



Machine Learning Approach for the Early Prediction of the Risk of Overweight and Obesity in Young People

Balbir Singh^(✉) and Hissam Tawfik

Leeds Beckett University, Calverley Street, Leeds LS1 3HE, UK
{b. singh, h. tawfik}@leedsbeckett.ac.uk

Abstract. Obesity is a major global concern with more than 2.1 billion people overweight or obese worldwide which amounts to almost 30% of the global population. If the current trend continues, the overweight and obese population is likely to increase to 41% by 2030. Individuals developing signs of weight gain or obesity are also at a risk of developing serious illnesses such as type 2 diabetes, respiratory problems, heart disease and stroke. Some intervention measures such as physical activity and healthy eating can be a fundamental component to maintain a healthy lifestyle. Therefore, it is absolutely essential to detect childhood obesity as early as possible. This paper utilises the vast amount of data available via UK's millennium cohort study in order to construct a machine learning driven model to predict young people at the risk of becoming overweight or obese. The childhood BMI values from the ages 3, 5, 7 and 11 are used to predict adolescents of age 14 at the risk of becoming overweight or obese. There is an inherent imbalance in the dataset of individuals with normal BMI and the ones at risk. The results obtained are encouraging and a prediction accuracy of over 90% for the target class has been achieved. Various issues relating to data preprocessing and prediction accuracy are addressed and discussed.

Keywords: Obesity · Data driven model · Young people · Classification · Machine learning · Ensemble learning

1 Introduction

Obesity in general is a growing epidemic that affects every age group. In the UK alone, the economic impact of dealing with obesity and related illnesses is almost same as the other major issues affecting the economy such as smoking and armed conflict as reported in a publication by McKinsey Global Institute on Overcoming Obesity [1]. The report discusses that more than 30% of the world population is either overweight or obese. If this growth rate prevails, the proportion of people in overweight and obese category is likely to rise up to 41% by the year 2030. The report concludes by recommending behaviour change interventions which can result in saving money in the long run as a result of reduced healthcare and increase in productivity levels. Implementation of this type of intervention could save almost £1 billion for the National

Health Service (NHS) in the United Kingdom alone. The evidence suggests that the behavioural change interventions to combat obesity need further investigation to find workable solutions rather than waiting for a perfect solution. In the United Kingdom, the data presented in Health Survey for England reports that the percentage of obese children between the age of 2 and 15 has increased significantly since 1995 [2]. 16% of boys and 15% of girls in this age group were classed as obese. 14% of both genders were classed as overweight. This results in 30% of boys and 29% of girls being either overweight or obese. In 1995, 11% of boys and 12% of girls of 2–15 years of age were obese. A range of studies have however been monitoring the trend in obesity growth. One in specific included data from review included data from 467,294 children from Australia, China, England, France, Netherlands, New Zealand, Sweden, Switzerland and USA [3]. It was reported that the prevalence of childhood obesity may be plateauing worldwide. Another survey identified 52 obesity studies worldwide from 25 countries and reported some stability in obesity [4]. Similar claims were made by [5] and [6]. Some reports even suggest that childhood obesity may well be declining as a cumulative result of increased physical activity, television viewing decline and reduction in sugary drink consumption [7]. However, [4] also reported that this stability should be observed with caution since previous stable phases were followed by further increases in prevalence of obesity.

The Health Survey England (HSE) series was established to keep an eye on changes in nation's health [8]. The survey was designed to acquire information about certain health conditions and other risk factors affecting them. It was reported that in 2013, 26% of men and 24% of women were obese. 41% of men and 33% of women were overweight but not obese. Looking at these combined figures gives cause for concern since 67% of men and 57% of women are above their normal weight for their height.

