

# Fuzzy Bat based Cluster Center Selection Algorithm (FBCCSA) Improved K-Means Algorithm for Type 2 Diabetes Mellitus Prediction

P.V. SankarGanesh, P. SriPriya

**Abstract:** A present days lot of people are preconception by diabetic mellitus because of its constantly increasing happening day to day life. The all the diabetic patients recognize care about health quality or risk aspect before identify the diagnosis. Gestational diabetes are affected Women's during pregnancy period. Earlier stage identified to solve the problems. Modern days data mining techniques are introduced for type 2 diabetes mellitus. This proposed techniques to solve the problem and have accuracy of the prediction model get used for to more dataset. The newly designed K-means algorithm is predicted resolve the problem. Fuzzy bat algorithm also known as Fuzzy Bat based Cluster Center Selection Algorithm. Initially center values is produced randomly and directly affect the result of clusters. The proposed Fuzzy bat algorithm center values are selected by measuring the difference between the diabetes samples and the within cluster sum of squared errors. If the distance is smaller than the clusters are formed as well as center are also selected from diabetes samples. An Improved K-means Algorithm are to be combined to FBCCSA. This Proposed Paper improved k means algorithm with Logistic Regression (LR) algorithm for classification. A Sample Dataset were Collected to determine the proposed technique, which were taken from the University of California Irvine (UCI) for Analysis. LR model has provides higher accuracy of prediction than those of other methods such as Hybrid prediction model (HPM) and Logistic Regression (LR). The proposed model guarantees that the quality of dataset is adequate and results show good performance.

As a result, the model is shown to be useful for the realistic health administration of diabetes.

**INDEX TERMS:** Data mining, Diabetes mellitus, Clustering, Fuzzy Bat Algorithm (FBA), Logistic Regression (LR), University of California Irvine (UCI), Pima Indians Diabetes Dataset,

## I. INTRODUCTION

Diabetes mellitus every now and again gradually began to as diabetes, is the most significant worldwide medical issue that impacts a great many individuals in the world. Type 2 diabetes mellitus is an incessant illness where the body's glucose digestion is diminished and glucose levels are raised..

It is especially seen to be impacted by Lifestyle exercises. For example, a Drinking, work out, dietary propensities and others. T2DM alongside other ceaseless sicknesses, for example, hypertension, heftiness, dyslipidemia, and angina influences personal satisfaction anticipation.

The present moment and long haul antagonistic impacts related to T2DM in patients with cardiovascular hazards are notable. Subsequently, early finding and anticipation of T2DM are imperative to forestalling various genuine and conceivably perilous entanglements in patients with cardiovascular hazard. Late examinations have demonstrated that improving the way of life and medicine intercessions can forestall diabetic confusions, and it perhaps can forestall to the beginning of T2DM

Gestational diabetes early sign of type 2 diabetes rises in explicit women who can't overcome the insulin obstacle that makes during pregnancy. In any case, women who have gestational diabetes have extended threats of making Type 2 diabetes further not far off. Ladies are impacted during Pregnancy Time. Perceive the past Stage.

So as to inquire about the high-hazard gathering of DM, we need to use propelled data innovation. Therefore, the innovation of information mining is a reasonable preparation field for us. Information mining, in addition, notable as Knowledge Discovery in Databases (KDD), is well-characterized as the computational procedure of finding new examples in enormous datasets including strategies at the crossing point of man-made consciousness, AI, measurable information and framework database. The center utilization of these methodologies is design acknowledgment, forecast, affiliation, and bunching. Information mining involves a succession of steps at risk consequently or semi-naturally so as to mine and find invigorating, obscure, concealed highlights from enormous amounts of information. The high caliber of information and the appropriately applied strategy are two huge parts of information mining. Be that as it may, prescient models dependent on the bunching and arrangements are effectively utilized and have been built up in late work.

AI can help individuals to settle on an essential choice about diabetes mellitus presenting to their day by day physical assessment information, and it can fill in as a kind of perspective for specialists for AI process, the best approach to pick the legitimate geologies and the right classifier are the most significant confusions. As of late, different calculations are broke down and applied to anticipate diabetes, which likewise incorporates the conventional AI technique, for example, strategic relapse For AI strategy, how to choose the perform characterization productively turns into the most significant issues.

