By ensembling Bootstrap over KNN, a unique approach is proposed for improving the accuracy of diabetic healthy diet recommendations

Srinivas Manikanta [1], A.Rama [2*]

[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, TamilNadu, India, 602105.

[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,Saveetha University, chennai, TamilNadu, India, 602105.

ABSTRACT

Aim: The Aim of this paper is to enhance the accuracy in diabetic healthy diet recommendation system by Bootstrap ensembling over KNN. **Materials and Methods :** Bootstrap ensembling(N=10) and KNN algorithm (N=10) n was iterated at different times for predicting accuracy percentage of accidents happened. Two sample groups are taken into consideration and tested, G-power is calculation which contains two different groups, alpha (0.05), power (90%) and environment ratio**. Results:** It was observed that the Bootstrap Ensembling algorithm obtains accuracy as 70.02%. The Bootstrap Ensembling technique appears to have better significance than the support vector machine algorithm . **Conclusion**: The result proves that the Bootstrap Ensembling technique approaches with varying seed value have significant improvement in diet recommendation.

Keywords : Diabetes, Diet ,KNN , Bootstrap Ensembling , Accuracy.

INTRODUCTION

Diabetes is one of the fastest growing diseases in the world. To control this, a proper diet and regular exercise is necessary. Healthy diet is the most essential approach to prevent disease. This paper mainly describes the personal diabetic diet recommendation. The System developed is capable of generating a dietetic plan and suggests suitable food according to the user's health conditions and recommends a healthy diet for diabetes (Batra, Roy, and Panda 2020)**.** Diabetes is characterizedby means ofexcessive blood sugar over extended periods. Diabetes causesheadaches, where acute headaches can encompass hyperosmolar hyperglycemic state, diabetic ketoacidosis, or may be death. Serious long-time periodheadachesencompasspersistent kidney ailment, foot ulcers, harm to the eyes, cardiovascular ailment, and stroke. Diabetes happensbecause ofboth the incapacity of the pancreas to supplysufficient insulin,or the frame cells improperly responding to the insulin produced. Type 1, Type 2, and Gestational diabetes are the 3importantsorts of diabetes mellitus, even thoughthere may bea set of differentprecise types. Type 1 diabetes effects from the failure of pancreas to supplysufficient insulin because of the lack of beta cells as a result

of an autoimmune response. Type 2 diabetes starts off evolving with the insulin resistance, a circumstancewherein the cells fail to wellreply to insulin (Pitocco et al. 2012).

As the ailmentcontinues progressing, a loss of insulin may also additionally occur. A mixture of inadequateexercising and immoderateframe weight leads to drastic or unusual side effects. Gestational diabetes is 0.33, animportant form that may happenin a pregnant ladywithoutprecedingrecords of diabetes developing excessive blood sugar levels (Malaeb et al. 2019). Adequate dieting, with appropriatenutrients and normalexercising are very crucial in stopping or controlling diabetes. Prevention and remedy of diabetes by maintaining a wholesomeweight loss plan, normalbodilyexercising, everydayframe weight, and additionallyfending off use of tobacco. A low-fats weight loss plan, low-calorie weight loss plan, paleolithic weight loss plan, very low carbohydrate weight loss plan, uncooked food, and/or ketogenic weight loss plan can assistsave you or control diabetes (Pawlak 2017).(Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020)

Food is the basic need of every individual. Proper health diet selection is necessary to prevent diabetes. Diet selection varies from person to person, the intake of food selection to be selected according to the user's health condition (Lee 2014). The human body usually needs sugar for energy; however, too much sugar in blood can consequently damage the body, especially diabetes. Diabetes prevention would be the proper nutrition and healthy diet which balances sugar to the optimal level and maintains a healthy weight (Phanich, Pholkul, and Phimoltares 2010). The Food Pyramid is recommended to every diabetic patient. Food items are analyzed mainly based on the nutritional value for the dietetic plan. This paper recommends the diet for each patient menu with a list of recommended dishes on a daily basis.

In early times, many researchers have recommended a healthy diet for diabetes. There are 97 articles which were published in IEEE Xplore digital library, 108 articles published in Science direct and 162 articles from Google Scholar. Among all the articles and journals the most cited papers is from Pawlak Research Unit (Pawlak 2017). In order to maintain diabetes, a person should follow and maintain body weight goals by focusing on Healthcare, blood glucose levels, Insulin, blood pressure, glycemic index and lipid levels. Delay or prevent complications of diabetes healthy diet, taken as the prevention of any medical treatment (Association and AMERICAN DIABETES ASSOCIATION 1987).

