of factors contributing to obesity. The use of World Health Organization (WHO) categories for BMI is standard and allows for easy comparison with global health standards. Descriptive statistics are mentioned, but it would be interesting to know if any advanced statistical analyses were performed, such as regression analysis to identify significant predictors of obesity.Obesity levels were categorized by WHO data: underweight = less than 18.5; normal = 18.5 to 24.9; overweight = 25.0 to 29.9; obesity I = 30.0 to 34.9; obesity II = 35.0 to 39.9;obesity III = higher than 40.

One of the major causes for the development of obesity refers to high caloric intake, a decrease of energy expenditure due to the lack of physical activity, alimentary disorders, genetics, socioeconomic factors, and/or anxiety and depression.That'san interesting snippet about obesity research. It's not surprising that multiple factors contribute to its development. High caloric intake, lack of physical activity, genetics, and mental health all play significant roles. It's commendable that they collected data from multiple countries to make their studymore diverse. Did they find any notable correlations or patterns among the variables and obesity in their dataset?

2.2 Decision trees (DT)

A decision tree is a classification procedure that recursively partitions a dataset into subdivisions.It typically consists of a root node, internal nodes, and terminal nodes (leaves).Nodes represent tests on one or more attributes, and terminal nodes display the results of decisions. Decision trees are employedto support decision-making processes. They assign probabilities to each choice based on the context of the decision. Each node in a decision tree has a single parent and two or more descendants. Nodes are created after dividing the data, and terminal nodes represent the final decision outcomes. Decision trees are a supervised approach for classification. The structure is formed by nodes (representing tests) and terminal nodes (showing decision results).

Decision trees can be implemented through various algorithms.C4.5 and random forest are mentioned as examples of algorithms used for implementing decision trees. Decision trees are known for their simplicity and interpretability. They provide a clear representation of the decision-making process.

2.3 Support Vector Machines (SVM)

. SVM is characterized by a strong theoretical foundation. It has demonstrated excellent empirical results in various applications.SVM has been applied by different agents in various tasks, such as digit recognition (referencing Tong and Koller, 2001), object recognition (referencing Vapnik, 1998, and Papageorgiou et al., 1998), text classification , and human activities recognition .The systems mentioned are based on the statistical learning system developed .This system proposed a mathematical model for regression and classification problems.SVM is noted for its major advantage of having powerful tools and algorithms.

These tools and algorithms are capable of efficiently and rapidly finding solutions. It's worth mentioning that SVM is a supervised machine learning algorithm used for both classification and regression tasks. It works by finding the hyperplane that best separates data points of different classesin a high-dimensional space. The ability to handle high-dimensional data and the flexibility to use different kernel functions are among the factors that contribute to SVM's popularity in various domains. Additionally, SVM is known for its good generalization performance, even in cases where the number of features is greater than the number of samples (the so-called "curse of dimensionality").

2.4 Simple K-Means

K-Means is indeed a popular method for grouping datapoints based on similarity. The idea behind K-Means is to partition your data into 'k' clusters, where each data pointbelongs to the cluster with the nearest mean. This process helps identify inherent patterns or groupings in the data without the need for labeled examples.

When applied well, clustering algorithms like K-Means can reveal insightful patterns and relationships within a dataset. The notion of generating high-quality groups with high intra- class similarities and low inter-class similarities aligns with the goal of forming cohesive clusters.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 10, Issue 4, Nov 2022

The reference to the usefulness of clustering algorithms in data mining highlights their versatile application in exploring and solving various problems. It's like having a tool that can autonomously group similar entities within your data, aiding in tasks ranging from pattern recognition to anomaly detection.

