# Synthetic Data Generation for Food Waste Pelletizer using TGANs, TVAEs

Pattan Afrid Ahmed
*SCOPE*
*Vellore Institute of Technology*
Chennai, India – 632014.
pattanafrid.ahmed2020@vitstudent.ac.in

Dr. Priyadarshini J
*SCOPE*
*Vellore Institute of Technology*
Chennai, India – 632014.
priyadarshini.j@vit.ac.in

***Abstract*** – **Prepared and raw food waste materials often establish challenges to the food makers in its dumping and prompt transportation. The proposed project aims at clean dumping of the food wastes with a significantly reduced space and its usage as cattle-feed. The proposed system consists of both hardware and software portions. The hardware offer to perform an integrated mechanical grinding, induction and compacting processes. The level of operation and other process parameters are controlled by fuzzy logic. Synthetic data is generated using TGANs, TVAEs which are unsupervised learning techniques and can be trained using less data. The compact system can be used for making cattle feeds from the bio-wastes at the user outlets.**

***Keywords:*** *grinding, induction, fuzzy logic.*

## I. INTRODUCTION

According to FAO reports, Every year, around 1.3 billion tonnes of food intended for human consumption is wasted or thrown, accounting for almost one-third of all food produced worldwide. Because Sub-Saharan Africa's net food production is so low, consumers in producing nations squander the most food (222 million tonnes) per year (230 million tonnes). Each year, the total amount of food lost or wasted corresponds to about half of the world's population annual cereals crop. (2.3 billion tonnes in 2009/2010).

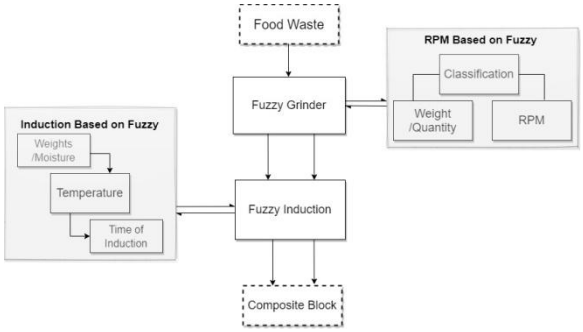
Automated food waste pelletizer goes through three different stages. First, the organic waste is thrown into the bin. After reaching threshold weight or height, it starts grinding the waste and converts it into paste. Then the paste is further processed to heating, where the moisture is evaporated and the paste is converted into a block or semi-solid form.

It mainly consists of three major components: collecting food waste, mechanical grinding, pelletizing with the help of induction heating.

*Collecting food waste*: Organic food waste like vegetable peels, banana leaves, meat, bones are collected or thrown into a pelletizer. It has a threshold limit of 5kg weight and 0.8m height. When it reaches the threshold it automatically moves to the next step.

*Mechanical grinding*: When the dustbin reaches the threshold limit then grinding the waste starts. We have segregated the dataset for different materials with different RPM. When it comes to variable speed, BLDC motor works fine with good efficiency and control using fuzzy logic. For BLDC motor fuzzy logic is applied to adopt the suitable RPM for different weights and densities of the material.

*Pelletizing with induction heating*: The grinded material obtained after mechanical grinding is in the form of paste or pulp. This material is then heated using the Electromagnetic Induction (EMI). The time for induction heating is derived from the moisture content in the paste. Using fuzzy logic, we achieve the temperature and the time for the induction heating.




Fig(1). System Architecture

*Synthetic data generation:* It is the process of generating artificial data, rather than from real-world experiments. Synthetic data is generated algorithmically, and approximating original data to production or operational data, statistical model validation, and, increasingly, machine learning model training or simulation model validation. This helps in keeping privacy of sensitive data like health care, bank, financial data, etc., or to reduce the time and cost of building data for experiments.

Generative algorithms are used for synthetic data generation like GANs(Generative Adversarial Networks), VAEs(Variational Auto-encoders). We'll be discussing in detail how these work, basic architecture in this following report. Initially, we prepared 20 objects of experimental data that was generated by manual experiments. Using this data approximately 150 objects of synthetic data are generated which can further be used for simulation.