**Revised Manuscript Received on December 5, 2019.**

P.V. SankarGanesh, Research Scholar, VISTAS Pallavaram, Chennai-117, India, [sankarphd2017@gmail.com](mailto:sankarphd2017@gmail.com)

P. SriPriya, Professor and Department of Computer applications, VISTAS Pallavaram, Chennai-117, India, [sripriya.phd@gmail.com](mailto:sripriya.phd@gmail.com)

Right now, System utilized K-implies calculation for disposing of the undesired information. Diminish the handling time. Data grouping is a solo arrangement strategy. Information bunching is an unaided characterization technique targets making gatherings of articles, or groups so that items in a similar group are fundamentally the same as and questions in various groups are very unmistakable. Group investigation is one of the novel customary themes in the information mining field. It is an essential advance toward energizing information disclosure. Bunching is the methodology of collection information objects into a lot of disjoint classes, called groups. Presently questions inside a class have a high likeness to one another meanwhile protests in independent classes are increasingly not at all like. It is notable that the choice of starting group habitats can altogether influence K-implies' bunching results. The motivation behind this work is to propose an improved, proficient technique to deal with such a test..

As per the information provided by the inner-class distance among samples within the alike cluster is made-up to be smaller than the inter-class distance among clusters, the algorithm will dynamically adjust initial cluster centers that are randomly selected. Subsequently, such adjusted early cluster centers will be highly representative in the sense that they are dispersed among as many illustrations as possible.

## II. LITERATURE REVIEW

As of late, utilizing the information mining strategy alongside expanding recurrence to foresee the plausibility of infection. Numerous calculations and toolbox have been made and considered by specialists. These have featured the gigantic capability of this exploration field. Right now, scarcely any significant works that are firmly identified with the proposed issue are exhibited. (Han & Rodriguez, 2008)arranged Rapid-I's Rapid Miner to decide choice tree-based diabetes forecast models from a Pima Indians Diabetes Data Set, which gathers the data of patients with and without creating diabetes. Coming about because of the information mining strategy, the trade will accentuation on the information preprocessing, containing property recognizable proof and determination, exception expulsion, information standardization, and numerical discretization, visual information investigation, shrouded connections disclosure, and a diabetes forecast model development.

(Patil & Joshi, 2010)suggested a Hybrid Prediction Model (HPM) which plays out a Simple K-implies grouping calculation ordained at ensuring chosen class names of given information (mistakenly arranged occasions are expelled, for example, design mined from unique information) and subsequently applying the arrangement calculation to the outcome set. C4.5 calculations are used to shape the last classifier model by methods for the k-overlap cross-approval technique. The precisions achieved have been in the range between 59.4–84.05%. In any case, the proposed HPM achieved an arrangement precision of 92.38%. (Ahmad, Mustapha, Zahadi, & Masah, 2011) foreseen an expectation precision of multilayer perceptron in neural systems in the logical inconsistency of tree-based calculations, specifically, the ID3 and J48 calculations on Pima Indian diabetes mellitus informational collection. The characterization analysis is practiced utilizing calculations in WEKA to decide the class diabetes or non-diabetes with the informational collection of

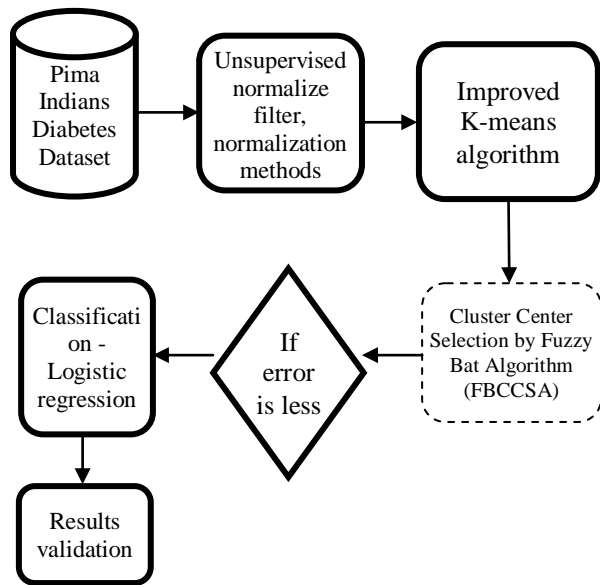
768 patients. Results indicated that a pruned J48 tree performed with a higher precision of 89.3%.