3. EXPERIMENTATION

A dataset created by a reference (Ref. [14]) was selected for predicting or detecting obesity levels in young people. Choosing an appropriate dataset is crucial for the success of any data mining project.The selected dataset underwent a process of preparation and transformation. This likely involved cleaning the data, handling missing values, addressing atypical data points, and ensuring balanced classes. Checking the correlation between attributes is essential to understand how variables relate to each other. Weka, a popular tool for machinelearning and data mining, was used for applying supervised and unsupervised data mining techniques.The choice of tools depends on the specific algorithms and analyses needed for the project. Since the study involves predicting or detecting obesity levels, supervised learning techniques would likely include classification algorithms. These algorithms learn from labeled data, where the outcome (obesity level) is known, to make predictions on new, unseen data. Unsupervised learning techniques might be used for exploring inherent patterns or structures in the data without labeled outcomes. Clustering algorithms, such as K-Means, or dimensionality reduction techniques could be part of this phase. Weka, being a comprehensive tool, likely facilitated the application of various algorithms without the need for extensive coding. The overall goal of the study seems to be developing a predictiveor detection model for obesity levels in young people using a combination of supervised and unsupervised data mining techniques. The process involves careful dataset selection, thorough data preparation, and the application of appropriate algorithms using a tool like Weka. This kind of research is essential for understanding the potential of computational intelligence in addressing health-related issues and can contribute significantly to public health efforts.

3.1 Experimental Analysis

The original dataset with 498 samples and 16 features was divided into training, testing, and validation sets.25% of the original dataset was randomly selected to generate the testing dataset.The remaining 75% was used for training.

The training dataset was then divided into two parts:

- trainForVal: Used for training the model.
- testForVal: Used for validating and optimizing model parameters.

20% of the training set was randomly selected to generate the testForVal set, and the remaining 80% was used for trainForVal. The functions of testForVal and trainForVal were used for parameter optimization.This suggests that model hyperparameters were fine-tuned using the validation set to achieve optimal performance.The testing and training sets were utilized to assess the performance of the model using the optimized hyperparameters.This involves evaluating how well the model generalizes to new, unseen data (testing set) and how well it fits the training data.Table 1 provides information on the distribution of the number of samples in each dataset by classes.This is crucial for understanding the balance of the dataset, especially when dealing with classification problems.

The entire process, including dataset splitting, validation, parameter optimization, and model assessment, was implemented using Python software.

Dataset	Underweight	Normal Weight	Overweight Level I	Overweight Level II	ObesityType I	ObesityType II	ObesityType III
training	28	212	43	9	2	42	38
testing	6	75	4	2	1	16	20
validation	22	168	35	7	1	36	30
validation	6	44	8	2	1	6	8

Table 1. The number of observations forobesity level categories in training ,testing, and validation datasets

3.2 Neural Network(NN) Optimization

A neural network model with one hidden layer was developed using the Keras library.

Neural networks are chosen for their configurability, especially with hyperparameters, and their adaptability to handle complex, nonlinear relationships in data. The NN model consists of three layers: input, hidden, and output.The number of neurons in the input layer is equal to the number of features in the dataset (16 for the original model and 5 for models with selected features).The output layer has seven neurons, corresponding to the number of classes in the dataset. The number of neurons in the hidden layer is optimized. Hyperparameters play a crucial role in NN performance. The learning rate and the number of neurons are highlighted as examples of key hyperparameters.Grid search and random search are mentioned as optimization techniques to find the

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 10, Issue 4, Nov 2022

best combination of hyperparameters. Grid search is noted to be more superior than random search but can be demandingon computational resources.Mean accuracy and F1-score measures were used to evaluate the performance of predictive models.The F1-score, being a mean measure of sensitivity andspecificity, is highlighted as an important metric for testing model validity.The Brier score is calculated to examine the models' calibration and discrimination, with lower values indicating superior model performance. The Brier score also penalizes overfitting.

A process flow diagram (Figure 1) is mentioned for a better understanding of the proposed model system. Bayesian optimization is considered a superior choice for seeking hyperparameters compared to grid search and random search.The involvement of Gaussian processes allows Bayesian optimization to consider prior results, leveraging past calculations to identify better sets of hyperparameters.It requires fewer repetitions and operates faster than other techniques, making it more efficient.Bayesian optimization remains reliable even when dealing with non-convex issues, where globally optimal solutions are challenging to obtain. Bayesian optimization was chosen to optimize hyperparameters such as the number of neurons in the hidden layer (n_unit_dense), learning rate (LR), number of epochs (epoch), and batch sizes (batch).The model was trained using the trainForVal set.The best hyperparameter set was selected based on the accuracy of the testForVal set. Bayesian optimization allows the specification of parameter intervals (minimum and maximum values) and considers any value within those intervals.The skopt library in Python was used for implementing Bayesian optimization.The gp_minimize function from the library was employed.Parameters like acq_func (acquisition function) and n_calls (number of calls) were set to "EI" (Expected Improvement) and 300, respectively.