*Fuzzy Logic:* It is an approach for variable processes that allow multiple values to process through the same variable. In a broad sense, it can be best understood in the context of set membership functions which are defined by "high", "medium", "low" etc,. In a slim sense, a fuzzy logic system aims to formalize approximate reasoning that is thought as vagueness. We'll be using fuzzy logic to control the speed of the motor depending upon the weights and the threshold

height limit of the bin. After grinding the organic waste, a paste like form is obtained. We need to reduce this paste into the powder or semi-solid form by induction heating process. For induction heating, we'll apply fuzzy to control the input temperature such that the moisture content in the waste determines the time or temperature required to heat it resulting in evaporation of moisture content and can form as a powder or semi-solid.

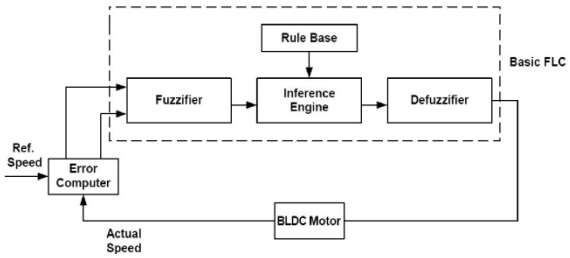
*Simulation:* Using python, fuzzy control system simulation on the generated data should be done. Fuzzy set and rules will be defined by trial and error method such that it should fire desired outputs. So that simulation with respect to speed and the temperature on a fuzzy system can be done.

## II. LITERATURE SURVEY

**[1]** The basic idea of the project by *Harshada Ithape*[2] and team designed a smart garbage monitoring system in which it will automatically give the information about the level in dustbins to the municipal corporation using a webpage. In this project they are separating the dry and wet garbage into two different dustbins using a DC motor. The garbage filled level in the dustbin will be sensed by ultrasonic sensors and upload the information on the webpage.

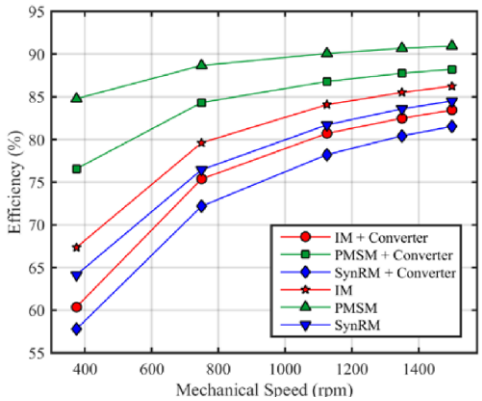
**[2]** The project proposed by *K.Harika*[4] and team is an e-monitoring system that puts forth an embedded system and web-based software assimilated with IOT technology. The filling level of the garbage bin in the dust bin and its original level height could be sensed monitored by the ultrasonic sensor. Programming in the Arduino UNO is done in such a way that once a particular level of filling is sensed, an information message is sent requesting to clean and update the status of the bin.

**[3]** The research Paper by R.K.Karambe, D.H.Gahane[6] reports the automation of grinder applying fuzzy logic. The rule base is fed with input values Material Type and Quantity and gives the output in terms of Range as a defuzzifier which fires the rule base. This research work will maximize the capacity of existing fuzzy logic control (FLC) systems in the process of automation of Grinder with potential benefits. In this paper, it depicts a fresh idea to control the grinding speed, so that it can perform at its optimum speed as well as save electricity. Two sensors are used, one is used to measure the Material type and second is used to detect Quantity of Material. The amplification of the sensors are connected with the two fuzzifiers of the fuzzy logic control system and defuzzified which ranges the control and give the output.




Fig(2). Flow chart of fuzzy on BLDC

**[4]** Research by *Fernando Bento, Jorge*[8] is the comparative analysis of the most significant AC electric drives, based on three distinct motor types: squirrel cage induction motor (IM), permanent magnet synchronous motor (PMSM), and synchronous reluctance motor (SynRM), to analyze the performance from the most popular drive systems available in industry.




Fig(3). Mechanical speed(rpm) vs efficiency(%)

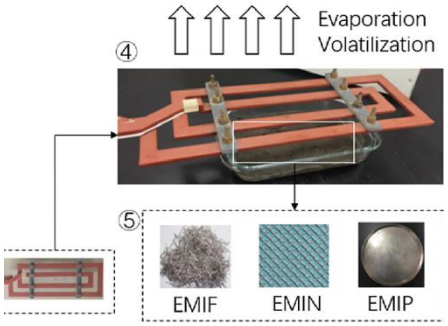
The above comparison in *fig.3* makes a general evaluation of the drives (PMSM, SynRM, IM) which are supported by their experimental results with parameters like motor or drive efficiency, voltage and current supply, stator temperature. The graph also shows the better performances with the use of power converters used in the control of motors. And Significant increase in efficiencies without relevant changes in the operational ranges.

