Application of Machine Learning Techniques to Predict Teenage Obesity Using Earlier Childhood Measurements from Millennium Cohort Study

Balbir Singh
School of Built Environment, Engineering and Computing,
Leeds Beckett University, Leeds, United Kingdom
B.Singh@leedsbeckett.ac.uk

Anatoliy Gorbenko
School of Built Environment, Engineering and Computing,
Leeds Beckett University, Leeds, United Kingdom
A.Gorbenko@leedsbeckett.ac.uk

Anna Palczewska
School of Built Environment, Engineering and Computing,
Leeds Beckett University, Leeds, United Kingdom
A.Palczewska@leedsbeckett.ac.uk

Hissam Tawfik
Department of Electrical Engineering, University of
Sharjah, Sharjah, UAE
htawfik@sharjah.ac.ae

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 the risk of developing serious illnesses such as type 2 diabetes, respiratory problems, heart disease, stroke, and even death. It is essential to detect childhood obesity as early as possible since children who are either overweight or obese in their younger age tend to stay obese in their adult lives. This research utilises the vast amount of data available via UK's millennium cohort study to construct machine learning driven framework to predict young people at the risk of becoming overweight or obese. The focus of this paper is to develop a framework to predict childhood obesity using earlier childhood data and other relevant features. The use of novel data balancing technique and inclusion of additional relevant features resulted in sensitivity, specificity, and F1-score of 77.32%, 76.81%, and 77.02% respectively. The proposed technique utilises easily obtainable features making it suitable to be used in a clinical and non-clinical environment.

CCS CONCEPTS

• **Computing methodologies**; • **Machine learning**; • **Learning paradigms**; • **Supervised learning**; • **Supervised learning by classification**; • **Cross-validation.**;

KEYWORDS

Body mass index, classification, machine learning, obesity, millennium cohort study

ICCBDC 2023, August 17–19, 2023, Manchester, United Kingdom

ACM ISBN 979-8-4007-0733-9/23/08.
https://doi.org/10.1145/3616131.3616139

ACM Reference Format:
Balbir Singh, Anatoliy Gorbenko, Anna Palczewska, and Hissam Tawfik. 2023. Application of Machine Learning Techniques to Predict Teenage Obesity Using Earlier Childhood Measurements from Millennium Cohort Study. In *2023 7th International Conference on Cloud and Big Data Computing (ICCBDC 2023), August 17–19, 2023, Manchester, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3616131.3616139

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 [7]. 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 the overweight and obese category is likely to rise to 41% by the year 2030. The report concludes by recommending behaviour change interventions which can result in saving money in the long run because 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 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 [12]. 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 [24]. 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 [26]. Similar claims were made by [2] and [22]. 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 [29]. However, [26] 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 [21]. 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 [27]. Obesity is strongly linked with other negative health conditions such as type 2 diabetes, cardiovascular diseases, cancers and even death [6], [11], [3], [9]. 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 apply, and evaluate machine learning algorithms to classify adolescents at a risk of becoming overweight or obese using early childhood BMI as input features. 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 [5]. 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).

2 RELATED RESEARCH

Only a small number of studies conducted to predict childhood obesity were identified through literature searches which applied models such as decision tree, logistic regression, random forest, artificial neural networks, and gradient boosting machines to datasets from surveys other than the MCS data. This is possibly due to the complexity associated with data collected by the millennium cohort study since it is very challenging to create a coherent dataset of desired features from 6 separate surveys. Furthermore, several studies reported reasonable prediction accuracies, but their precision metrics were low indicating the presence of a large number of false positives in their predictions as reported by Hammond et al. [15], and Pang et al. [25]. Hence, it is important to utilise some complex prediction algorithm using machine learning methods or even ensemble of predictors to implement a robust and accurate prediction system rather than using simpler techniques such as linear regression or other statistical methods as suggested by Michie et al. [20].