Childhood obesity is of great public concern as up to 90% of the overweight and obese childhood population will continue to be obese as adults [9]. Obesity is strongly linked with other negative health conditions such as type 2 diabetes, cardiovascular diseases, cancers and even death [10–13]. Considering all this, it is obvious that there is a pressing need to identify individuals at a risk of developing obesity as early as possible so that some preventative measures can be put in place as early as possible. The purpose of this study is to carry out apply, analyse and evaluate machine learning algorithms to classify adolescents at a risk of becoming overweight or obese using early childhood BMI as input features to machine learning. The BMI in adults is defined as a ratio of body mass in kilograms to the square of individual's height in meters. In adults, BMIs of over 25 kg/m² and 30 kg/m² are classed as overweight and obese respectively. This formula cannot be applied to children and adolescents since their body mass index changes significantly with age [14]. It varies between 13 kg/m² and 17 kg/m² from birth to an age of 1. It then decreases to 15.5 kg/m² at the age of 6 and increases to 21 kg/m² at the age of 20. Therefore, normal overweight and obesity thresholds can't be applied to children and adolescents. For this reason, this study uses overweight and obesity cut-off points recommended by International Obesity Task Force (IOTF). This paper is organised as follows:

- **Related Work:** This section discusses the application of machine learning techniques in health and approaches tackling obesity.
- **Methods:** This section deals with the data preprocessing, application of ML algorithms for the classification of imbalanced data, techniques used for the treatment of data imbalance and the evaluation of suitable algorithms for improved prediction accuracy.
- **Discussion:** Performance of ML algorithms and their effect on prediction accuracy are discussed.
- **Conclusion and further work:** The paper concludes the work carried out and the challenges experienced. Further work to build on this research is discussed.

2 Related Work

2.1 Machine Learning Applications in Health

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 [15]. A large number of examples of automated disease diagnosis using computational intelligence have been reported over the last decade or so [16–18]. ANNs have also been applied in various non-linear problem-solving scenarios including robotic decision making, swarm intelligence, aviation application and Artificial Intelligence (AI) in games [19]. Identification of diseases associated with the brain activity such as Parkinson's, Schizophrenia, and Huntington's disease from the Contingent Negative Variation (CNV) response in electroencephalograph has also been successfully implemented using ANNs [20]. Multilayer perceptron (MLP) and probabilistic neural network (PNN) have been utilized for the prediction of osteoporosis with bone densitometry [21]. A combination of several ML algorithms has also been successfully implemented to predict accurate amount of medication dosage required for patients suffering from sickle cell disease [22]. Administering accurate amount of medication based on patient's condition is of paramount importance. In this study, a range of combined machine learning techniques were investigated to accurately predict the correct amount of medication required. The algorithms investigated included Random Forests, Support Vector Machines and a few variants of Neural Networks. It was reported that a combination of Multilayer Perceptron Neural Networks trained with Levenberg-Marquardt algorithm and Random Forests provided the best results. This work builds on authors' earlier research where regression was used to predict age 14 BMI using the BMI values from earlier childhood [23].

2.2 Machine Learning Techniques for Tackling Obesity

Only a small number of studies conducted to predict childhood obesity were identified through literature searches. This is possibly due to the complex nature of the problem. Hence, it is important to utilise some complex prediction algorithm using machine learning methods or even ensemble of ensemble predictors to implement a robust and accurate prediction system rather than using simpler techniques such as linear regression or other statistical methods [24]. It is proposed that such a system should be able to:

- at least match human decision maker's behaviour
- be sufficiently general and handle a wide range of variability
- be applicable in practical situations with a proven degree of success