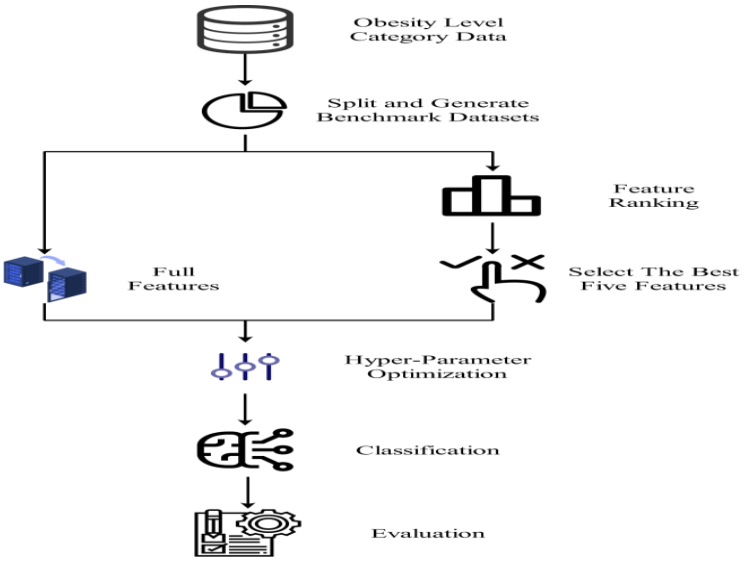

Figure1. Demonstration of an end-to-end workflow diagram of the proposed system

Hyperparameters	Lowest	Highest	Optimum
n_unit_dense	20	5000	30
LR	10^{-10}	10^{-1}	0.013
epoch	20	1500	1051
batch	1	32	16

TABLE2.THE HYPERPARAMETER OPTIMIZATION DETAILS OF THE NN MODEL USED FOR THE OBESITY LEVEL ESTIMATION.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 10, Issue 4, Nov 2022

3.3 Results

Table 3 presents the performance measure of each trial. This table shows the F1-scorefor each class and the accuracy. In addition, the mean of 10 trials and the standard deviation(SD) between the trials are also shown in Table 3.

According to the results in this table, our model obtained 93.06% mean accuracy. When class-based results are examined, it can be seen that all classes except "ObesityTypeII" were well detected.

.

Trial	Accuracy	F1-Score"over **Weight"**	F1-Score "Normal **Weight** "	F1-Score "Overweight **LevelI"**	F1-Score "Overweight **LevelII"**	F1-Score "Obesity Type **I"**	F1-Score "Obesity Type **II"**	F1-Score "Obesity Type **III"**
1	92.74 %	88.89 %	95.36 %	96.29%	100.0%	100.0 %	78.57 %	90.90 %
2	93.55 %	88.89 %	96.00 %	100.0%	100.0%	0%	80.00 %	90.00 %
3	92.74 %	88.89 %	94.74 %	100.0%	100.0%	100.0 %	74.07 %	95.24 %
4	96.77 %	94.12 %	97.99 %	100.0%	100.0%	100.0 %	90.32 %	95.24 %
5	95.97 %	94.12 %	97.33 %	100.0%	100.0%	100.0 %	87.50 %	94.74 %
6	92.74 %	88.89 %	95.36 %	96.29%	100.0%	100.0 %	78.57 %	90.90 %
7	94.35 %	94.12 %	96.69 %	100.0%	100.0%	100.0 %	80.00 %	90.00 %
8	91.13 %	88.89 %	93.51 %	100.0%	100.0%	100.0 %	66.67 %	94.74 %
9	87.90 %	84.21 %	93.33 %	88.00%	100.0%	0%	66.67 %	80.00 %
10	92.74 %	80.00 %	95.89 %	96.30%	100.0%	100.0 %	87.50 %	85.71 %
mean	**93.06 %**	**89.10 %**	**95.62 %**	**97.68%**	**100 %**	**80%**	**78.98 %**	**90.74 %**
SD	**2.34**	**4.27**	**1.43**	**3.62**	**0**	**40**	**7.76**	**4.62**