Machine learning techniques have been applied in several studies for obesity prediction. Dugan et al. applied ML to data collected from children before the age of 2 to predict future obesity [10]. In that study, they used data collected on children prior to the second birthday using a clinical support system. They reported using the ID3 algorithm for decision trees without pruning. The other algorithms tested by them 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. Novak & Bigec [23] 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. Adnan et al. [1] discussed several algorithms for predicting childhood obesity. They recommended the suitability of ANNs, Naïve Bayes and Decision Trees. Genetic Algorithms have also been tried as in the study done by Adnan et al. [16] to improve the prediction accuracy. 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 [31] 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 [13] 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). They also suggested the investigation of Deep Learning techniques for future work. Some studies used traditional statistical methods to predict obesity. Steur et al. [28] used Prevention and Incidence of Asthma and Mite Allergy (PIAMA) study dataset on Dutch children born in 1996/67. The limitation of this method was a poor positive predictive or precision score highlighting a significant amount of false positives in the prediction results. The others using statistical methods to predict overweight, and obese children were Druet et al.[8], Graversen et al. [14], Levine et al. [17], Manios et al. [18], and Weng et al. [30].

This research will focus to investigate, design, apply, and critically evaluate the application of machine learning techniques for the analysis and prediction of risk of excess weight in young people using data from the millennium cohort study.

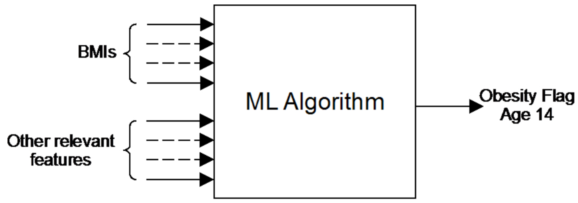

Figure 1: Predicting teenage obesity flag using historic BMIs and other relevant features

3 MACHINE LEARNING ALGORITHMS RELEVANT TO THIS STUDY

Machine learning algorithms can be divided into three broad categories: Unsupervised learning, supervised learning, and reinforcement learning. For this study, only a range of relevant and commonly used supervised classification algorithms have been employed. They include Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), artificial neural networks based Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Naive Bayes (NB), and Extreme Gradient Boosting Machines (XGB).

4 THE PROPOSED METHOD

A range of classification metrics were explored to measure the relative performance of a range of classifiers. For example, a confusion matrix is a good indicator of individual class accuracy and additional metrics such as precision, recall, F-scores, specificity, and receiver operator characteristics can be derived from a confusion matrix and will be incorporated. Other techniques such as hyper parameter tuning, and rgularisation will be employed to determine the robustness of classifiers and to avoid overfitting.

One important aspect of the investigation is to predict children at risk around their teenage years using the input features from the adiposity rebound age. This is an important aspect in children's growth since children who are overweight or obese around the adiposity rebound age, they tend to stay obese in their adult lives as reported by Mead et al. [19]. During this part of the growth curves, children's BMI has a non-linear relationship to their age. Therefore, normal statistical methods are unable to predict children at risk with a high degree of accuracy. Thus, there is another reason to use machine learning algorithms to successfully and robustly make predictions using additional features such as children's birth weight, parents BMI, gender, and parents' socio-economic status, etc. to identify children at the risk of developing obesity as shown in Figure 1.

5 DATASET SELECTION, CLEANING, AND AUGMENTATION

5.1 The MCS dataset

For any successful machine learning application, the selecting of appropriate data subsets and data pre-processing are the two most crucial tasks. Selecting an appropriate dataset is just as important as maximising the machine learning algorithm accuracy. The data for this study are used from UK's Millennium Cohort Study (MCS) (Smith & Joshi, 2002a). 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. Over 24000 families were contacted to carry out the survey during the first wave which was conducted between 2001 and 2002. 18552 families were successfully recruited for the first survey when the babies were nine months old. Five more surveys are available when the children were aged 3 years (MCS2, 2004/5 n=15,590), 5 years (MC3, 2006 n=15,246), 7 years (MCS4, 2008 n=13 587), 11 years (MCS5, 2012 n=13,287) and 14 years (MCS6, Jan 2015 to April 2016, n=11726). These survey sweeps are tabulated in Table 1.