Reference [25] applied machine learning techniques to data collected from children before the age of 2 to predict future obesity. In this study, they used data collected on children prior to the second birthday using a clinical support system. They reported an accuracy of 85%, sensitivity of 89%, positive predictive value of 84% and negative predictive value of 88% using the ID3 algorithm for decision trees without pruning. The other algorithms tested were Naïve Bayes, Random Trees, Random Forests, C4.5 decision trees with pruning and Bayes Net. Several other studies also reported the use of machine learning algorithms used for predicting obesity. Reference [26] suggested that Radial Basis ANNs (RBANNs) are far more efficient than classical Back Propagation ANNs (BPANNs) but very large datasets would be required to train such systems. This study discussed algorithms only; results were not reported. Reference [27] discussed several algorithms for predicting childhood obesity. They recommended the suitability of ANNs, Naïve Bayes and Decision Trees. Several optimisation techniques have also been applied to achieve better prediction accuracy. For example, Genetic Algorithms were employed by [28] to improve the prediction accuracy to 92%. However, it must be noted that they used a very small sample size of 12 subjects. A comprehensive study, possibly the best one identified so far, was carried out by [29] to apply machine learning techniques to predict childhood obesity. They compared the performance metrics of several machine learning prediction algorithms. They compared logistic regression with six data mining techniques: Decision Trees, Association Rules, Neural Networks, Naïve Bayes, Bayesian Networks and Support Vector Machines. They considered prediction sensitivity the most important element in predicting obesity for their study. The highest reported sensitivity for their work was 62% in the case of Naïve Bayes and Bayesian Networks. This research group used a limited range of demographics (gender) and biometrics (weight, height and BMI) and the subjects were 2 year old children. It is envisaged that the prediction accuracy can be further improved by using a different set of parameters, using big data and other machine learning techniques such as deep learning to handle big data. Reference [30] applied machine learning techniques to measure and monitor physical activity in children. They evaluated Multilayer Perceptrons (MLPs), Support Vector Machines, Decision Trees, Naïve Bayes and $K = 3$ Nearest Neighbour algorithms. It was reported that MLPs

outperformed all the other algorithms yielding an overall accuracy of 96%, sensitivity of 95% and specificity of 99%. It should be noted that the sample size in this case was also relatively small (22 participants). The investigation of Deep Learning techniques for future work was proposed.

3 Methods

3.1 Data

The data for this study are used from UK's Millennium Cohort Study (MCS) [31]. The MCS is an ESRC funded research project that followed every child born in year 2000 and 2001. This study is the most recent of Britain's well-known study that followed national longitudinal birth cohorts. The very first MCS sweep surveyed 18,818 babies and 18,552 families related to those children. This wave was conducted in 2001 and 2002 when the babies were 9 months old. Subsequent sweeps were carried out at the age of 3 years, 5 years, 7 years, 11 years and 14 years to examine each child's growth as shown in Table 1:

Table 1. Millennium Cohort Study survey sweeps.

Millennium Cohort Study profiles						
Survey	MCS1	MCS2	MCS3	MCS4	MCS5	MCS6
Age	9 months	3 years	5 year	7 years	11 years	14 years

3.2 Classification Challenge

As discussed in the Introduction section, the BMI in adults is defined as a ratio of body mass in kilograms to the square of individual's height in meters. There are predetermined BMI values setting the boundaries for normal, overweight and obese levels. In children and adolescents, these levels are calculated differently and vary with their age and gender. The data for each survey sweep were collected over the course of approximately 12 months. This means that the children in a given sweep having exact same BMI will have different obesity levels depending upon their age and gender. This makes the classification problem significantly challenging since the target obesity levels at the age of 14 don't have linear relationship with the earlier measurements.

3.3 Data Preprocessing

A vast amount of data about children, their raising families and physical surroundings was collected including weight, height, obesity flags, socio economic conditions and indices of multiple deprivations. For the purpose of this study, body mass index values from 3 years, 5 years, 7 years and 11 years of age were used as input features to the machine learning algorithm and obesity flags for age 14 such as Normal and At risk (Overweight and Obese) were used as target variables as show in Fig. 1. The obesity flags for 14 year old subjects are determined by the following factors:

- Age at the time of survey
- Weight in kilograms and height in meters and
- Gender

These obesity flags are available in the dataset and are determined using growth charts published by the IOTF.