MATERIALS AND METHODS

The research was carried out in the Open Source lab, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. In order to add value to the diet choices for diabetes patients, the study looked at two types of classifiers: Bootstrap Ensembling and KNN algorithms. The Bootstrap

Ensembling algorithm is in group 1 with a sample size of ten, while the K-nearest neighbor (KNN) algorithm is in group 2 with a sample size of ten, and the two algorithms are compared for higher accuracy and precision score values in order to choose the best algorithm. Clinical.com was used to create the pre-test analysis, which had a G power of 83 %, a threshold of 0.05 percent, a confidence interval of 90 %, and a mean and standard deviation of 90 %. This dataset was taken from the kaggle open source website. The Bootstrap Ensembling algorithm was chosen for implementation in this study, and it was compared to the KNN algorithm.

Bootstrap Ensembling :

The bootstrap is an effective statistical approach for estimating an amount from an information sample. This is simplest to recognize if the amount is a descriptive statistic together witha median or a general deviation.

Algorithm :

Ensemble Bootstrap Pseudocode

Parameters: learning-rates: $\{\alpha t\}_{t\geq 1}$

Initialize: Q-ensemble of size W : $\{ T^{i}\}^{K}_{i=1}$, s_0

for t = 0, . . . , T do

Choose action at = argmax_i [P W i=1 Ti (st, a)]

i at = explore(a_t)

s_{t+1}, rt ← env.step(s_t, a_t)

Sample an ensemble member to update: k_t ∼U([W])

$a\hat{} * = \text{argmax}_a\, Q^{kt}$ (st+1, a)

T_{kt} (st, at) ← $(1 - \alpha t)T^{kt}$ (st, at)

$+\alpha t\ (rt + \gamma T^{EN\backslash kt}$ (st+1, $a\hat{} *$)

end for

Return $\{Wi\}^{K}_{i=1}$

Steps to perform the analysis of Bootstrap algorithm :

- Import the dataset
- Explore the data and analyze dataset how it looks
- Pre-process the data
- Split the data into attributes and labels
- Divide the data into training and testing sets
- Train the Boot strap Ensembling algorithm
- Make some recommendations
- Evaluate the results of the algorithm

KNN :

K-nearest neighbor classifier is one of the introductory supervised classifiers. KNN copes with the samplepopularityissues and additionally the pleasantpicks for addressing a number of the classassociated tasks. This model is to expect the goal label with aid of locatingthe closest neighbor elegance. The closest elegancecan bediagnosed by the use of space measures like Euclidean distance.

Pseudocode :

Calculate "z(y, y_i)" i =1, 2, ….., n; where z denotes the Euclidean distance between the points.

Arrange the calculated n Euclidean distances in non-decreasing order.

Let q be a +ve integer, take first q distances from this sorted list.

Find those q-points corresponding to these q-distances.

Let q_i denotes the number of points belonging to the i^{th} class among q points i.e. $q \geq 0$

If $q_i > q_j\ \forall\ i \neq j$ then put y in class i.

Traditional Mediterranean healthy diet is primarily based totally on:

• High consumption of vegetables (< 250 g/die).
• Fruits/nuts.
• Cereals (wealthy in fiber).
• Animal sources of proteins (fish, meat, eggs).
• Olive oil is an important fat source .
• Low to mild milk and dairy products.
• Low alcohol consumption (Pitocco et al. 2012).

Statistical Analysis :

The data collected in this study was analyzed using IBM SPSS version 21 statistical software. For both proposed and current algorithms, 10 iterations with a maximum of 20 samples were performed, with the anticipated accuracy noted for each iteration for accuracy analysis. The T-test was performed using the value obtained from the Independent Sample iterations. Diet is one of the dependent factors. Carbohydrates, iron, vitamins, and glucose are the independent factors (Pawlak 2017). These numbers have been subjected to a comprehensive review in order to make food recommendations for diabetics.

RESULTS

Table 1 represents mean and standard deviation of the 2 algorithms. Table 1 shows the findings of the group statistics on all variables, because it uses the ensemble approach for recommending diet selection. Bootstrap ensembling algorithm acquires best accuracy and standard deviation when compared to KNN. As a result of these transformations, it obtains the best boundary between the viable outcomes. Because of the relevance of equality of variance, the probability value states that the results in the research effort are significant and correlated with each other (Pawlak 2017). The table demonstrates the difference in accuracy of both Bootstrap Ensembling and KNN. The accuracy comparison of the Bootstrap Ensembling and KNN algorithm is shown in the graph (fig. 1). The results of the independent sample t-test are shown in Table 2. Because of its efficient classification feature based on the Ensembling technique , the algorithm outperforms KNN. The outcome of the study shows Ensembling AdaBoost with 70.02 % higher accuracy than KNN 63.2%.