The dataset had a significant number of missing values for the body mass index measurements. Table 2 shows the total number of cases and missing values.

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. Each variable has some missing values. However, different cases have different missing variables. Since this is health data, and the BMI can fluctuate from survey to survey, there is no accurate way of replacing the missing data. Besides, considering the significantly large size of the dataset, it would be useful to drop all those cases where there are any missing values. There are 7237 instances which have no missing values after dropping the cases with any missing values.

5.2 Outliers and missing values treatment

After examining the dataset closely, it is evident that some of the values seem to fall beyond the first and third quartiles. This suggests that there are a significant number of outliers in the data set. Some of the values differ from the mean significantly which suggests that these values may have resulted from human error at the time of data collection. Since it is a health dataset and each observation belongs to a real person, it would be beneficial to drop those cases with outliers. This will still result in a dataset that is of a significantly appropriate size, suitable to be used for machine learning. The outlier values will be identified using the Inter Quartile Range (IQR) and the first and third quartiles. A plot of dataset BMI values with outliers is shown in Figure 2. These outliers are identified and those cases containing these are removed from the dataset. The cleaned dataset is shown in Figure 3.

5.3 Data Augmentation

The dataset is severely imbalanced. The number of cases in the at-risk class (overweight or obese) is much smaller than the normal class made up of participants with healthy BMI classification. To deal with this class imbalance issue, one approach which is normally used is to synthesise new class instances for the minority class using SMOTE [4] 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 Figure 4. However, for robustness, a trained network must be tested on an original, imbalanced dataset.

Table 1: Millennium Cohort Study sweeps

Survey	MCS1	MCS2	MCS3	MCS4	MCS5	MCS6
Age	9 months	3 years	5 years	7 years	11 years	14 years

Table 2: Millennium Cohort Study univariate analysis of BMI values

Survey age	non-zero cases	Mean	Std Dev	Missing values		No. of extremes[1]	
				Count	Percent	Low	High
3 years	14452	16.7700	1.68150	5063	25.9	138	432
5 years	15459	16.0585	2.93298	4056	20.8	266	235
7 years	13816	16.6453	2.36758	5699	29.2	29	654
11 years	13047	19.2474	3.68190	6468	33.1	9	577
14 years	11110	21.4508	4.14464	8405	43.1	4	528

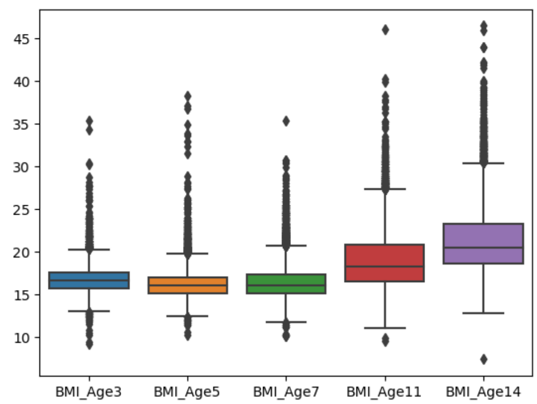

Figure 2: BMI box plot showing outliers

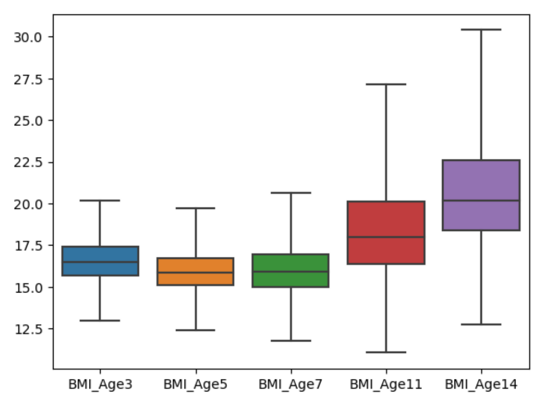

Figure 3: BMI box plot showing after outliers removal

For this reason, a sample of the original imbalanced data can be kept for testing the ML algorithm accuracy. The same machine learning algorithms used in the case of imbalanced data classification can then be used to classify synthetically balanced dataset and results can be compared. One issue with this approach is that the SMOTE