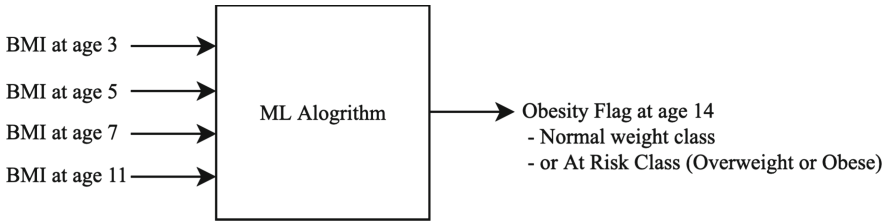


Fig. 1. Machine learning based obesity prediction model

The dataset has a large number of missing values and outliers. There are 11110 non missing values for the obesity flag for age 14 (MCS6) survey. Since the obesity flag for the age 14 survey is being used as a target variable, imputing the missing values for this variable are not going to be useful for this study. A reliable predictor variable is required to implement a robust machine learning algorithm. Therefore, to be able to use reliable data, instances of missing values for the age 14 obesity flag are deleted. As it can be seen from the count field of Table 2, a large number of data values are still missing for the input variables.

Table 2. Missing values

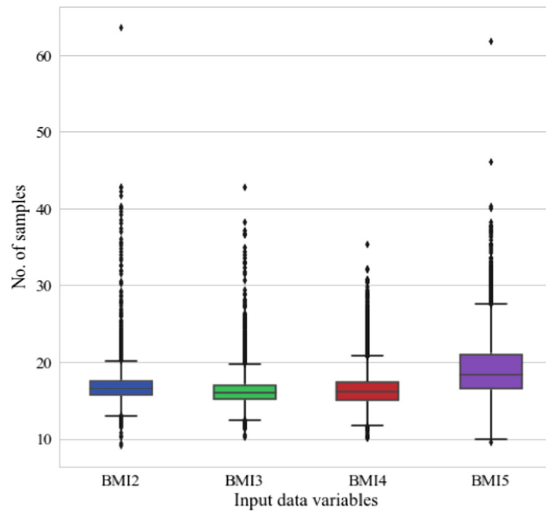
BMI	Non null values	Missing values
BMI2 (Age 3)	9525	1585
BMI3 (Age 5)	10382	728
BMI4 (Age 7)	10147	963
BMI5 (Age 11)	10398	712

The dataset analysis also indicates that the minimum and maximum data values are significantly higher than the first and third quartiles, highlighting the presence of outlier data points as shown in Table 3.

Table 3. Description of uncleaned data

	BMI2	BMI3	BMI4	BMI5
Count	9525	10382	10147	10398
Mean	16.78	16.25	16.51	19.08
Std	2.01	1.83	2.24	3.55
Min	9.1	10.19	10.11	9.48
25%	15.7	15.16	15.09	16.51
50%	16.6	16	16.08	18.3
75%	17.5	17.01	17.38	20.95
Max	63.6	42.8	35.36	61.72

As highlighted earlier, the dataset has a considerably large number of outliers which need to be identified and dealt with in order to produce meaningful results. One way of visualising and estimating outliers is to use boxplots for all of the input variables. By choosing the boxplot whisker at an appropriate level, the data points that fall beyond these points can be identified as outliers. A boxplot diagram for the input variables BMI2, BMI3, BMI4 and BMI5 is shown in Fig. 2.

**Fig. 2.** Boxplot of uncleaned dataset

All data values beyond first and third quartile are identified and imputed using the mean value of each of the input variable. The boxplot for the cleaned data is shown in Fig. 3:

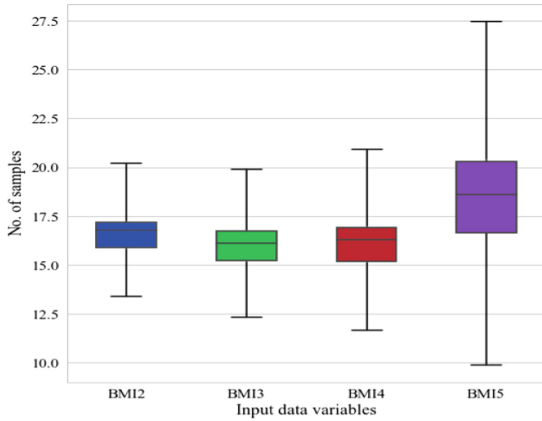


Fig. 3. Boxplot of cleaned dataset

3.4 Machine Learning Algorithm Implementation

The dataset is composed of 11110 instances and has three classes based on the obesity label. There are 8160 normal cases, 2126 overweight and 824 obese. Majority of the instances falling under the normal category, 19% under overweight and only just over 7% belonging to the obese category, makes the dataset highly imbalanced. As overweight and obese categories are both at risk, therefore instances belonging to these two classes have been combined and have been labeled as ‘At risk’. Although this reduces the imbalance a little, but majority of the algorithms are only able to classify the majority class with a high degree of accuracy.