DISCUSSIONS

The American Dietetic Association states that deliberate vegetarian diets, such asgeneral vegetarian or vegan diets are healthy, nutritionally adequate, and mightofferfitness prevention and remedy of positive diseases. The body condition of Type 2 diabetic patients does not work properly and it leads to impairment (disability). Recommendation of vegetarian food plansuch as protein, v-3 fatty acids, iron, zinc, iodine, calcium, and nutrients D and B12 (Mangels, Messina, and Messina 2011) might prevent from Type 2 diabetes. The American Diabetes Association states inside the 2014 Clinical Practice Guidelines that plant-primarily based total diets enhance metabolic manipulation in topics with diabetes (Davidson and Hsia 2017) This research paper recommends the proper diet selection for diabetic patients and what should be taken based on food groups. Also, it was proved that Bootstrap Ensembling outperforms existing KNN

algorithm with an accuracy of 70.02% , whereas existing KNN considers (60%) accuracy .

Vegetarian diets providenutritional interventions in each prevention and remedy of type 2 diabetes. According to authenticguidelines, transition to a vegetarian food planmust be supervised with the aid of a certifieddoctor and a skillful registered dietician. Proper and careful consideration of vegetarian diets are nutritionally adequate, powerful for weight and glycemic index manipulation, bestow metabolic and cardiovascular blessings, and decrease diabetes complications (Mangels, Messina, and Messina 2011). Larger medical trials are needed toaffirm the effectiveness of vegetarian diets and to tell their use in nutritionalpointers for the prevention and remedy of type 2 diabetes.

A Case Fatality rate (CFR) is the original technique of recommender systems, and suggests items to the active user that other users with similar preferences as mentioned in the earlier studies (Ricci, Rokach, and Shapira 2015). The limitations of the study is increased risk of heart disease, dehydration.In a recent study by Mika, they focused on challenges for nutrition recommender systems. Food recommender systems can be considered as two following types: The systems that recommend recipes for healthier meals, and the systems that suggest healthier food items. By following a proper dietetic plan we can maintain a healthy body and we can prevent heart diseases.The first type itself is divided into two categories. The diet plan suggests which type of food the person likes.

CONCLUSION:

The paper is mainly focused on a food recommendation approach and calculating daily personalized meal plans for the diabetic patients, according to their nutritional diet and food preferences. The recommended system identifies which type of food to be given to which patient based on the infection and different provisions like age, sexual orientation, weight, calories, protein, fat, sodium, fiber, cholesterol. Diabetes recommendation system using healthy diet by comparing Ensembling AdaBoost was successfully implemented. It can be slightly improved based on the random data sets analysis in future. The outcome of the study shows Ensembling AdaBoost with 70.02 % higher accuracy than KNN 62.3%.

DECLARATIONS:

Conflict of Interests

No conflict of interest

Authors Contribution

Author NNC was involved in data collection, data analysis, manuscript writing. Author DV was involved in the Action process, Data verification and validation, and Critical review of manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this research work successfully.

Funding: We thank the following organization for providing financial support that enabled us to complete the study.