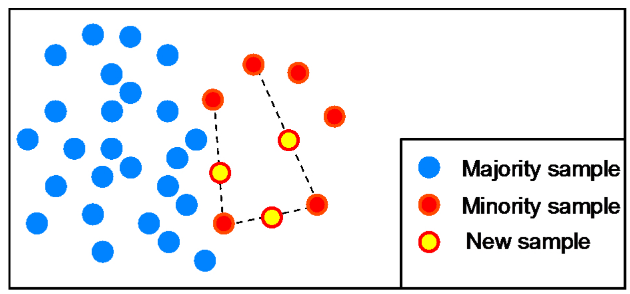

Figure 4: Generating synthetic data points using SMOTE

algorithm generates data values which are not 'real'. Dealing with health data requires rigorous training and testing on 'real' data. Therefore, the augmentation of data for this reason was based on 'natural' way of dealing with the imbalance rather than generating 'artificial' data. To implement this natural approach, the majority class was randomly subdivided into subsections, each having a size very close to the minority class. Each subsection of data was trained and tested using a 10-fold cross validation.

6 EXPERIMENTAL RESULTS

6.1 Classification to predict obesity flags using the unbalanced data

This section deals with the classification of overweight and obese children (at risk category) using the unbalanced dataset and commonly used relevant machine learning algorithms. The proportion of overweight and obese children is much lower than the normal weight children (normal class). This makes it extremely difficult to predict children at the risk of developing obesity with a high degree of accuracy. Because of the severe data imbalance, even a small proportion of false positives gives rise to a seriously degraded at risk class accuracy metrics. Some of the robust algorithms such as Decision Tree (DT), Random Forest (RF), and Extremely Gradient Boosted machines (XGB) are likely to perform marginally better than the others because of the built in penalties for misclassification.