Table 4. Classification results of imbalanced data.

Metric	Algorithm	Normal	At Risk	Weighted average
Precision	KNN	0.87	0.71	0.83
	J48 pruned tree	0.87	0.78	0.85
	Random forest	0.88	0.72	0.84
	Bagging	0.89	0.73	0.85
	SVM	0.87	0.74	0.84
	MLP	0.85	0.75	0.82
	Voting	0.87	0.74	0.84
Recall	KNN	0.91	0.63	0.83
	J48 pruned tree	0.94	0.62	0.85
	Random forest	0.91	0.66	0.84
	Bagging	0.91	0.68	0.85
	SVM	0.92	0.63	0.84
	MLP	0.93	0.54	0.83
	Voting	0.92	0.63	0.84

A range of popular classification algorithms were employed to classify adolescents of age 14 at a risk of becoming overweight or obese using BMIs measured at ages 3, 5, 7 and 11 years. The results from the best performing algorithms are tabulated in Table 4. Although the overall accuracy is well over 80% for all of the algorithms used, because of the class and BMI imbalance issues, the individual class accuracy for the At risk class is still very poor as indicated by the recall metric shown in Table 4. For example, the weighted average for recall in the case of J48 decision tree algorithm is 85% and the normal class has a significantly higher accuracy of 94%, the individual class accuracy for the ‘At risk’ class is only 62%.

Apparently, higher sensitivity or recall score for the minority class is of paramount importance in order to detect individuals at the risk of becoming overweight or obese as early as possible.

3.5 Proposed Approach

As discussed above, it is important to classify the minority class with a high degree of accuracy. To deal with the this class imbalance issue, it was proposed to synthesise new class instances for the minority class using SMOTE [32] and train ML algorithms on synthetically increased data. SMOTE generates additional data points using KNN approach by inserting synthetic data points on lines joining K nearest neighbours as shown in Fig. 4.

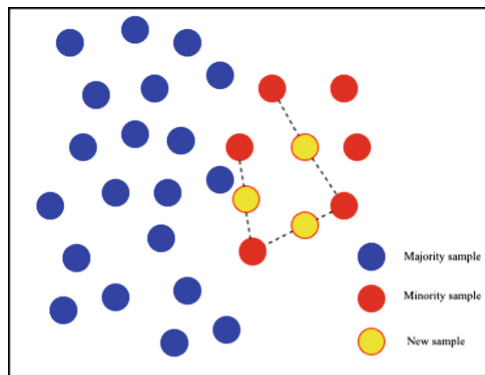


Fig. 4. Generating synthetic data points using SMOTE

A SMOTE oversampling of 138% was applied to the minority class to make the number of cases in the minority class equal to the number of cases in the majority class. However, for robustness, a trained network must be tested on an original, imbalanced dataset. For this reason, 30% sample of the original imbalanced data was kept for testing the ML algorithm accuracy. The same machine learning algorithms used in the case of imbalanced data classification were used to classify synthetically balanced dataset. The precision and recall score metrics are shown in Table 5 and plotted in Fig. 5.

Table 5. Classification results of balanced data.

