

# A Machine Learning Approach to Select the Type of Intermittent Fasting in Order to Improve Health by Effects on Type 2 Diabetes

Shula Shazman

*Department of Mathematics and Computer Science, The Open University, Derech Hauniversita 1, Raanana, Israel*

**Keywords:** Machine Learning, Decision Tree, Type 2 Diabetes, Insulin Resistance, Precision Medicine, Intermittent Fasting, Alternate Day Fasting.

**Abstract:** Intermittent fasting (IF) is the cycling between periods of eating and fasting. The main types of IF are: complete alternate-day fasting; time-restricted feeding (eating within specific time frames such as the most prevalent 16:8 fast, with 16 hours of fasting and 8 hours for eating); religious fasting such as the Ramadan (occurs one month per year, with eating taking place only after nightfall). IF can be effective in reducing metabolic disorders and age-related diseases by bringing about changes in metabolic parameters associated with type 2 diabetes. Questions do remain, however, about the effects of the different types of IF as a function of the age at which fasting begins, gender and severity of type 2 diabetes. In this paper we describe a machine learning approach to selecting the best type of IF to improve health in type 2 diabetes. For the purposes of this research, the health outcomes of interest are changes in fasting glucose and insulin. The different types of intermittent fast offer promising non-pharmacological approaches to improving health at the population level, with multiple public health benefits.

## 1 INTRODUCTION

Diabetes has become prevalent with changes in lifestyle, threatening to reduce life expectancy for humans around the globe. According to the International Diabetes Federation (IDF), there were 425 million people in the world with diabetes in 2017 – close to 1 in 11 people (Diabetes Atlas 8<sup>th</sup> edition, 2017).

There are two main types of diabetes – type 1 and type 2, both of which can lead to chronically high blood sugar levels. People with type 1 diabetes barely produce insulin at all, while those with type 2 diabetes produce insulin but do not respond to it as they should. Ninety to ninety-five percent of people living with diabetes have type 2 diabetes.

Type 2 diabetes is generally characterized by insulin resistance (IR), where the body does not fully respond to insulin. IR is now used as a screening index for primary prevention of type 2 diabetes. Using the Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) equation, IR can be estimated from fasting glucose and insulin levels. A high score of HOMA-IR indicates significant insulin resistance usually found in people with type 2

diabetes (Tang et al. 2015; Sharma and Fleming 2012).

Despite the awareness of the need for early diagnosis, prevention and treatment of diabetes, the IDF estimates that there will be 642 million people living with the disease by 2040, and another half as many who will be living with undiagnosed diabetes, at unknowing risk of its disabling, life-threatening complications (Diabetes Atlas 8<sup>th</sup> edition, 2017).

The cornerstone of type 2 diabetes management is a healthy diet, increased physical activity and maintaining healthy body weight. Oral medication and insulin are also frequently prescribed to help control blood glucose levels. A new precision medicine approach is also necessary for treatment of diabetes in addition to traditional approaches.

Daily calorie restriction regimens are still the most common diet strategies implemented for improving HOMA-IR (Wilding 2014). These are effective for weight loss in some individuals, but many people find this type of diet difficult, as it requires vigilant calorie counting on a daily basis, and the sense of never being able to eat freely throughout the day results in dieter frustration.

These impediments to the calorie restriction approach have brought about the introduction of

another approach termed intermittent fasting (IF), which has proven promising in achieving reduction in HOMA-IR, although not in all cases. IF is a form of time-restricted eating; it differs from calorie restriction in that the individual is only required to restrict energy intake during a portion of the day (typically 16 hours), and allows for free food consumption in the non-restricted hours. Alternate day fasting is a subclass of IF, consisting of a 'fast day' alternating with a 'feed day' (*ad libitum*, which is eating food as much as desired).