TABLE 3. PERFORMANCE EVALUATION CRITERIA FOR OBESITY PREDICTION OF THE TRAINED NEURAL NETWORK MODEL FOR EACH TRIAL

4. CONCLUSIONS

Eating habits and physical activity levels emerged as crucial risk factors for predicting obesity.These findings align with existing knowledge about the fundamental role of diet and exercise in weight management.The ML algorithms helpedin understanding how the identified risk factors contribute to changes in weight.This insight can deepen our understanding of the complex interactions between lifestyle factors and obesity.The study suggests that strategies for preventing and treating overweight and obesity should focus on increasing participation in physical activity and regulating eating habits.This insight provides actionable guidance for public health initiatives and individualized interventions.The emphasis on physical activity and healthy eating habits highlights the importance of promoting a holistic and sustainable approach to weight management.

5. REFERENCES

[1] .Kivrak, M. Deep Learning-Based Prediction of Obesity Levels According to Eating Habits and Physical Condition. J. Cogn. Syst. 2021, 6, 24–27.

[2] .Hernández Álvarez, G.M. Prevalencia de Sobrepeso y Obesidad, y Factores de Riesgo, en Niños de 7-12 Años, en una Escuela Pública de Cartagena Septiembre-Octubre de 2010. 2011. Available online: https://repositorio.unal.edu.co/handle/unal/7739 (accessed on 17 January 2023).

[3] Obesity and Overweight. Available online: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (accessed on 1 January 2023).

[4] Cecchini, M.; Vuik, S. The Heavy Burden of Obesity; OCED: Paris, France, 2019.

[5] Colditz, G.A. Economic costs of obesity and inactivity. Med. Sci. Sport. Exerc. 1999, 31, S663–S667.

[6] Gil-Rojas, Y.; Garzón, A.; Hernández, F.; Pacheco, B.; González, D.; Campos, J.; Mosos, J.D.; Barahona, J.; Polania, M.J.; Restrepo, P. Burden of disease attributable to obesity and overweight in Colombia. Value Health Reg. Issues 2019, 20, 66–72.

[7] Oshinubi, K.; Rachdi, M.; Demongeot, J. Analysis of reproduction number R0 of COVID-19 using current health expenditure as gross domestic product percentage (CHE/GDP) across countries. Healthcare 2021, 9, 1247.

ISSN 2347–3657
Volume 10, Issue 4, Nov 2022

[8] Van Baal, P.H.M.; Polder, J.J.; de Wit, G.A.; Hoogenveen, R.T.; Feenstra, T.L.; Boshuizen, H.C.; Engelfriet, P.M.; Brouwer, W.B.F. Lifetime medical costs of obesity: Prevention no cure for increasing health expenditure. PLoS Med. 2008, 5, e29.

[9] Gülü, M.; Yapici, H.; Mainer-Pardos, E.; Alves, A.R.; Nobari, H. Investigation of obesity, eating behaviors and physical activity levels living in rural and urban areas during the covid-19 pandemic era: A study of Turkish adolescent. BMC Pediatr. 2022, 22, 405.

[10] .Reidpath, D.D.; Burns, C.; Garrard, J.; Mahoney, M.; Townsend, M. An ecological study of the relationship between social and environmental determinants of obesity. Health Place 2002, 8, 141–145.

[11] Cohen, D.A.; Finch, B.K.; Bower, A.; Sastry, N. Collective efficacy and obesity: The potential influence of social factors on health. Soc. Sci. Med. 2006, 62, 769–778

[12] Lamerz, A.; Kuepper-Nybelen, J.; Wehle, C.; Bruning, N.; Trost-Brinkhues, G.; Brenner, H.; Hebebrand, J.; Herpertz-Dahlmann, B. Social class, parental education, and obesity prevalence in a study of six-year-old children in Germany. Int. J. Obes. 2005, 29, 373–380.

[13] Ng, M.; Fleming, T.; Robinson, M.; Thomson, B.; Graetz, N.; Margono, C.; Mullany, E.C.; Biryukov, S.; Abbafati, C.; Abera, S.F. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014, 384, 766–781.

[14] Ford, E.S.; Mokdad, A.H. Epidemiology of obesity in the Western Hemisphere. J. Clin. Endocrinol. Metab. 2008, 93, S1–S8.