(Songthung, 2016)proposed a grouping to mine a broad dataset accumulated from 12 medical clinics in Thailand during 2011-2012 with 22,094 records of the screened populace who are females age 15 years or more established. RapidMiner Studio 7.0 is proposed with Naive Bayes and Chi-squared Automatic Interaction Detector (CHAID) Decision Tree (DT) classifiers to anticipate high-chance people. Results demonstrate that the utilization of order presented right now of hand-processed scoring can improve the forecast exhibition with an expansion in inclusion. Overall, hardly any investigations of calculation assessment and model building up for DM expectations have been consummate by these related works. Notwithstanding, the forecast precision and information legitimacy were not sufficiently high for a practical application.

(Wu, Yang, Huang, & He, 2018)proposed a series of preprocessing procedures, the model is comprised of two parts, the improved K-means algorithm and the logistic regression algorithm. As an outcome, the model is presented to be valuable for the genuine health management of diabetes. Although the improved model is not so complicated, it attained well effect according to plenty of experiments. Wei et al [22] analyzed risk factors of T2DM based on the Frequent Pattern (FP)-growth and Apriori algorithms. Analyze the run time and recorded data, time and support degree. FP-growth algorithm has a higher efficiency in analyzing risk factors of T2DM.

## III. PROPOSED METHODOLOGY

Proposed algorithm introduces cluster center selection Fuzzy bat algorithm Proposed work introduces a new cluster center selection by Fuzzy Bat Algorithm (FBA) is known as FBCCSA. In the proposed FBA, center values are selected by measuring the difference between the diabetes samples and the within cluster sum of squared errors. If the distance is lesser than the clusters are formed as well as center are also selected from diabetes samples. Based on this improved K-means algorithm is combined to FBCCSA. The proposed model is comprised of two parts, the improved K-means algorithm with FBCCSA and the Logistic Regression (LR) algorithm for classification. Pima Indians Diabetes Dataset collected from University of California Irvine (UCI) was utilized to analysis. Pima Indians Diabetes Dataset and MATLAB are used for implementation. Figure 1 demonstrations the overall architecture of the proposed system.



**Fig 1. Architecture of Proposed IKM- FBCCSA and LR system for diabetes prediction**

### 1.1. Dataset Description

Pima Indian Diabetes Dataset comprises of data on 768 patient datasets (268 tested\_positive cases and 500 tested\_negative examples) originating from a populace close to Phoenix, Arizona, USA. Tested\_ positive and tested\_negative entitles whether the patient is diabetic or not, correspondingly. Each example is involved 8 traits, which are for the most part numeric. This information contains individual wellbeing information just as results from therapeutic assessments. The properties in the dataset are recorded as follows,

- Number of times pregnant (preg)
- Plasma glucose fixation at 2 h in an oral glucose resilience test (plas)
- Diastolic pulse (pres)
- Triceps skin overlay thickness (skin)
- 2-h serum insulin (insu)
- Body mass list (bmi)
- Diabetes family work (pedi)
- Age (age)
- Class variable (class)

### 1.2. Data Preprocessing

The nature of the information, to an enormous degree, influences the aftereffect of the forecast. This implies information preprocessing assumes a significant job in the model (Orabi & Kamal, 2016). Right now some fitting techniques so as to enhance the first dataset..

At first, we have assessed each characteristic's medicinal outcome and its relationship to DM. We verified that the quantity of pregnancies has little association with DM (Patil & Joshi, 2010). Thus, we have changed over this numeric property into an ostensible trait. The worth 0 indicates non-pregnant and 1 connotes pregnant. The trouble of the dataset was decreased by this procedure.