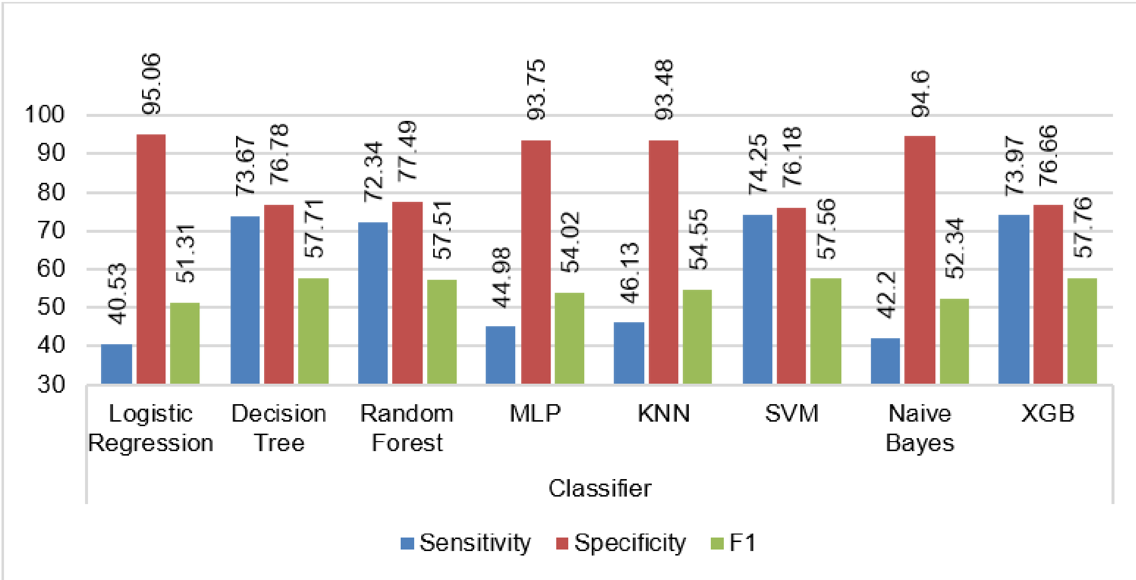

Figure 5: Predicting age 14 obesity flag using age 7 BMI values

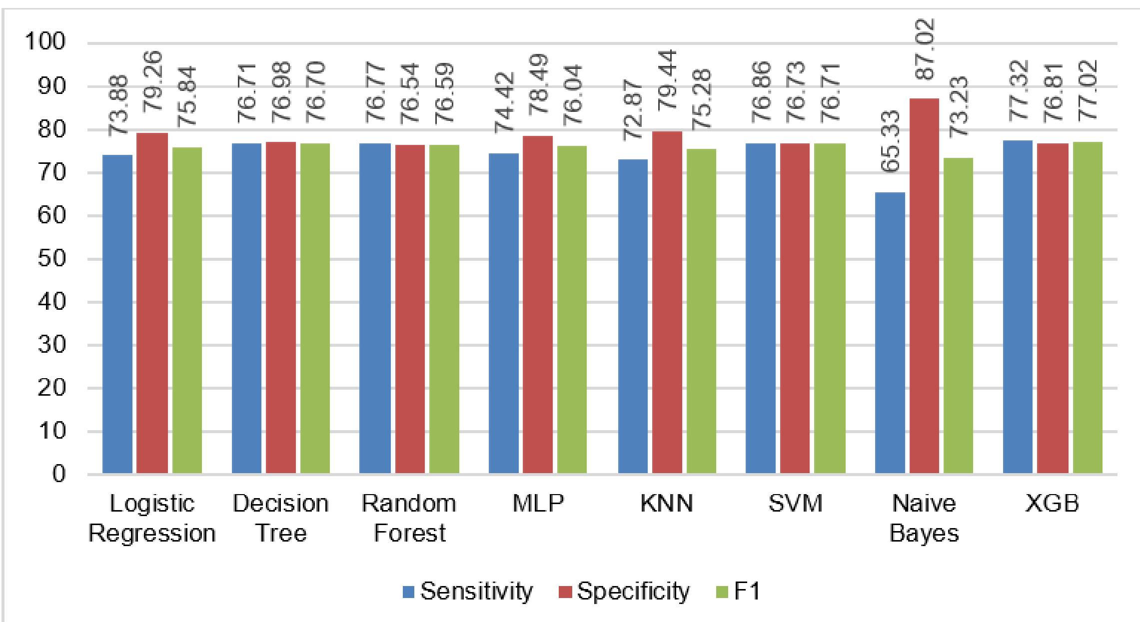

Figure 6: Predicting age 14 obesity flag using age 7 BMIs and other relevant features

The following sections evaluate the performance of the relevant machine learning models using the unbalanced dataset. The timelapse between the input data age and the target age is 7 years, which is relatively large. The prediction accuracies using the unbalanced data are just over 70%. The F1 score is still below 60% indicating the presence of a large number of false positives. The results for this test are plotted as bar graph in Figure 5.

6.2 Obesity classification employing data balancing with proposed under-sampling techniques using the random under-sampling algorithm

This section discusses the prediction of obesity flags using the historic BMI values and other relevant features after under-sampling the majority class. As discussed previously, the proportion of at risk class cases is much smaller than the normal weight subjects. Also, there are a significant number of cases with missing values for several features. A significant number of cases have some feature values which can be categorised as potential outliers. There are several stages involved in implementing this framework including missing values treatment, outliers treatment, feature selection, under-sampling of the majority class, hyper parameter tuning, 10-fold cross validation (CV) and finally averaging the results produced by the CV process for each of the sub datasets.