Metric	Algorithm	Normal	At Risk	Weighted average
Precision	KNN	0.93	0.57	0.83
	J48 pruned tree	0.95	0.57	0.85
	Random forest	0.91	0.63	0.84
	Bagging	0.92	0.62	0.84
	SVM	0.95	0.56	0.85
	MLP	0.96	0.51	0.84
	Voting	0.96	0.56	0.85
Recall	KNN	0.78	0.83	0.79
	J48 pruned tree	0.76	0.88	0.79
	Random forest	0.84	0.76	0.82
	Bagging	0.82	0.80	0.82
	SVM	0.74	0.89	0.78
	MLP	0.68	0.92	0.74
	Voting	0.74	0.91	0.78

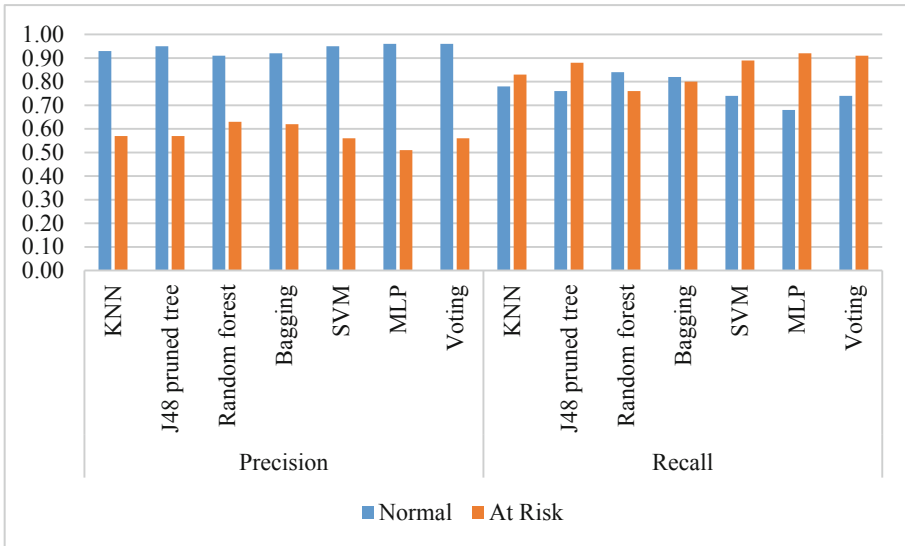


Fig. 5. Precision and recall plot of Normal and At Risk classes

The recall metric shows a significant increase for the ‘At risk’ class accuracy. Although the normal class accuracy has decreased somewhat, but the increase in the minority class accuracy outweighs this loss since detecting individuals at a risk of becoming overweight or obese is far more important than misclassifying a normal individual.

The recall metric comparison is plotted in Fig. 6 as a bar graph where the output of each algorithm is plotted before and after oversampling the minority class.

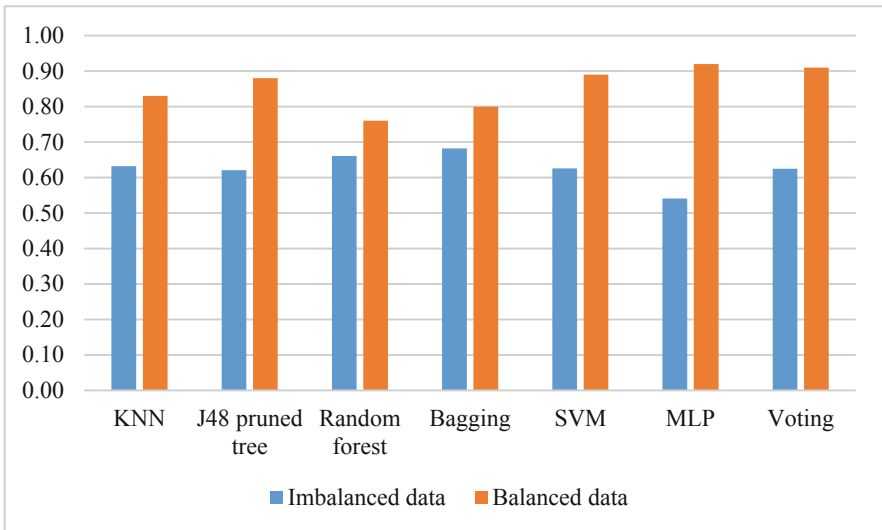


Fig. 6. Comparison of minority class accuracy for imbalanced and balanced datasets