Next, there are hardly any absent and inappropriate qualities in the dataset because of blunders or deregulation. The majority of the erroneous test results were brought about by

these negligible qualities. For a delineation, in the first dataset, the estimations of diastolic circulatory strain and weight record couldn't be 0, which demonstrates that the genuine worth was absent. To diminish the impact of aimless qualities, we utilized the methods from the preparation information to supplant every single missing worth.

### Unsupervised Normalize Filter

The unaided standardize channel for credit was utilized to standardize all the information into the segment [0, 1] by utilizing (1), where  $x'$  is the mean or normal incentive for the variable and  $s$  is the standard deviation for the variable. Worth is the new standardized worth.

$$Value = \frac{Value - x'}{s} \quad (1)$$

### Min-Max Normalization

Normalization of Min-Max method which is very easy to determine by the system which provided the pre-defined boundary (Yadav & Bharadwaj, 2012), (Patro, 2015). As per Min-Max were determined by normalization method

$$N = \left( \frac{A - \min_A}{\max_A - \min_A} \right) * (D - C) + C \quad (2)$$

Let N be the Min-Max Normalized information provided by the predefined by boundary in terms of [C,D]. Assume that A be the orginial data, due to that B may be used as a Mapped data

### Z Score Normalization

Z-score Normalization strategy is given the standardized qualities or scope of information from the first unstructured information through the ideas, for example, mean and standard deviation (Patro, 2015). So the unstructured information can be standardized by means of z-score parameter, according to determined equation

$$v'_i = \left( \frac{v_i - \bar{E}}{\text{std}(E)} \right) \quad (3)$$

### 1.2 Data Classification

The model comprises of twofold level calculations. In the primary level, improved K-implies calculation with FBCCSA is proposed to evacuate mistakenly grouped information. The advanced dataset was used as a contribution for the next level. At that point, the calculated relapse calculation is proposed to group the rest of the information.

#### i) Improved K-Means Algorithm With FBCCSA

i) Improved K-implies calculation with FBCCSA  
Group examination targets apportioning the information perceptions into divergent bunches so perceptions inside the indistinguishable bunch are more determinedly identified with one another than those doled out to various bunches (H, gan, ma, & wu, 2008). The K-implies is one of the most well-known group calculations. It is a normal separation based group calculation, and the separation is utilized as a proportion of likeness, i.e., the littler separation between objects shows the more prominent similitude.

1) Show all items. Select K is 2 from gave N as the quantity of beginning bunch place.

2) Compute distance among each object and cluster midpoint. Cluster every object to the nearest cluster according to the distance using (4)

$$S_i^{(t)} = \left\{ \forall j, 1A_j A_k x_p : \left| x_p - m_i^{(t)} \right|^2 \leq \left| x_p - m_j^{(t)} \right|^2 \forall j, 1 \leq j \leq k \right\} \forall j, 1A_j A_k \quad (4)$$

3) Recalculate every cluster center to verify whether they are changed using (5)

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (5)$$

4) Repeat the process until the new cluster center are same as the original data, still the step 2 and step 3 should be processed.

**Algorithm Cluster Center Selection by Fuzzy Bat Algorithm (FBCCSA)**

**Input:**  $x_i = (x_1, \dots, x_i)^t \rightarrow$  Bat population (diabetes dataset)

**Output:**  $x^{best}$  optimal CCS value and min  $(f(x)) \rightarrow$  optimal solution

- (1) Population initialization  
 $x_i = quasi\_Bat\_Init(x_i)$
- (2) Evaluate fitness as SSE by equation (7)
- (3) Compute  $(x^{best})$  and find current global best
- (4) While  $(Iter \leq Iter_{max})$ 
  - 4.1. For  $i = 1, \dots, N_{pop}$  do  
Update frequency by equation (8)  
Update velocity by equation (9)  
Update position by equation (10)
  - 4.2. If  $rand(0,1) > r_i^{(t)}$  then  
Update the new CCS value solution  
End if
  - 4.3. If  $f_{new} = Compute\ x_j^t$
  - 4.4. Evaluate = Evaluate + 1
  - 4.5. If  $f_{min}^{new} < f_{old} \& rand(0,1) < A_i^{(t)}$  then  
 $x_i = x_j^t, f_{old} = f_{min}^{new}$
  - End if
  - 4.6.  $f_{min} = \min(f(x^{best}))$
  - 4.7. End for
- (5) End while

ii) **Logistic Regression (LR)**

The classification algorithm intended to launch a model that can map data items to a given category, based on the existing data. It was used to mine significant data items from the model or to predict the tendency of data. In most cases, the dependent variable of the logistic regression algorithm is binary-classification. It shows that the logistic regression algorithm is continuously utilized to resolve two-category problem.