The results to predict the age 14 obesity status using the BMI values from age 7 and other relevant features are plotted in Figure 6. The chart shows that most ML algorithms perform well except for the Naïve Bayes, which tends to favor the normal class. There is a significant increase in the F1-score indicating the reduction in false positives.

7 CONCLUSION AND FUTURE WORK

Although BMI in adults is a good indicator of obesity, children's obesity flags not only depend on their BMI but also their age and sex. In this study we focused on predicting obesity flags using not only the BMI values but also other relevant features such as parents BMI, IMD values, ethnicity, birth weight, gender and housing status were combined with BMI values. Most ML models evaluated in this study performed well with the exception of Naïve Bayes. Using the proposed method and the XGB classifier, the sensitivity, specificity, and F1-scores improved from 73.97%, 76.66%, and 57.76% respectively to 77.32%, 76.81%, and 77.02%. The results indicate that a significant increase in the F1-score highlights the model's ability to minimize the false positives. This suggests that this technique can be used both for clinical assessment and population monitoring. The future work will involve the inclusion of explainable AI to highlight the contribution of features putting an individual at the risk of developing obesity. Additionally, health professionals and public participants will be consulted to assess the suitability of this technique in clinical and non-clinical environments.

ACKNOWLEDGMENTS

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] Adnan, M.H.B.M. *et al.* 2010. A survey on utilization of data mining for childhood obesity prediction. *8th Asia-Pacific Symposium on Information and Telecommunication Technologies.* (2010).
[2] Blüher, S. *et al.* 2011. Age-specific stabilization in obesity prevalence in German children: A cross-sectional study from 1999 to 2008. *International Journal of Pediatric Obesity.* 6, sup3 (Jun. 2011), e199-206. DOI:https://doi.org/10.3109/17477166.2010.526305.
[3] Butland, B. *et al.* 2007. *Tackling Obesities: Future Choices – Project report 2nd Edition.*
[4] Chawla, N. V *et al.* 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research.* (2002). DOI:https://doi.org/10.1613/jair.953.
[5] Cole, T.J. *et al.* 2000. Establishing a standard definition for child overweight and obesity worldwide. *BMJ 2000:320:1-6.* (2000).
[6] Dietz H William 1998. Health Consequences of Obesity in Youth: Childhood Predictors of Adult Disease. *Pediatrics.* (1998).
[7] Dobbs, R. *et al.* 2014. Overcoming obesity: An initial economic analysis. *McKinsey Global Institute.*
[8] Druet, C. *et al.* 2012. Prediction of childhood obesity by infancy weight gain: An individual-level meta-analysis. *Paediatric and Perinatal Epidemiology.* 26, 1 (Jan. 2012), 19–26. DOI:https://doi.org/10.1111/j.1365-3016.2011.01213.x.
[9] D.S., F. *et al.* 2007. Cardiovascular Risk Factors and Excess Adiposity Among Overweight Children and Adolescents: The Bogalusa Heart Study. *Journal of Pediatrics.* (2007).
[10] Dugan, T.M. *et al.* 2015. Machine Learning Techniques for Prediction of Early Childhood Obesity. *Applied Clinical Informatics.* (2015). DOI:https://doi.org/10.4338/aci-2015-03-ra-0036.
[11] Engeland, A. *et al.* 2003. Body mass index in adolescence in relation to total mortality: 32-Year follow-up of 227,000 Norwegian boys and girls. *American Journal of Epidemiology.* (2003). DOI:https://doi.org/10.1093/aje/kwf219.
[12] Fat, L.N. 2014. Children's body mass index, overweight and obesity. *Health Survey for England: Chapter 10.* (2014).