4 Discussion

It is clear from Table 5 and Figs. 5 and 6 that synthetically enhancing the minority class significantly increases the sensitivity or recall metric for the minority class. Some of the algorithms that didn't perform satisfactorily in the case of imbalanced data, their performance is also improved significantly when the dataset is synthetically balanced. For example, the Multi-Layer Perceptron (MLP) resulted in a minority class accuracy of 54% when the dataset was imbalanced but jumped to 92% in the case of balanced data. It should be noted that the performance of all the algorithms was tested on a sample of unaltered, imbalanced dataset. This dataset had 2448 in the normal class and 885 in the At Risk class. The precision value for the normal class for the MLP algorithm is 96%. This means, out of all those predicted as 'Normal', only 4% were predicted as 'At Risk'. A recall value of 92% means that out of all those labeled as 'At Risk', only 8% were predicted as 'Normal'. This results in an F1 score of 93.96%.

5 Conclusion and Further Work

Obesity in general but more importantly childhood obesity is of great global concern since a majority of obese children grow up to obese adults. Many studies have employed statistical methods to predict the probabilities of children growing into adults as obese. In this paper we evaluated several popular machine learning algorithms to

accurately predict adolescents at risk of becoming overweight or obese at the teenage stage. Issues relating to low prediction accuracy because of data imbalance have been considered and dealt with using Synthetic Minority Oversampling Technique (SMOTE). An ensemble of classifiers capitalises on the prediction accuracies of individual classifier algorithms. For robustness, a sample of unaltered data is used to test the prediction accuracies. The results obtained are encouraging and the authors are continuing to develop further on this research. Future work is planned to include longitudinal and cross sectional data relating to participants to achieve even higher accuracies. It is also proposed to use predict obesity state at the age of 14 using the obesity flags from earlier ages since the BMI is age and gender dependent for ages from 2 to 20.

Acknowledgements. The authors are grateful to ‘The Centre for Longitudinal Studies, Institute of Education’ for the use of data used in this study and to the ‘UK Data Archive Service’ for making them available. However, they are in no way responsible for the analysis and interpretation of these data.

References

1. Dobbs, R., et al.: Overcoming obesity : an initial economic analysis, McKinsey Global Institute (2014)
2. Fat, L.N.: Children’s body mass index, overweight and obesity. *Heal. Surv. Engl.* (2014). Chapter 10
3. Olds, T., et al.: Evidence that the prevalence of childhood overweight is plateauing: data from nine countries. *Int. J. Pediatr. Obes.* **6**(5–6), 342–360 (2011)
4. Rokholm, B., Baker, J.L., Sørensen, T.I.A.: The levelling off of the obesity epidemic since the year 1999 - a review of evidence and perspectives. *Obes. Rev.* **11**, 835–846 (2010)
5. Blüher, S., et al.: Age-specific stabilization in obesity prevalence in German children: a cross-sectional study from 1999 to 2008. *Int. J. Pediatr. Obes.* **6**(sup3), e199–e206 (2011)
6. Moss, A., Klenk, J., Simon, K., Thaiss, H., Reinehr, T., Wabitsch, M.: Declining prevalence rates for overweight and obesity in German children starting school. *Eur. J. Pediatr.* **171**, 289–299 (2011). <https://doi.org/10.1007/s00431-011-1531-5>
7. Wabitsch, M., Moss, A., Kromeyer-Hauschild, K.: Unexpected plateauing of childhood obesity rates in developed countries. *BMC Med.* **12**, 17 (2014)
8. Moody, A.: Adult anthropometric measures, overweight and obesity. In: *Health and Social Care Information Centre* (2013)
9. Singh, A.S., Mulder, C., Twisk, J.W.R., Van Mechelen, W., Chinapaw, M.J.M.: Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obes. Rev.* **9**, 474–488 (2008)
10. Dietz, W.H.: Health consequences of obesity in youth: childhood predictors of adult disease. *Pediatrics* **101**, 518–525 (1998)
11. Engeland, A., Bjørge, T., Sjøgaard, A.J., Tverdal, A.: Body mass index in adolescence in relation to total mortality: 32-year follow-up of 227,000 Norwegian boys and girls. *Am. J. Epidemiol.* **157**, 517–523 (2003)
12. Butland, B., et al.: *Tackling Obesities: Future Choices – Project Report*, 2nd edn. (2007)