The significant utilization of trial is to figure whether one individual is diabetic or not, which is an exemplary parallel order issue. Straight relapse issue has the option to just figure a consistent worth. It maintains solid affectability all over the genuine number field. The prescient estimation of the arrangement issue must be 0 or 1, so it may set a basic point. The yield is 1 if the worth is more prominent than the edge, in any case, the yield is 0. The yield variable scope of calculated relapse is consistently somewhere in the range of 0 and 1. Strategic relapse is a relapse model that decreases the forecast range and constrains the expectation incentive to [0, 1]. In light of direct relapse, the strategic relapse includes a layer of sigmoid capacity (non-linearity). The highlights are first added directly and afterward anticipated utilizing the sigmoid capacity. The fundamental equations of the calculated relapse in (17), (18), and (19).

$$\Pr(Y = +1|X) \beta \cdot x \& \Pr(Y = -1|X) = 1 - \Pr(Y = +1|X) \quad (6)$$

$$\sigma(X) := \frac{1}{1 + e^{-X}} \in [0,1] \quad (7)$$

$$\Pr(Y = +1|X) \sigma(\beta \cdot x) \& \Pr(Y = -1|X) = 1 - \Pr(Y = +1|X) \quad (8)$$

The Y indicates that the patient is diabetic. x independent variables represent the 8 attributes in the original dataset. Every dependent variable x is assigned a coefficient value called  $\beta$  representing the weight. After being analyzed by the logistic regression algorithm, the dataset showed every variable's value of weight. Different weights represent diverse correlation between x and Y. Once the regression model has been settled, it is efficient to input new data and predict whether the outcome is positive or negative. Set the logistic regression algorithm as the final step.

**IV. EXPERIMENTAL RESULTS**

Using the MATLAB toolkit, it was convenient for us to study the result of the experiment through a visualized interface. Analyzed and evaluated the IKM- FBCCSA (LR) model based on the following aspects. The results are measured between the proposed and existing methods (HPM, AMMLP model and IKM-LR). The results are measured in terms of the precision, recall, F-measure, accuracy, Mathews Correlation Coefficient (MCC) and Kappa statistic. The table 1 shows the original confusion matrix and table 2 shows the confusion matrix of the diabetes dataset.

**Table 1. Confusion Matrix**

		Detected	
		Positive	Negative
Actual	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative(TN)

**Table 2. Confusion matrix**

A	B	CLASSIFIED
377	7	Predicted Positive
20	185	Predicted Negative





Exactness is the extent of the appropriately characterized cases to the whole number of misclassified cases and effectively grouped cases. Recall the extent of accurately requested cases to all outnumber of sloppy cases and effectively grouped cases. In aggregation, we applied the F-measure to join the review and exactness which is viewed as a decent marker of the connection between them.

Exactness is the extent of the whole amount of expectations that were effectively determined.

**Precision (P):** Precision is characterized as the rates of anticipated class which has a place with the positive class that was right, as decided to utilize the condition

$$\text{Precision(P)} = \frac{TP}{TP+FP} \quad (9)$$

**Recall (R) :** is used to determine the observed class and predicted class which may be positive case may be identified are as follows using the equation(10):

$$\text{Sensitivity(Sen)} = \text{Recall(R)} = \frac{TP}{TP+FN} \quad (10)$$

**F-measure (F):** are determined the combination of precision and harmonic mean and calculate the traditional F-measure are as follows

$$\text{F-measure} = 2 \cdot \left( \frac{P \cdot R}{P+R} \right) \quad (11)$$

**Classification Accuracy (CA):** termed to determine the percentage of total amount of prediction calculate by both positive and negative were identified. (Songthung, 2016):

$$\text{Classification accuracy (CA)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

Where True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

The Mathews Correlation Coefficient (MCC) is used as a measure of the quality of binary classifications, calculated by (24).