**TABLE I-** *Performance analysis*

| | IM | PMSM | SynRM |
|---|---|---|---|
| **Power (kW)** | 2.2 | 2.2 | 2.2 |
| **Speed (rpm)** | 1435 | 1500 | 1500 |
| **Frequency (Hz)** | 50 | 75 | 50 |
| **Torque (Nm)** | 14.6 | 14 | 14 |
| **Voltage (V)** | 400 | 400 | 400 |
| **Current (A)** | 4.56 | 4.4 | 5.7 |
| **Efficiency (%)** | 87 | 91.6 | 89.5 |
| **Frame Size** | 100L | 90L | 100L |
| **Frame Material** | Cast Iron | Cast Iron | Aluminium |
| **Weight (kg)** | 33 | 17 | 25 |
| **Efficiency Class** | IE3 | IE4 | IE4 |

**[5]** In this paper by *Adil Usmanl* and *Bharat Singh Rajpurohit*[9] an attempt has been made to analyze the result of PI Controller, by using the appropriate $K_p$ and $K_j$ value in order to get the constant control on the speed of a Brush Less Direct Current (BLDC) motor. Further an attempt as an assessment has been made in designing an alternative controller to minimize steady state error and obtain better results for the control of the speed parameter of a BLDC Motor. A Fuzzy Logic Controller (FLC) has been designed to compare the PI Controller output with the FLC output.

**[6]** *Yongjie Xue, Chen Wang*[12] and team proposed thermal drying of sewage sludge is not only an effective waste treatment method, but it is also a necessary stage in the resource utilization process. The study employed electromagnetic induction heating to investigate the drying properties of sewage sludge. During the drying process, the temperature and heat transmission were measured. The combination of sewage sludge and three types of induction media can be efficiently dried by applying electromagnetic induction heating. Electromagnetic Induction Fiber (EMI-F), electro-magnetic Induction Plate (EMI-P), and electromagnetic Induction Net (EMI-N) are the three media materials identified. EMI-Fiber and EMI-N are two common materials that are defined as inner induction media that are blended and constructed with sludge. EMI-Plate is elucidated as an exterior induction medium that appears like a plate or container with a thin layer of sludge.




Fig(4). types of electromagnetic induction

**[7]** *Lei Xu*[16] and his team worked on neural networks using the *tanh* activation function, TGANs can successfully produce values with a distribution that is centered. It is determined that the probability distribution may well be constructed directly using *softmax* due to the low cardinality. However, category values must be converted to binary variables using a one-hot-encoding representation with noise. It transform **T** vector with **d** after prepossessing the discrete and continuous columns to **V**, **U**, **D** vectors. The LSTM with attention technique is used to generate the appropriate row. Random variable **z**, weighted context vector with previous hidden, and embedding vector are the inputs for each step of the LSTM.

$$a_t = \sum_{k=1}^{t} \frac{\exp \alpha_{t,k}}{\sum_j \exp \alpha_{t,j}} h_k. \qquad \text{(1)}$$

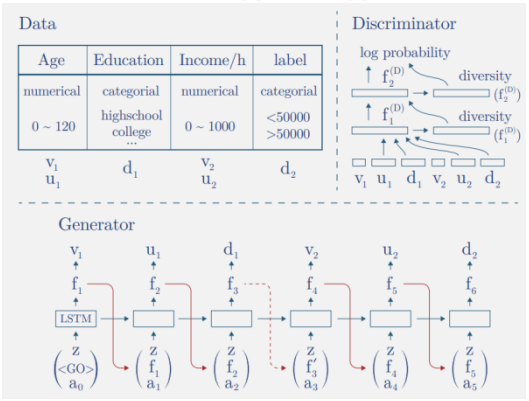
The Multi-Layer Perceptron (MLP) is implemented, alongwith *LeakyReLU* and *BatchNorm*. Concatenated vectors were applied in the first layer, coupled with mini-batch diversity using LSTM feature vectors. The loss function is the sum ordinal log loss function with the *KL divergence* term of input variables.