Previous studies and reviews provide an overview of IF regimens (Patterson and Sears, 2017; Malinowski et al., 2019; Ganesan et al., 2018; Barnosky et al., 2014), and summarize the evidence for their health benefits. Furthermore, they discuss physiological mechanisms by which IF might lead to improved health outcomes. They have not provided a clear answer, however, to the question of whether IF is always able to reduce HOMA-IR; that is, the conditions (age, gender, basal fasting glucose level, etc.) needed to make the IF effective for reducing HOMA-IR have not yet been deciphered. Moreover, no previous IF study has reported results per individual; results were reported on a group level only.

In today's era of precision medicine, the current study has been motivated to answer the question of whether a patient with prediabetes or diabetes could benefit from an intervention, reducing HOMA-IR or even eliminating the disease altogether. This study suggests a recommendation system based on individual data from human fasting intervention studies, where the health outcomes of interest are changes in metabolic parameters associated with type 2 diabetes. The system presented, based on a machine-learning approach, predicts which type of IF treatment can improve an individual's health by reducing insulin resistance and preventing or curing type 2 diabetes.

The results of this study provide a set of rules which can assist individual patients and their physicians in selecting the best IF intervention for their personal case.

## 2 METHODS

This study aims to predict whether a specific IF intervention would reduce the insulin resistance of an individual with prediabetes. The approach contains four basic steps: identifying required data, preparing and pre-processing, modeling the data and finally, training and testing.

### 2.1 Identifying Required Data

In order to answer the question of this study I asked for the individual data from authors of 25 published papers that performed randomized clinical trials investigating the IF effects on type 2 diabetes parameters. I received the individual data from 6 out of 25 papers (Halberg et al., 2005; Harvie et al., 2011; Harvie et al., 2013; Clifton et al., 2004; Chowdhury et al., 2016a; Chowdhury et al., 2016b). The rest of authors responded that they could not send the data due to participant confidentiality.

### 2.2 Preparing and Pre-processing the Data

#### 2.2.1 Selecting Individuals

From all the data received, 254 individuals with basal fasting glucose above 5 mmol/L (90 mg/dL) or BMI (Body Mass Index) above or equal to 25 were selected. The selection criteria were established since they indicate possible prediabetes (IDF Diabetes Care. Volume 42, Supplement 1, January 2019). The IDF's 2019 cutoff for fasting glucose indicating prediabetes is 100 mg/dL; we set the cutoff at 90 mg/dL. (The table containing the data may be found in Supplement 1 at the following link: <https://github.com/shulash3/intermittentFasting/blob/master/Supplementary1.xlsx>).

#### 2.2.2 Calculating HOMA-IR

The Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) has been proven to be a very sensitive test for indicating prediabetes (Sharma and Fleming 2012). Using the HOMA-IR equation, insulin resistance can be estimated from fasting glucose and insulin levels.

$$\text{HOMA-IR} = \text{Fasting Glucose} * \text{Fasting Insulin} \quad (1)$$

A high score of HOMA-IR indicates significant insulin resistance, usually found in people with type 2 diabetes.

For each of the 254 individuals, we calculated the HOMA-IR twice using Equation 1 – once for the basal values of fasting glucose and insulin and once for the values after the intervention. The difference between them represents the insulin resistance reduction.

#### 2.2.3 Intermittent Fasting Interventions

The dataset included 9 different types of interventions, e.g. continuous energy restriction – a

seven day-a-week trial; intermittent energy restriction – a two day-a-week trial allowing eating freely in the remaining 5 days; daily morning fasting; or fasting every second day. Part of the interventions contained specific diets such as carbohydrate restriction; high carbohydrate; or high monounsaturated.

Table 1: IF regimens.

Intervention name	Details	Reference
CER	Continuous Energy Restriction – 7-day-a-week trial; eating restricted calories every day.	Harvie et al. 2011
IER	Intermittent Energy Restriction 2-day-a-week trial; eating restricted calories only two days a week.	Harvie et al. 2011
DMF	Daily Morning Fasting; start eating at noon and finish at 20:00.	Chowdhury et al. 2016a and Chowdhury et al. 2016b
FESD	Fasting Every Second Day; eating only four days a week.	Halberg et al. 2005
IECR	Intermittent Energy and Carbohydrate Restriction; eating restricted calories only two days a week.	Harvie et al. 2013
IECR+PF	Intermittent Energy and Carbohydrate Restriction + free Protein and Fat; eating restricted calories only two days a week.	Harvie et al. 2013
DER	Daily Energy Restriction; eating restricted calories every day.	Harvie et al. 2013
High Carb	High Carbohydrate weight loss diet; eating restricted calories every day.	Clifton et al. 2004
High Mono	High Monounsaturated weight loss diet; eating restricted calories every day.	Clifton et al. 2004

Table 1 summarizes the different IF regimens included in this study. The reference to each regimen is shown in the table for further details.