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \quad (13)$$

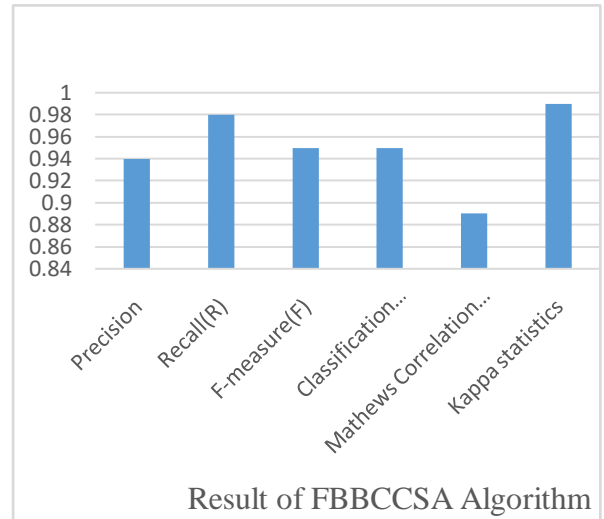
### Kappa statistic

A statistical Measure of Kappa which determined the parameter which is more significant to calculate the consistency of the model. Its mainly focused on the associates the outcomes of the model with outcomes produced by arbitrary between zero and one. 1 represents the true value and where zero represents the false value

$$K = [P(A) - P(E)] / [1 - P(E)] \quad (14)$$

### Result of FBBCCSA Algorithm

<b>Precision</b>	0.94
<b>Recall(R)</b>	0.98
<b>F-measure(F)</b>	0.95
<b>Classification Accuracy(CA)</b>	0.95
<b>Mathews Correlation Coefficient(MCC)</b>	0.89
<b>Kappa statistics</b>	0.51
<b>Total Number of Instances</b>	768



## V. CONCLUSION AND FUTURE WORK

This work aimed to establish an appropriate prediction model for the high-risk T2DM group. Novel model algorithm is proposed this work which consists of double-level algorithms, i.e., the improved K-means with cluster center selection (CCS) and Logistic Regression (LR) algorithms. In the first level, improved K-means algorithm with FBBCCSA is proposed to remove incorrectly clustered data. It consists of two steps: clustering and CCS. In the CCS stage is that inserts a program to record and sort the value called 'Within cluster sum of squared errors' by ascend order with Fuzzy Bat Algorithm (FBA). In FBA, a new method of population initialization approach is performed using lower difference sequence which makes use of quasi-random sequence to create the initialization of the swarm by the data points in the diabetes samples. Next modification is performed by focusing on local search method. Fuzzy membership function is added to local and global value detection of CCS value. This Proposed work mainly focused on both cluster and a method of particular class which belongs to predict the accuracy and also proven to provide approximately to T2DM. These proposed work will eliminated the unwanted original data which ensures the high quality of predicted data.

## REFERENCES

1. Ahmad, A., Mustapha, A., Zahadi, E., & Masah, N. . (2011). Comparison between neural networks against decision tree in improving prediction accuracy for diabetes mellitus. In International Conference on Digital Information Processing , Communications, 537-545.
2. Alghamdi, M., Al-Mallah, M., Keteyian, S., Brawner, C., & Ehrman, J. . (2017). Predicting



**Dr.P. Sripriya**, Professor and Department of Computer applications, VISTAS Pallavaram, Chennai-117, India, sripriya.phd@gmail.com