$$f_1^{(D)} = \text{LeakyReLU}(\text{BN}(W_1^{(D)}(v_{1:n_c} \oplus u_{1:n_c} \oplus \mathbf{d}_{1:n_d}))), \qquad (2)$$

$$f_i^{(D)} = \text{LeakyReLU}(\text{BN}(W_i^{(D)}(f_{i-1}^{(D)} \oplus \text{diversity}(f_{i-1}^{(D)})))), i = 2:l, \qquad (3)$$

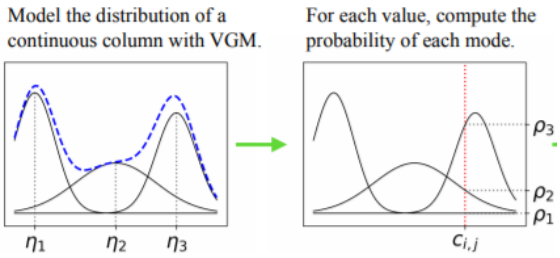
Loss function is defined to optimize the KL divergence of the generated data to the generator.

$$\mathcal{L}_G = -\mathbb{E}_{z\sim\mathcal{N}(0,1)} \log D(G(z)) + \sum_{i=1}^{n_c} \text{KL}(u_i', u_i) + \sum_{i=1}^{n_d} \text{KL}(\mathbf{d}_i', \mathbf{d}_i), \qquad (4)$$



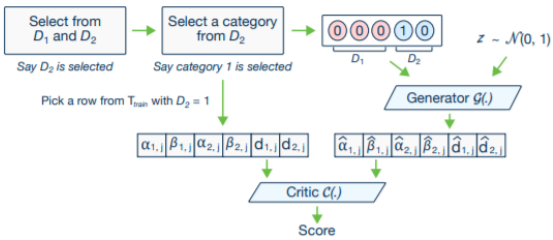

Fig(5). TGAN architecture

**[8]** *Lei Xu, Kalyan*[17] and team worked on TGAN and proposed Conditional TGAN to deal with the non-Gaussian and multimodal data. Variational Gaussian Mixture Models (VGMM) are used to generate continuous data in the tabular vectors. It first detects a number of modes in data and then fit GMM. After the vector is normalized with each mode, it works similar to TGAN.



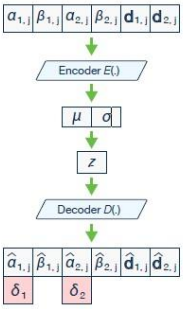

Fig(6). Mode specific normalization

Conditional vector, Generator loss, and training-by-sampling approaches are the three essential components of the approach. The model structure does not consist of LSTM, instead it is trained with WGAN loss with gradient penalty.




Fig(7). CTGAN model architecture

They also proposed a model Variational Auto-encoder, which outperformed CTGAN. The major key between TVAE and CTGAN is that, generator network does not have direct access to the real data distribution during training, unlike TVAE. It works by approximating the data distribution. While, TVAEs just take the original data distribution in the latent space and try to recreate or generate the synthetic data which is more close to the original data distribution.




Fig(8). TVAE model architecture

## III. ADVANTAGES

IoT Based garbage monitoring system helps a lot to attain the clean and sanitized pollution less surroundings and environment in developing smart cities. The automated dustbin monitoring device detects the full state of the garbage bins with sensors that indicate their status and issue alert commands. Thus eliminating the need of intermittent checks and overflowing of these garbage bins. This would help in keeping the environment as well as cities clean. This makes the garbage collection more effective, efficient and operative. The utility of the system is to help operate the Grinder at safer speed, so one can use Grinder without having knowledge of internal construction and technical knowledge of grinder. With this, we can increase the life of Grinder and save electricity. A Fuzzy Logic Controller (FLC) has been designed to compare the output of Proportional Integral Controller (PIC) with the FLC output. The results are then compared, interpreted, and comprehended. As a result, the FLC is found to be far more versatile than the normal-linear PIC.

By using CTGANs with VGM models along Wasserstein GAN loss penalty, we were able to generate data which closely approximated the original data. This data can be further used for simulation of fuzzy logic or fuzzy control system simulation.