### 2.2.4 Selecting the Features

The initial vector of features for every individual is shown in Figure 1A. The vector is composed of details regarding the individual (age, gender, weight, ethnicity, basal BMI, basal fasting glucose, fasting glucose after intervention, basal fasting insulin and fasting insulin after intervention) and details regarding the intervention (intervention name and duration).

Figure 1B describes the training vector, which is the vector after removing the features 'fasting glucose after intervention' and 'fasting insulin after intervention'. The calculation of the HOMA-IR difference is added to the training vector as follows: if the intervention is successful we expect a reduction in HOMA-IR; thus, if the HOMA-IR difference is greater than zero the assignment in the 'HOMA-IR difference' column is set to TRUE otherwise it is FALSE.

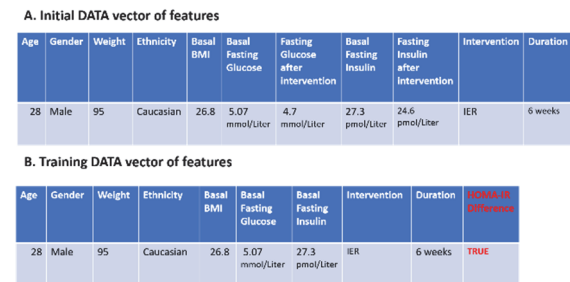


Figure 1: Vectors of initial and training features.

## 2.3 Modeling the Data

Data mining tools such as classification, clustering, association and neural networks solve problems by analyzing large volumes of data. Classification is possibly the most frequently used data mining technique. In this study we address a classification problem. Classification is the process of finding a set of models that describes and differentiates data classes and concepts, for the purpose of being able to use the model to predict the class whose label is unknown. There are many algorithms that can be used for classification, e.g. decision trees, neural networks, logistic regression and others. However, the decision tree classification with the Waikato Environment for Knowledge Analysis (Weka) is the simplest way to mine information from a database. Furthermore, decision trees are a way of representing a sequence of rules that lead to a class or value. A decision tree is a flowchart-like tree structure.

The decision tree algorithms J48, LMT (Logistic Model Tree), Random Forest and Random Tree as well as the Logistic Regression and Naïve Bayes classifiers were tested on the data in this study.

## 2.4 Training and Testing

The 254 samples in the training data were trained by the J48 decision tree (Weka 3.8.3). The implementation of the J48 decision tree in Weka 3.8.3 can handle categorical and numerical attributes like those found in our mixed dataset (Sewaiwar and

Verma 2015). The optimal number of features as a function of sample size is proportional to  $\sqrt{n}$  for highly correlated features (Hua et al. 2004). The features in the study shown here are highly correlated;  $\sqrt{254} = 15.9$  while the number of features is 9 (i.e. 9 attributes for 254 individuals is reliable).

Two test approaches were selected to validate the model – the Leave-One-Out and the 10-Fold cross-validations.