13. Freedman, D.S., Mei, Z., Srinivasan, S.R., Berenson, G.S., Dietz, W.H.: Cardiovascular risk factors and excess adiposity among overweight children and adolescents: the Bogalusa heart study. *J. Pediatr.* **150**, 12–17 (2007)
14. Cole, T.J., Bellizzi, M.C., Flegal, K.M., Dietz, W.H.: Establishing a standard definition for child overweight and obesity worldwide. *BMJ* **320**, 1–6 (2000)
15. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
16. Szolovits, P., Patil, R.S., Schwartz, W.B.: Artificial intelligence in medical diagnosis. *Ann. Intern. Med.* **108**(1), 80–87 (1988)
17. Ishak, W.H.W., Siraj, F.: Artificial intelligence in medical application: an exploration. *Health Inform. Eur. J.* **16**, 1–9 (2008)
18. Jarvis-Selinger, S., Bates, J., Araki, Y., Lear, S.A.: Internet-based support for cardiovascular disease management. *Int. J. Telemed. Appl.* **2011**, 9 (2011)
19. Kumar, K., Thakur, G.S.M.: Advanced applications of neural networks and artificial intelligence: a review. *Int. J. Inf. Technol. Comput. Sci.* **4**, 57 (2012)
20. Jervis, B.W., et al.: Artificial neural network and spectrum analysis methods for detecting brain diseases from the CNV response in the electroencephalogram. *IEE Proc. - Sci. Meas. Technol.* **141**, 432–440 (1994)
21. Mantzaris, D.H., Anastassopoulos, G.C., Lymberopoulos, D.K.: Medical disease prediction using artificial neural networks. In: 8th IEEE International Conference on BioInformatics and BioEngineering, BIBE 2008 (2008)
22. Khalaf, M., et al.: Machine learning approaches to the application of disease modifying therapy for sickle cell using classification models. *Neurocomputing* **228**, 154–164 (2017)
23. Singh, B., Tawfik, H.: A machine learning approach for predicting weight gain risks in young adults. In: Conference Proceedings of 2019 10th International Conference on Dependable Systems, Services and Technologies, DESSERT 2019 (2019)
24. Michie, D., Spiegelhalter, D.J., Taylor, C.C., Campbell, J. (eds.): *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, Upper Saddle River (1994)
25. Dugan, T.M., Mukhopadhyay, S., Carroll, A., Downs, S.: Machine learning techniques for prediction of early childhood obesity. *Appl. Clin. Inform.* **6**, 506–520 (2015)
26. Novak, B., Bigec, M.: Application of artificial neural networks for childhood obesity prediction. In: Proceedings - 1995 2nd New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, ANNES 1995 (1995)
27. Adnan, M.H.B.M., Husain, W., Damanhoori, F.: A survey on utilization of data mining for childhood obesity prediction. In: 8th Asia-Pacific Symposium Information Telecommunication Technologies (2010)
28. Hariz, M., Adnan, B.M., Husain, W., Aini, N., Rashid, A.: Parameter identification and selection for childhood obesity prediction using data mining. In: 2nd International Conference on Management and Artificial Intelligence (2012)
29. Zhang, S., Tjortjis, C., Zeng, X., Qiao, H., Buchan, I., Keane, J.: Comparing data mining methods with logistic regression in childhood obesity prediction. *Inf. Syst. Front.* **11**, 449–460 (2009)
30. Fergus, P., et al.: A machine learning approach to measure and monitor physical activity in children. *Neurocomputing* **228**, 220–230 (2017)
31. Smith, K., Joshi, H.: The Millennium Cohort Study. *Popul. Trends* **107**, 30–34 (2002)
32. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)