[13] Fergus, P. *et al.* 2017. A machine learning approach to measure and monitor physical activity in children. *Neurocomputing.* (2017). DOI:https://doi.org/10.1016/j.neucom.2016.10.040.
[14] Graversen, L. *et al.* 2015. Prediction of adolescent and adult adiposity outcomes from early life anthropometrics. *Obesity.* 23, 1 (Jan. 2015), 162–169. DOI:https://doi.org/10.1002/oby.20921.
[15] Hammond, R. *et al.* 2019. Predicting childhood obesity using electronic health records and publicly available data. *PLoS ONE.* 14, 4 (Apr. 2019). DOI:https://doi.org/10.1371/journal.pone.0215571.
[16] Hariz, M. *et al.* 2012. Parameter Identification and Selection for Childhood Obesity Prediction Using Data Mining. *2nd International Conference on Management and Artificial Intelligence* (2012).
[17] Levine, R.S. *et al.* 2012. Identifying infants at risk of becoming obese: Can we and should we? *Public Health.* 126, 2 (2012), 123–128. DOI:https://doi.org/10.1016/j.puhe.2011.10.008.
[18] Manios, Y. *et al.* 2013. Childhood Obesity Risk Evaluation based on perinatal factors and family sociodemographic characteristics: CORE Index. *European Journal of Pediatrics.* 172, 4 (Apr. 2013), 551–555. DOI:https://doi.org/10.1007/s00431-012-1918-y.
[19] Mead, E. *et al.* 2016. Predicting future weight status from measurements made in early childhood: A novel longitudinal approach applied to Millennium Cohort Study data. *Nutrition and Diabetes.* 6, (2016). DOI:https://doi.org/10.1038/nutd.2016.3.
[20] Michie, D. *et al.* eds. 1994. *Machine Learning, Neural and Statistical Classification.* Ellis Horwood.
[21] Moody, A. 2013. Adult anthropometric measures, overweight and obesity. *The Health and Social Care Information Centre.* (2013). DOI:https://doi.org/10.1016/j.jamda.2013.05.021.
[22] Moss, A. *et al.* 2011. Declining prevalence rates for overweight and obesity in German children starting school.
[23] Novak, B. and Bigec, M. 1995. Application of artificial neural networks for childhood obesity prediction. *Proceedings - 1995 2nd New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, ANNES 1995* (1995).
[24] Olds, T. *et al.* 2011. Evidence that the prevalence of childhood overweight is plateauing: data from nine countries. *International Journal of Pediatric Obesity.* 6, 5–6 (Oct. 2011), 342–360. DOI:https://doi.org/10.3109/17477166.2011.605895.
[25] Pang, X. *et al.* 2019. Understanding early childhood obesity via interpretation of machine learning model predictions. *Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019* (Dec. 2019), 1438–1443.
[26] Rokholm, B. *et al.* 2010. The levelling off of the obesity epidemic since the year 1999 - a review of evidence and perspectives. *Obesity Reviews.*
[27] Singh, A.S. *et al.* 2008. Tracking of childhood overweight into adulthood: A systematic review of the literature. *Obesity Reviews.*
[28] Steur, M. *et al.* 2011. Predicting the risk of newborn children to become overweight later in childhood: The PIAMA birth cohort study. *International Journal of Pediatric Obesity.* 6, 2–2 (Jun. 2011). DOI:https://doi.org/10.3109/17477166.2010.519389.
[29] Wabitsch, M. *et al.* 2014. Unexpected plateauing of childhood obesity rates in developed countries. *BMC Medicine.*
[30] Weng, S.F. *et al.* 2013. Estimating Overweight Risk in Childhood From Predictors During Infancy. *Pediatrics.* 132, 2 (Aug. 2013), e414–e421. DOI:https://doi.org/10.1542/peds.2012-3858.
[31] Zhang, S. *et al.* 2009. Comparing data mining methods with logistic regression in childhood obesity prediction. *Information Systems Frontiers.* (2009). DOI:https://doi.org/10.1007/s10796-009-9157-0.~