### 3 RESULTS

#### 3.1 HOMA-IR Reduction

All results of the six different classifiers – J48, LMT, Random Forest, Random Tree, Logistic Regression and Naïve Bayes – are shown in Table 2. The Area Under the Curve (AUC) of the 10-Fold test is shown in the first row of the table while the second row contains the data of the Leave-One-Out test. The AUC values of the 10-Fold test range between 0.67 and 0.75 while those of the Leave-One-Out range between 0.65-0.80. For both tests the AUC ranges are very small; we therefore conclude that for this case all six classifiers perform similarly. Finally, the J48 (C4.5) decision tree (Weka 3.8.3) is selected to model the data of this study. Although the advantage of Random Forest is to prevent overfitting by creating random subsets of the features and building smaller trees and then combining the subtrees, J48 is found to produce the most accurate prediction among the decision tree algorithms (Sewaiwar and Verma 2015). Furthermore, J48 is self-explanatory and easy to follow. The J48 decision tree is a predictive machine-learning model which selects a target value (HOMA-IR difference TRUE or FALSE) of an individual and an intervention based on the training vectors available. In the J48 decision tree, the different features (age, gender, weight, etc.) are denoted by the internal nodes of a decision tree, the branches between the nodes tell us the possible values that these features may have in the experimental samples (gender: male/female, etc.), while the terminal nodes tell us the final value of the dependent variable (TRUE or FALSE assigned for HOMA-IR difference).

The result of testing the J48 decision tree's model using the 10-Fold cross validation test show that the model predicts correctly in 72% of the cases, and the AUC is 0.7. Furthermore, the Leave-One-Out test achieves 78% accuracy and an AUC of 0.8. The results suggest that the model can predict correctly in

78% of the cases whether an intervention would help an individual improve their type 2 diabetes risk parameters by reducing HOMA-IR.

Table 2: AUC for different classifiers.

	J48	LMT	Random Forest	Random Tree	Logistic	Naive Bayes
10-Fold	0.7	0.75	0.75	0.67	0.79	0.73
Leave-One-Out	0.8	0.74	0.74	0.66	0.79	0.72

The visualization of the complete J48 decision tree is found in Figure 2A, with detailed views shown in Figures 2B, 2C, 2D and 2E.

A larger figure of 2A can be found in Supplement 2, at the following link: <https://github.com/shulash3/intermittentFasting/blob/master/Supplementary2.png>. In Figure 2A the relative positions of Figures 2B, 2C, 2D, 2E are visible.

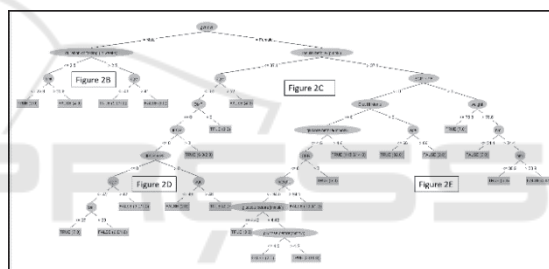


Figure 2A: Visualization of complete J48 decision tree.

The first node in the tree, as shown in Figure 2A, is the gender feature, indicating that this attribute is the most informative one for the decision. Interestingly we also notice in Figure 2A that for males the most important feature in determining whether an intervention would be effective is the fast duration while for females the basal fasting insulin level is reported as the most important feature. In figures 2B-2E TRUE (colored green) indicates success in reducing HOMA-IR while FALSE (colored red) indicates no reduction.

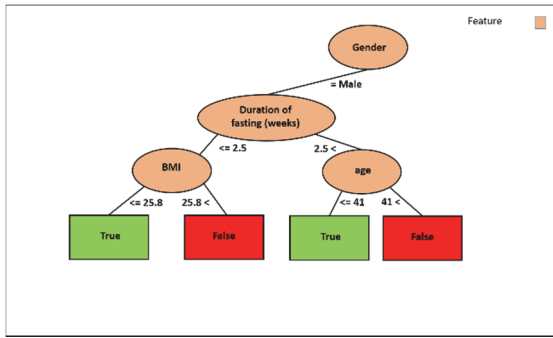


Figure 2B: Male sub-decision tree.

It can be observed from Figure 2B that men are indifferent to any of the intervention types, but the duration of the intervention plays an important role. Short duration of intervention and lower BMI or long duration of intervention and younger age lead to the success of the intervention (reducing HOMA-IR). Reasonably, attributes like lower BMI and younger age make it easier to reduce HOMA-IR.