- diabetes mellitus using SMOTE , ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project. PLoS one.
3. ansal, A., & Sharma, M. . (2017). Improved k-mean clustering algorithm for prediction analysis using classification technique in data mining. International Journal of Computer Applications,, 0975-0987.
  4. Guo, Y., Li, Y., & Wang, G. . (2014). Application of artificial neural network to predict individual risk of type 2 diabetes mellitus. Journal of Zhengzhou University , 180-183.
  5. H, D., gan, g., ma, c., & wu, j. (2008). Data clustering: Theory, algorithms, , applications . International Statistical Review, 141-143.
  6. Han, J., & Rodriguez, J. . (2008). Discovering decision tree based diabetes prediction model. In International Conference on Advanced Software Engineering , Its Applications, 99-109.
  7. Joshi, K. . (2013). Modified K-Means for better initial cluster centres. International Journal of Computer Science , Mobile Computing, IJCSMC, 219-223.
  8. Kaur, D. . (2013). Enhancement in the Performance of K-means Algorithm. International Journal of Computer Science , Communication Engineering, 29-32.
  9. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning , data mining methods in diabetes research. Computational , structural biotechnology journal, 104-116.
  10. Lee, B. J., & K. J. (2016). Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry , triglycerides based on machine learning. IEEE J. Biomedical Health Information, 39-46.
  11. Majhi, S. . (2018). Optimal cluster analysis using hybrid K-Means , Ant Lion Optimizer. Karbala International Journal of Modern Science, 347-360.
  12. Marcano-Cedeño, A., & Torres, J. . (2011). A prediction model to diabetes using artificial metaplasticity. International Work-Conference on the Interplay Between Natural , Artificial Computation, 418-425.
  13. Orabi, K., & Kamal, Y. . (2016). Early predictive system for diabetes mellitus disease. Industrial Conference on Data Mining, 420-427.
  14. Patil, B., & Joshi, R. . (2010). Hybrid prediction model for type-2 diabetic patients. Expert systems with applications, 8102-8108.
  15. Patro, S. . (2015). Normalization: A preprocessing stage. arXiv preprint arXiv.
  16. Pavate, A. . (2015). Risk prediction of disease complications in type 2 diabetes patients using soft computing techniques. Fifth International Conference on Advances in Computing , Communications, 371-375.
  17. Romero, S., Garcia-Egido, A., Escobar, M., Andrey, J., Corzo, R., Perez, V., et al. (2013). Impact of new-onset diabetes mellitus , glycemic control on the prognosis of heart failure patients: a propensity-matched study in the community. . International journal of cardiology, 1206-1216.
  18. S, Y. X. (2010). A new metaheuristic bat-inspired algorithm in Nature inspired cooperative strategies for optimization. Springer, Berlin.
  19. Song, J., & Li, X. . (2015). An Optimized k-means Algorithm for Selecting Initial Clustering Centers. International Journal of Security , Its Applications, 177-186.
  20. Songthung, P. . (2016). Improving type 2 diabetes mellitus risk prediction using classification. 13th International Joint Conference on Computer Science , Software Engineering , 1-6.
  21. Ta, S. (2014). Diagnosis and classification of diabetes mellitus. Diabetes care, S81-S90.
  22. Tang, R., Fong, S., & Yang, X. . (2012). Integrating nature-inspired optimization algorithms to K-means clustering. In Seventh International Conference on Digital Information Management (ICDIM 2012), 116-123.
  23. Wei, Z., & Guangjian, Y. . (2016). Analysis for risk factors of type 2 diabetes mellitus based on FP-growth algorithm. China Medical Equipment, 45-47.
  24. Wexler, D., Grant, R., Wittenberg, E., Bosch, J., Cagliero, E., Delahanty, L., et al. (2006). Correlates of health-related quality of life in type 2 diabetes. Diabetologia, 1489-1497.
  25. Wu, H., Yang, S., Huang, Z., & He, J. . (2018). Type 2 diabetes mellitus prediction model based on data mining. Informatics in Medicine Unlocked, 100-107.
  26. Xin-She, Y. (2011). Bat algorithm for multi-objective optimization. International Journal of Bio-Inspired Computation, 267-274.
  27. Yadav, S., & Bharadwaj, B. . (2012). Mining Education data to predict student's retention: a comparative study. International Journal of Computer Science , Information Security, 113-117.

## AUTHORS PROFILE



**P. V. Sankarganesh**, Research Scholar, VISTAS Pallavaram, Chennai-117, India, [sankarphd2017@gmail.com](mailto:sankarphd2017@gmail.com).

Retrieval Number: D10031284S519/2019©BEIESP  
DOI:10.35940/ijrte.D1003.1284S519

Published By:  
Blue Eyes Intelligence Engineering  
& Sciences Publication