From the above papers, we can deduce that PMSM has greater efficiency in AC motors. When compared to IM, PMSM has a slight advantage over weight, performance and efficiency. BLDC is similar to PMSM where the input current supply is in DC form. This BLDC has greater advantage over conventional DC because of low maintenance and better speed control.

The research work introduces electro-magnetic induction heating methods for sludge drying and a few of drying characteristics that are investigated. With the practical work, EMI (electro-magnet induction) different heating methods could be used for sludge drying. Specifically, 200 V of operating voltage for EMI-P small sample, the wetness content of 25 g of sludge reduced from 84% to 19.6% under 18 min with the rate of drying is much higher when compared to other media.

## IV. DISADVANTAGES

The IoT Based Garbage Monitoring System collects and distinguishes between dry and wet garbage, which is good but the wet garbage still poses a problem for workers who clean and also the one responsible for recycling it. As it would give out more stinking smell and might be more infectious, dangerous and make it difficult for the worker to keep it and transport it. So every time a truck comes it has to deal with different kinds of garbage which might vary in sizes and shapes.

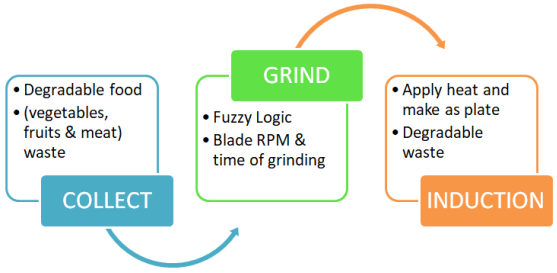
The Smart Garbage monitoring and clearing system indicates if bin is empty or full but it does nothing to process the garbage inside the Dustbin, and so every time a truck which comes it has to deal with different kinds of garbage which might be vary in sizes and shapes and make it difficult for the worker to keep it and transport it. Which smart bin created by our team will be overcome by grinding the garbage in small pieces so that it becomes really easy to handle and transport, also to remove the humidity and moisture our bin will apply appropriate induction by which handling becomes easier since it will become semi-solid or solid state. There is no proper overall automation or coordination for the system and it is entirely locally based, with no larger connectivity.

Using simple generative algorithms TGAN did not solve the problem as they were not able to generate proper data objects (row) with continuous data. Even CTGANs with LSTM layers are generating values which are out of the scope.

The entire system is only a simulation and not a real time application and there is no further process such as heating to make the waste viable. While fuzzy logic is used to determine speed of motor based on quantity and type of material, and to determine the time and the temperature that should be applied on the paste. Furthermore, MATLAB is more of a mathematical derivatives software and it is hard to handle real time data properly and efficiently.

## V. METHODOLOGY

The Smart Bin has grinding and induction feature which work with the required speed and the amount of heat through fuzzy logic and remove the moisture or wetness from the garbage and make it semi-solid or solid depending upon the garbage, which then will be much more easier to handle and process both for the cleaners and recycle workers.




Fig(9). step-by-step flow diagram

Fuzzification, an important Artificial Intelligence concept to calculate rotor speed and induction temperature over a large range of values using a set of rules. This ensures proper automation for the entire setup. Furthermore, the dustbin can send on the waste for grinding and heating at any point, not just when it gets full, and also doesn't depend on anything else to clear the bin when it reaches maximum capacity. Automated, integrated with a fuzzy logic based system for detecting types and weight of waste to then process it efficiently. Processing of garbage is being carried out more efficiently.

**TABLE II** –Sample rule base for fuzzy

| Category | Weight | Height | Speed |
|---|---|---|---|
| 1. | <=2 kg | 0.8 m | 300RPM |
| 2. | 2-4 kg | 0.8 m | 500RPM |
| 3. | 5=< | 0.8 m | 1000RPM |

BLDC motors has put back the usage of Brushed DC motors. PMSM have stumbled upon a more robust alternative to AC Induction motors. While a BLDC gives a trapezoidal wave and Permanent Magnet Synchronous Motor (PMSM) gives a sinusoidal wave. The complex sinusoidal waveform control calculations drove the advancement costs excessively high and requires a significantly more impressive processor than what's really needed for the generally straightforward trapezoidal control of a BLDC motor.