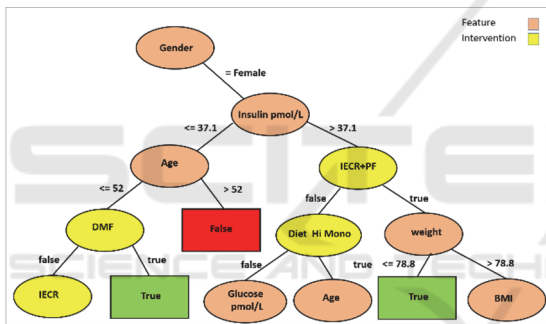


Figure 2C: Female sub-decision tree.

In Figure 2C the different interventions appear to be part of the decision nodes. The interventions are colored yellow while the features are light brown.

The tree view on the female side is more complex. This may be because there are more women in the dataset than men. In Figure 2C the different interventions appear to be part of the decision nodes. These are organized hierarchically beginning with DMF followed by IECR or beginning with IECR followed by the Hi Mono diet.

In Figure 2D we see a hierarchical structure of the interventions ordered by their success in improving HOMA-IR, beginning with DMF, IECR and then IECR+PF.

Interestingly in Figure 2E there is a node where lower BMI leads to an unsuccessful intervention. This evidence should be further investigated.

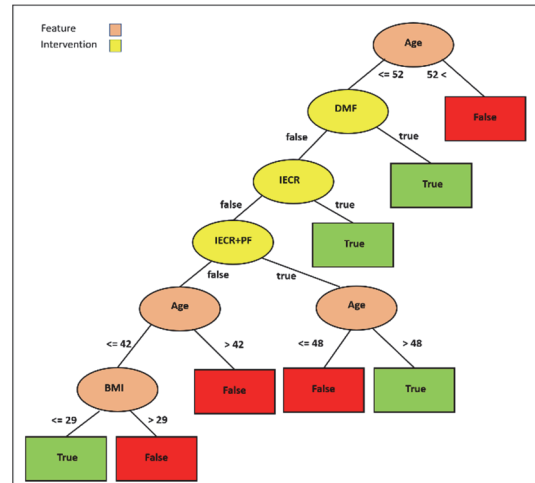


Figure 2D: Female left sub-decision tree.

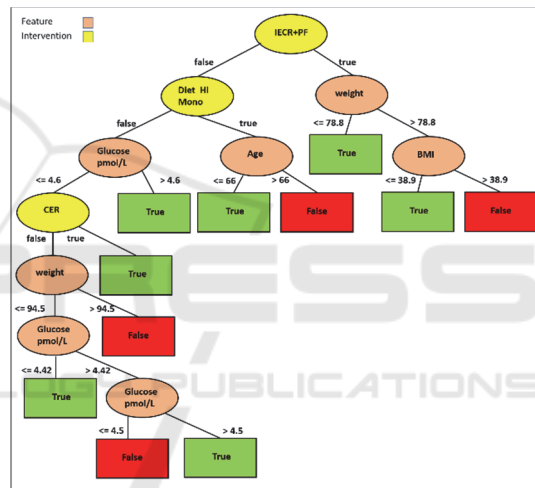


Figure 2E: Female right sub-decision tree.

### 3.2 Fasting Glucose or Fasting Insulin Reduction

Prediction results of fasting glucose reduction and fasting insulin reduction taken separately instead of HOMA-IR reduction are shown in Table 3.

Table 3: Summary of AUC results for improving type 2 diabetes risk parameters.

	HOMA-IR reduction	FASTING Glucose reduction	FASTING Insulin reduction
10-Fold Cross Validation test	0.7	0.6	0.55
Leave- One- Out test	0.8	0.6	0.6

The results in Table 3 show that the HOMA-IR improvement prediction is more effective than the prediction of the fasting glucose or the fasting insulin reduction taken separately. As shown in Equation 1, the HOMA-IR calculation is based on both fasting glucose and fasting insulin.

### 3.3 Random Classification

In order to validate that these results for HOMA-IR cannot be achieved randomly, I reordered the values in the HOMA-IR column in an arbitrary way. The proportion between the TRUE values and the FALSE values remained the same as in the original column. Then I trained and tested the data once more. The results of the random tests were much lower in AUC compared with the original data. The results for the 10-Fold cross validation test were 0.56 AUC compared with 0.7 in the original data. The results of the Leave-One-Out test were even more significant – 0.61 AUC compared with 0.8 in the original data. Those results suggest that the model predictions cannot be achieved randomly.