1. Kaashiv infotech, Anna nagar, Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Association, American Diabetes, and AMERICAN DIABETES ASSOCIATION. 1987. "Nutritional Recommendations and Principles for Individuals With Diabetes Mellitus: 1986." *Diabetes Care*. https://doi.org/10.2337/diacare.10.1.126.
2. Batra, Usha, Nihar Ranjan Roy, and Brajendra Panda. 2020. *Data Science and Analytics: 5th International Conference on Recent Developments in Science, Engineering and Technology, REDSET 2019, Gurugram, India, November 15–16, 2019, Revised Selected Papers, Part II*. Springer Nature.
3. Davidson, Mayer B., and Stanley H. Hsia. 2017. *Meeting the American Diabetes Association Standards of Care*. American Diabetes Association.
4. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. "Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine." *Energy & Fuels: An American Chemical Society Journal* 35 (12): 9930–36.
5. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. "Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea." *Journal of Nanomaterials* 2021 (July). https://doi.org/10.1155/2021/8115585.
6. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. "Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile." *Scientific Reports* 10 (1): 18179.
7. Lee, Ji Jeung. 2014. "How to Cook Healthy Diet for People with Diabetes." *The Journal of Korean Diabetes*. https://doi.org/10.4093/jkd.2014.15.3.167.
8. Malaeb, Samar, Caitlin Bakker, Lisa S. Chow, and Anne E. Bantle. 2019. "High-Protein Diets for Treatment of Type 2 Diabetes Mellitus: A Systematic Review." *Advances in Nutrition*. https://doi.org/10.1093/advances/nmz002.
9. Mangels, Reed, Virginia Messina, and Mark Messina. 2011. *The Dietitian's Guide to Vegetarian Diets*. Jones & Bartlett Learning.
10. Nandhini, Joseph T., Devaraj Ezhilarasan, and Shanmugam Rajeshkumar. 2020. "An Ecofriendly Synthesized Gold Nanoparticles Induces Cytotoxicity via Apoptosis in HepG2 Cells." *Environmental Toxicology*, August. https://doi.org/10.1002/tox.23007.
11. Parakh, Mayank K., Shriraam Ulaganambi, Nisha Ashifa, Reshma Premkumar, and Amit L. Jain. 2020. "Oral Potentially Malignant Disorders: Clinical Diagnosis and Current Screening Aids: A Narrative Review." *European Journal of Cancer Prevention: The Official Journal of the*

European Cancer Prevention Organisation 29 (1): 65–72.

12. Pawlak, Roman. 2017. "Vegetarian Diets in the Prevention and Management of Diabetes and Its Complications." *Diabetes Spectrum*. https://doi.org/10.2337/ds16-0057.
13. Perumal, Karthikeyan, Joseph Antony, and Subagunasekar Muthuramalingam. 2021. "Heavy Metal Pollutants and Their Spatial Distribution in Surface Sediments from Thondi Coast, Palk Bay, South India." *Environmental Sciences Europe* 33 (1). https://doi.org/10.1186/s12302-021-00501-2.
14. Pham, Quoc Hoa, Supat Chupradit, Gunawan Widjaja, Muataz S. Alhassan, Rustem Magizov, Yasser Fakri Mustafa, Aravindhan Surendar, Amirzhan Kassenov, Zeinab Arzehgar, and Wanich Suksatan. 2021. "The Effects of Ni or Nb Additions on the Relaxation Behavior of Zr55Cu35Al10 Metallic Glass." *Materials Today Communications* 29 (December): 102909.
15. Phanich, Maiyaporn, Phathrajarin Pholkul, and Suphakant Phimoltares. 2010. "Food Recommendation System Using Clustering Analysis for Diabetic Patients." *2010 International Conference on Information Science and Applications*. https://doi.org/10.1109/icisa.2010.5480416.
16. Pitocco, Dario, Francesca Martini, Francesco Zaccardi, and Giovanni Ghirlanda. 2012. "Antioxidants, Healthy Diet, and Diabetes." *Nutritional and Therapeutic Interventions for Diabetes and Metabolic Syndrome*. https://doi.org/10.1016/b978-0-12-385083-6.00024-3.
17. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). https://doi.org/10.1002/ep.13696.
18. Tesfaye Jule, Leta, Krishnaraj Ramaswamy, Nagaraj Nagaprasad, Vigneshwaran Shanmugam, and Venkataraman Vignesh. 2021. "Design and Analysis of Serial Drilled Hole in Composite Material." *Materials Today: Proceedings* 45 (January): 5759–63.
19. Uganya, G., Radhika, and N. Vijayaraj. 2021. "A Survey on Internet of Things: Applications, Recent Issues, Attacks, and Security Mechanisms." *Journal of Circuits Systems and Computers* 30 (05): 2130006.

TABLES AND FIGURES:

Table 1. Group Statistics of Bootstrap Ensembling with KNN by grouping the iterations with Sample size 10, Mean =70.146 , Standard Derivation = .12340 , Standard Error Mean = 0.03902. Descriptive Independent Sample Test of Accuracy and Precision is applied for the dataset in SPSS. Here it specifies Equal variances with and without assuming a T-Test Score of two groups with each sample size of 10.

	Group	N	Mean	**Std.Deviation**	**Std.Error Mean**

Accuracy	Bootstrap Ensembling	10	70.1460	.52507	.16604
	KNN	10	62.3450	1.34034	.42385

Independent Samples Test

Table 2. Independent Sample Test of Accuracy and Precision (Calculate P-value = 0.001 and Significant value= .087, Mean Difference= 13.500 and confidence interval = (10.64373-9.74166). Bootstrap Ensembling and KNN are significantly different from each other.