As discussed, a BLDC motor is easier to control the speed. Because of trapezoidal type Back-EMF, it's easy to define membership functions and rules which results in greater RPM control. Fuzzy logic controllers give better output when compared to conventional PI, PID's because of better response time. The inference may be concluded that the comparison was made with BLDC motor speed control using Fuzzy Logic Controller which has better and higher performance.

## VI. RESULTS

CTGANs are used to generate the synthetic data. Conditions are defined before training as the tabular data is relational and has low cardinality. Data training is performed using "reject sampling" method and the following aggregated score is obtained 0.552.

**TABLE III** –Scores of CTGANs

| metric | name | raw_score | normalized_score |
|---|---|---|---|
| LogisticDetection | LogisticRegression Detection | 2.482993e-01 | 0.248299 |
| SVCDetection | SVC Detection | 8.571429e-01 | 0.857143 |
| GMLogLikelihood | GaussianMixture Log Likelihood | -6.621698e+09 | 0.000000 |
| KSTest | Inverted Kolmogorov-Smirnov D statistic | 7.045455e-01 | 0.704545 |
| KSTestExtended | Inverted Kolmogorov-Smirnov D statistic | 7.045455e-01 | 0.704545 |
| ContinuousKLDivergence | Continuous Kullback–Leibler Divergence | 2.475332e-01 | 0.247533 |

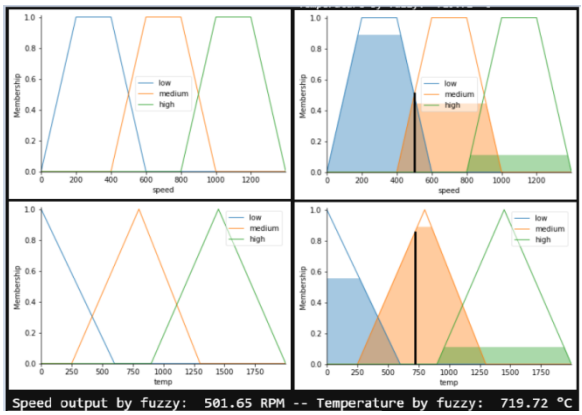
We also performed synthetic tabular data generation using VAEs and the reverberating algorithm CopulaGANs which is based on CDF (Cumulative Distribution Function). CopulaGANs are proposed to work well as the training of data is easy because CDF works on underlying CTGAN makes learning easy. But, it has generated data whose score are not any better than CTGANs, TVAEs. TVAE aggregated score is 0.669.

**TABLE IV** –Scores of TVAEs

| metric | name | raw_score | normalized_score |
|---|---|---|---|
| LogisticDetection | LogisticRegression Detection | 7.738095e-01 | 0.773810 |
| SVCDetection | SVC Detection | 8.690476e-01 | 0.869048 |
| GMLogLikelihood | GaussianMixture Log Likelihood | -2.871795e+09 | 0.000000 |
| KSTest | Inverted Kolmogorov-Smirnov D statistic | 7.348485e-01 | 0.734848 |
| KSTestExtended | Inverted Kolmogorov-Smirnov D statistic | 7.348485e-01 | 0.734848 |
| ContinuousKLDivergence | Continuous Kullback–Leibler Divergence | 2.382109e-01 | 0.238211 |

CTGANs are considered over TVAEs, if there is a privacy constraint on data. CTGANs are preferred while dealing with Healthcare, Financial data. But our primary goal is to generate data which is close to original. So we will be considering TVAEs synthetic data over CTGANs. Data is preprocessed and cleaned before simulation.

Lastly, simulation of a fuzzy control system using the generated dataset. Initially, defining the fuzzy set or membership functions, ranges and rules. Then input to the fuzzy is fed from the generated data and according to data, respective simulations are given as output. All these simulations are then compared to the original data and validated. As we have considered the temperature range from 0-2000$^{o}$C and speed range from 0-1400 rpm. For speed, trapezoidal membership function is considered because stabilized RPM outputs are expected and for temperature triangular membership function is considered as the output should be more precise. Below shown figures are the initial fuzzy graphs and the simulated graphs after giving input to it firing the desired output.