### 3.4 Feature Selection

An interesting question is whether all the features shown in Figure 1B are needed for the prediction. To test this a feature selection test was performed on the data. In each test a different feature was excluded. The AUC results are shown in Table 4.

Table 4: Features selection – AUC results of J48 Decision tree.

Excluded Feature	10-Fold Cross Validation test	Leave-One-Out test
None	0.7	0.8
Age	0.68	0.7
Gender	0.68	0.62
Weight	0.64	0.73
Ethnic	0.68	0.74
Basal BMI	0.69	0.77
Basal Fasting Glucose	0.65	0.73
Basal Fasting Insulin	0.62	0.6

The feature in every row of Table 4 other than the first, is excluded and AUC is calculated without this feature. None of the features is redundant, since the higher AUC is shown when all features are trained. Furthermore, J48 training and testing with data that does not have more than one feature (from the list of

all nine features) resulted in even lower AUC values than the values shown in Table 4.

## 4 CONCLUSIONS

Even a single fasting interval in humans (e.g., overnight) can reduce basal concentrations of metabolic biomarkers such as insulin and glucose, associated with chronic disease. For example, patients are required to fast for 8–12 hours before blood draws to achieve steady-state fasting levels for many metabolic substrates. Recent studies suggest that intermittent fasting regimens may be a promising approach to losing weight and improving metabolic health for people who can safely tolerate intervals of non-eating, or eating very little, for certain hours of the day, night, or days of the week. Furthermore, these eating regimens may offer promising non-pharmacological approaches to improving health at the population level with multiple public health benefits.

This study does not investigate weight loss; however, it offers a recommendation system based on data from several clinical trials for selecting the optimal intervention to improve the health of prediabetes individuals by reducing their type 2 diabetes risk parameters. The procedure in this study is built using a machine learning approach and is represented by a decision tree. The decision rules derived from the tree are shown in Figures 2B-2E and in the figure in Supplement 2 (which contains the entire picture of decision rules). First, we observe that males and females have a different set of rules, since the node gender comes first in the tree. Males are indifferent to the type of intervention; the success of the intervention in males, however, is dependent on IF duration. For example, if the duration of the intervention is less than or equal to 2.5 weeks then the success of the intervention depends on BMI. Males with a smaller BMI will be more likely to have a successful intervention. On the other hand, if the duration of the intervention is more than 2.5 weeks for males then age will be important to its success. Reasonably, younger age will serve as a benefit. As for females, most important for a successful intervention is the level of basal fasting insulin. In the case of a female with a basal fasting of less than or equal to 37.1 pmol/L (for moderate insulin resistance the fasting insulin should be in the range of 18–48 pmol/L) and age exceeding 52, there is no intervention in the dataset that can assist in improving HOMA-IR. Additional data from clinical trials can be useful for expanding the recommendation system and

applying it to a wider population. Furthermore, a wider dataset will make if possible, to answer a more interesting question, which is to predict what the best fasting approach would be considering one's age, gender, etc. An algorithm which would answer the above question would certainly assist physicians in providing personalized medical advice to their patients.

## ACKNOWLEDGEMENTS

I wish to thank Michelle Harvie, Nils Halberg, Flemming Dela, Peter M. Clifton, Eric Ravussin, Leonie Kaye Heilbronn and Enhad Chowdhury for their assistance in this study by sending the individual data from their published clinical trial papers.