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	LOWER	UPPER
Accuracy	Equal variance assumed	10.015	.005	17.137	18	.000	7.80100	.45522	6.84463	8.75737
	Equal variances not assumed			17.137	11.699	.000	7.80100	.45522	6.806339	8.79567

CHART BUILDER

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=Algorithm MEANCI(Acuuracy, 95)[name="MEAN_Acuuracy"
    LOW="MEAN_Acuuracy_LOW"                    HIGH="MEAN_Acuuracy_HIGH"]
MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: Algorithm=col(source(s), name("Algorithm"), unit.category())
  DATA: MEAN_Acuuracy=col(source(s), name("MEAN_Acuuracy"))
  DATA: LOW=col(source(s), name("MEAN_Acuuracy_LOW"))
  DATA: HIGH=col(source(s), name("MEAN_Acuuracy_HIGH"))
  GUIDE: axis(dim(1), label("Algorithm"))
  GUIDE: axis(dim(2), label("Mean Acuuracy"))
  GUIDE: text.title(label("Simple Bar Mean of Acuuracy by Algorithm"))
  GUIDE: text.footnote(label("Error Bars: 95% CI"))
  SCALE: cat(dim(1), include("1", "2"))
  SCALE: linear(dim(2), include(0))
```

```
  ELEMENT: interval(position(Algorithm*MEAN_Acuuracy), shape.interior(shape.square))
  ELEMENT: interval(position(region.spread.range(Algorithm*(LOW+HIGH))),
    shape.interior(shape.ibeam))
END GPL.
```

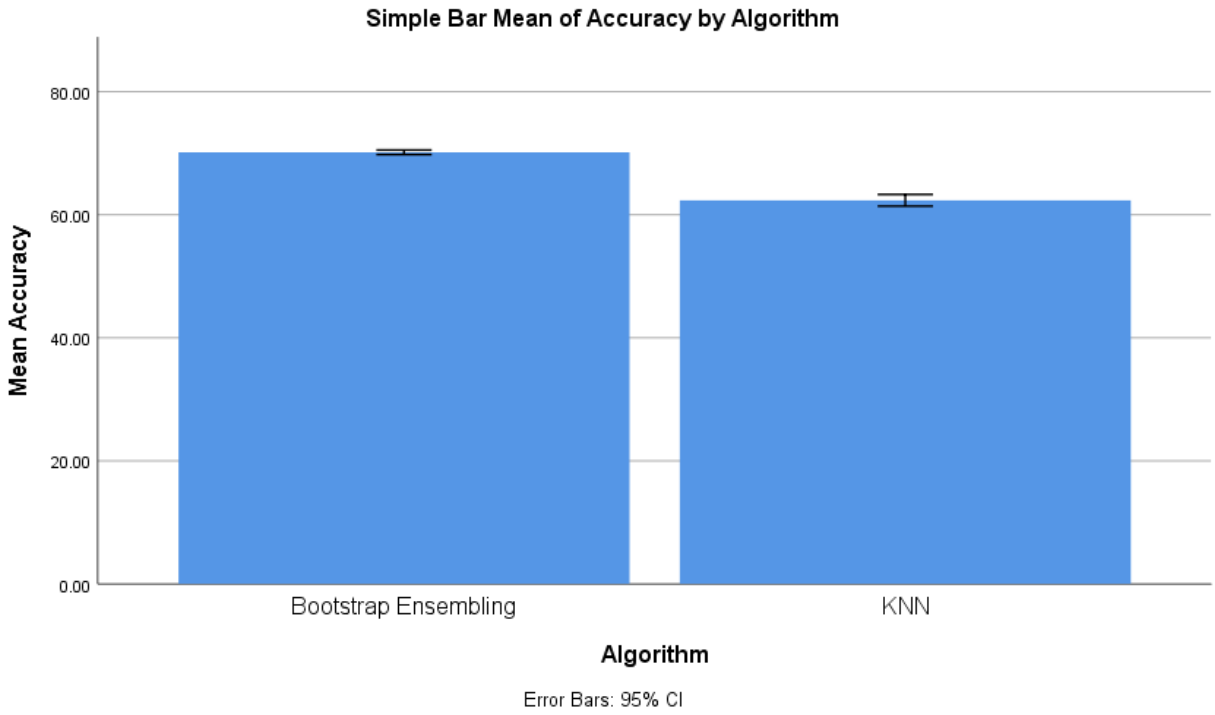

FIG-1 - Mean Accuracy of two algorithms