Fig(10). Simulation using Fuzzy Logic

From the above survey, we can deduce that EMI-P is most efficient when compared to others i.e,. EMI-N, EMI-F. Using python library Scikit-fuzzy, the input rules for different weights, heights and moisture which defines the speed required for grinding waste and induction temperature. Using this method of induction for research based smart dustbin, it will help in reduction of weight, volume of the organic waste which can be used further for transportation of cattle feed or agricultural uses.

## VII. CONCLUSION

Using the proposed BLDC motor and the EMIP induction which can be automated with the fuzzy. Generated data using over TVAEs is considered over CTGANs as the similarity distribution scores are more and the data is used for simulations with respect to speed and temperature. After the whole automated process of grinding, moisture reduction by induction heating and conversion to compost block can be achieved as an end product. This reduces the storage space, transportation cost, release of harmful gases and it can be fed to cattle or used for agriculture purposes.

## VIII. REFERENCES


[1] S Zavare, R Parashare , "Smart City waste management system using GSM" - International Journal Computer Science Trends Technology, 2017

[2] M. H. Thigale, Harshada Ithape, Shubhngi Lachyan, Ankita Mehare, "IoT Based Garbage Monitoring System", IJSRSET, Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 5 Issue 7, pp. 477-481, March-April 2019.

[3] Kumar, S. et al. "Smart garbage monitoring and clearance system using internet of things." 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials *(ICSTM)* (2017): 184-189.

[4] K.Harika, Muneerunnisa, V.Rajasekhar,P.Venkateswara Rao, L.J.N SreeLakshmi, IOT Based Smart Garbage Monitoring and Alert System Using Arduino UNO (IJIRCCE) February 2018.

[5] IoT Based Smart Garbage and Waste Collection Bin by S.S.Navghane, M S Killedar, Dr.V.M.Rohokale, (IJARECE) Volume 5, Issue 5, ISSN: 2278 – 909X May 2016.

[6] Automation of Grinder - An Introduction of Fuzzy Logic by R.K.Karambe, D.H.Gahane.

[7] Intelligent Waste Separator and Recycling using IoT and Evolutionary Fuzzy Logic Technology with energy efficient sensing during Hajj and Umrah crowd seasons by Salma Mahgoub Gaffer Elhag, ISSN (Online) 2393-8021, Vol. 6, Issue 2, February 2019.

[8] Bento, Fernando & Estima, Jorge & Cardoso, A.J.M.. (2017). Performance Comparison of Three Different AC Variable Speed Motor Drives.

[9] Usman, Adil & Singh, Bharat. (2016). Speed control of a BLDC Motor using Fuzzy Logic Controller. 1-6. 10.1109/ICPEICES.2016.7853304.

[10] Zanje, Tushar, Aniruddha Kadam, G. Vidhate, P. Yadav and G. Gurjar. "Speed Control of BLDC Motor using Fuzzy Logic*."* International journal of engineering research and technology(2018).

[11] Auliansyah, Fahrizal, Sutedjo, Ony Asrarul Qudsi and Indra Ferdiansyah. "Controlling Speed Of Brushless DC Motor By Using Fuzzy Logic Controller." 2020 International Seminar on Application for Technology of Information and Communication (iSemantic) (2020): 298-304.

[12] Yongjie Xue, Chen Wang, Zhenhua Hu, Yi Zhou, Gang Liu, Haobo Hou, Yue Xiao, Teng Wang, Jinping Li "Thermal treatment on sewage sludge by electromagnetic induction heating: Methodology and drying characterization, Waste Management", Volume 78, 2018, ISSN 0956-053X,

[13] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In International Conference on Machine Learning, (2017).

[14] Ramiro Camino, Christian Hammerschmidt, and Radu State. Generating multi-categorical samples with generative adversarial networks. In ICML workshop on Theoretical Foundations and Applications of Deep Generative Models, (2018).

[15] Edward Choi, Siddharth Biswal, Bradley Malin, Jon Duke, Walter F. Stewart, and Jimeng Sun. Generating multi-label discrete patient records using generative adversarial networks. In Machine Learning for Healthcare Conference. PMLR, 2017.

[16] Synthesizing Tabular Data using Generative Adversarial Networks - Lei Xu, Kalyan Veeramachaneni (2018)

[17] Modeling Tabular data using Conditional GAN - Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, Kalyan Veeramachaneni-arXiv:1907.00503 (2019).

[18] Synthesizing Tabular Data using Conditional GAN - Lei Xu, MIT (2020).