## REFERENCES

- Barnosky, A. R., Hoddy, K. K., Unterman, T. G. and Varady, K. A., 2014. Intermittent fasting vs daily calorie restriction for type 2 diabetes prevention: a review of human findings. *Translational Research Volume 164, Number 4*.
- Chowdhury, E. A., Richardson, J. D., Holman, G. D., Tsintzas, K., Thompson, D. and Betts, J. A. 2016. The causal role of breakfast in energy balance and health: a randomized controlled trial in obese adults. *Am J Clin Nutr 2016;103:747–56. Printed in USA*.
- Chowdhury, E. A., Richardson, J. D., Tsintzas, K., Thompson, D. and Betts, J.A. 2016. Effect of extended morning fasting upon ad libitum lunch intake and associated metabolic and hormonal responses in obese adults. *International Journal of Obesity (2016) 40, 305–311 Macmillan Publishers Limited*.
- Clifton, P. M, Noakes, M. and Keogh, J. B. 2004. Very Low-Fat (12%) and High Monounsaturated Fat (35%) Diets Do Not Differentially Affect Abdominal Fat Loss in Overweight, Nondiabetic Women. *Human Nutrition and Metabolism*.
- Ganesan, K., Habboush, Y. and Sultan, S. 2018. Intermittent Fasting: The Choice for a Healthier Lifestyle. *Cureus 10(7): e2947. DOI 10.7759/cureus.2947*.
- Halberg, N., Henriksen, M., Søderhamn, N., Stallknecht, B., Ploug, T., Schjerling, P. and Dela, F. 2005. Effect of intermittent fasting and refeeding on insulin action in healthy men. *J Appl Physiol • VOL 99 • DECEMBER 2005*.
- Harvie, M. N., Pegington, M., Mattson, M. P., Frystyk, J., Dillon, B., Evans, G., Cuzick, J., Jebb, S. A., Martin, B., Cutler, R. G., Son, T. G., Maudsley, S., Carlson, O. D., Egan, J. M., Flyvbjerg, A. and Howell, A. 2011. The effects of intermittent or continuous energy restriction on weight loss and metabolic disease risk markers: a randomized trial in young overweight women. *International Journal of Obesity (2011) 35, 714–727 Macmillan Publishers Limited*.
- Harvie, M. N., Wright, C., Pegington, M., McMullan, D., Mitchell, E., Martin, B., Cutler, R. G., Evans, G., Whiteside, S., Maudsley, S., Camandola, S., Wang, R., Carlson, O. D., Egan, J. M., Mattson, M. P. and Howell, A. 2013. The effect of intermittent energy and carbohydrate restriction v. daily energy restriction on weight loss and metabolic disease risk markers in overweight women. *British Journal of Nutrition (2013), 110, 1534–1547*.
- Hua, J., Xiong, Z., Lowey, J., Suh, E. and Dougherty, E. R. 2004. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics*.
- International Diabetes Federation 2017. Diabetes Atlas 8<sup>th</sup> edition.
- International Diabetes Federation. Diabetes Care. *Volume 42, Supplement 1, January 2019*
- Malinowski, B., Zalewska, K., Węsierska, A., Sokołowska, M. M., Socha, M., Liczner, G., Pawlak-Osińska, K. and Wiciński, M. 2019. Intermittent Fasting in Cardiovascular Disorders-An Overview. *Nutrients*.
- Patterson, R. E. and Sears, D. D. 2017. Metabolic Effects of Intermittent Fasting. *Annual Review of Nutrition*.
- Sewaiwar, P. and Verma, K.K. 2015. Comparative Study of Various Decision Tree Classification Algorithm Using WEKA. *International Journal of Emerging Research in Management & Technology ISSN: 2278-9359 (Volume-4, Issue-10)*.
- Sharma, S. and Fleming, E. 2012. Use of HbA1C testing to diagnose pre-diabetes in high risk African American children: A comparison with fasting glucose and HOMA-IR. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews 6 (2012) 157–162*.
- Tang, Q., Li, X., Song, P. and Xu, L. 2015. Optimal cut-off values for the homeostasis model assessment of insulin resistance (HOMA-IR) and pre-diabetes screening: Developments in research and prospects for the future. *Drug Discoveries & Therapeutics. 2015; 9(6):380-385*.
- Wilding, J. P. 2014. The importance of weight management in type 2 diabetes mellitus. *The Authors International Journal of Clinical Practice Published by John Wiley & Sons Ltd*.