Precision Pig Farming Image Analysis Using Random Forest and Boruta Predictive Big Data Analysis Using Neural Network and K- Nearest Neighbor

S.A. Shaik Mazhar
Research Scholar
Department of Computer Science
Vels Institute of Science, Technology & Advanced Studies,
Chennai, India
shaik72ms@gmail.com

G. Suseendran
Department of Information Technology
Vels Institute of Science, Technology & Advanced Studies,
Chennai, India
suseendar_1234@yahoo.co.in

2021 2nd International Conference on Intelligent Engineering and Management (ICIEM) | 978-1-6654-1450-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICIEM51511.2021.9445328

Abstract: **Conditions and monitoring for production are significant issues in livestock accuracy Agriculture, in which image measurement and smart data collection are required. Dynamical surveillance and review of this Article Man are suggested as a device for scientific identification and growth evaluation of pigs. The Watershed enhanced algorithm is adapted to each human animal's section in chronic occlusion, depending on the depth of the photos captured during flight Camera in the chosen area of interest. For swine's weight, the rate of development is calculated from the image-based calculations and predicted using a segmented linear fitting form. Related results will then be used to interpret and explain incidents. As real-time feedback to the farmers, it happens in the pig hen. Preliminary studies have demonstrated a high potential for precision farming methods for livestock farming to increase efficiency and animal health. In this paper, Machine Learning is used in IMAGE analysis using Random forest and boruta with Predictive Big Data analysis on the pig farming data using the neural network and k- nearest neighbor algorithm for advanced predictive data analysis of our pig farming agriculture.**

Keyword: Big data; Random forest; boruta; neural network; k-nearest neighbor.

I. INTRODUCTION

Machine learning is a branch of artificial intelligence. The World of computer science also uses statistical techniques to provide machines with the opportunity to understand data not being programmed. This field has an extensive lot of use for recognition Patterns and learning in various areas. Numerous ones Studies in machine learning algorithms are used to forecast performance based on new output Inputs or learning about the structure and relationship of the results [1]. Industrial animal raising is a global industry. The Size that most businesses dwarf. The annual carcass rating Meat demand is estimated at around USD 1 trillion. This, Huge scale represents a steady rise in meat consumption as developed nations become more dependent on animal products are a type of protein [2]. A long civilization of agriculture and food Community lets us with a special affinity for pork in our land. Compared to developing pig-farming countries, there is a considerable distance in the World regarding better reproduction, Organizational level, standard output, commodity Performance, manufacturing power, and economic advantages [3] [4].

Pork is the World's most consumed meat product, with over one billion pigs produced worldwide. Over 110 million tons of meat were processed last year (USDA, 2018). The health and output monitoring of pigs is significant, but it may be overwhelming since many farms are substantial [5]. Early identification and treatment of conditions related to human health and wellbeing animal husbandry is a significant concern. Pigs are particularly susceptible to multiple viruses and diseases—stress due to many pigs in the "closed" pig space [6]. Live weights are conventionally taken by driving the animal to the Land scale, which is mentally exhausting for all species and the shop clerk [8]. The main parameter for deciding the status of the pig body is the live weight. Weight variations are straightforward approaches to measure the health and growth of the pig. [7] [8].

Innovent Ltd has developed the QScan-based platform for popular infrared (IR) cameras to monitor pores and pigs to collect pigs' growth figures. Difficult to see the shapes of the pigs. This research aims to describe the image – the picture classification by the type of object that it includes using random forest algorithm and boruta [9-10]. Precision agriculture (PA) is a management term within farm production systems to track, control, and respond to uncertainty. PA relies rather than conventional management methods on particular crops or animals on. PA techniques for produce, dairy/beef, and aquaculture are widely applied. [11].

The composition of the body and its diversity in vivo should be estimated for a pig to assess its potential economic value. Advanced imaging approaches such as MRI, tests for CT scans, ultrasound, and dual-energy X-ray absorptiometry can be used to do this (DXA) [12]. We have a dataset and propose a tool that can reliably derive animal identity from top-down silverware by first segmenting animals into RGB-D frames and then extract a subset of local ASIFT coat descriptors to be enough characteristic across the organism. [13].

II. RELATED WORKS

Siroj Bakoev et al. [14]. Prediction of nine ML (random) forestry classification algorithms, K-Nearest Neighbors, Naive Bays, C50Tree, Vector Support Machines, was checked and compared. Generalized Linear Models, Study, and Analysis of Boost and Linear Discrimination) And the A limited range of critical The Random Forest and the K-

nearest Neightbours have been identified as the most successful algorithms for poor pig leg prediction. These measurements can be developed in the earlier phase of animal development. Measurement Essential pork extremity predictors for muscle length, rear fat, and average daily benefits have been established. This research reveals that the powerful and reasonably easy use of machine-learning algorithms to determine pig limb status depends on meat development rate and characteristics.

WeerapongThanapongtharm et al. [15] proposed a method based on a very comprehensive 2010 census, pig type (native pigs, pig breeding and fattening pigs), agricultural scale (small farming systems and large farming systems), and farm system type (farrow-to-finish, nursery, and finishing systems). Second, I think the goal was to investigate the statistical spatial relationship between these various forms of pig farming distribution and the selection of spatial variables representing feed supply and user access.

Qiming Feng et al. [16] — studied on the one line, the mathematical model consistent with the pig Farm periodicity is created, which makes pig breeding more uniform, which maintains the regular delivery of pigs. On the other hand, by studying the three-year outlook of the pig price curve and Comprehensive analysis of expenditures and profits, Management techniques to achieve the total value could be Sent. Scientific management at the end of the report, the plan of pig farms is laid out.

MaciejOczak et al. [17] proposed a thesis that involves evaluating a system for automatically detecting hostile behavior. Pigs feed into a neural network with a multilayer index. The experiment performed a mixed party of 11 male swine weighing an average of 23 kg. Events. Images of the software that calculates the operating index measured the actions of the animals. Five organizational index characteristics Recorded videos over 14 periods have been determined (average, maximum, minimum, total, and variance). A multilayer feed neural network has been trained and confirmed to detect high aggression and median aggression cases.

Zhang Bingzhen et al. [18] proposed a formula to compute the tree's similarity to solve this issue. They are centered upon the Uncertainty Matrix. The number of trees in various groups is taken into account by this approach, and the grouping of trees is accurate. And the wrong cases and the inaccurate rankings. Simultaneously, random model collection, in conjunction with the Decision Tree Success classification, Forest is finished using the "remove inferior" technique. The experimental findings indicate that the proposed approach has been used in three data sets with greater precision and higher classification consistency than the original algorithm. Therefore, the random Confusion Matrix-based Forest Picture Classification Model can be used. Enhance the random forest's capacity to distinguish.

III. PROPOSED METHOD

The real-time device consists of data calculation, Detection and segmentation of a pig, estimation and predicting of weights, and data analysis. In Fig.1, the computer workflow is shown. 1, where four major parts are included, i.e., acquisition of photographs, acquisition of images, as listed below, performance, data processing, and decision-making.

The time of flight camera is in the picture acquisition part. Could you place it in pen on top of the feeder? IFM's 03D313 3D camera is the moment that the camera we have used is used. GmbH is 176*132, the frame speed of the sensor is 15fps, and the working distance varies from 300 to 8000mm. Although the sensor depth resolution is small, the depth of the images is also insufficient. More suitable than the standard RGB Pictures for pig monitoring. Our suggested algorithm could solve the low-resolution problem.

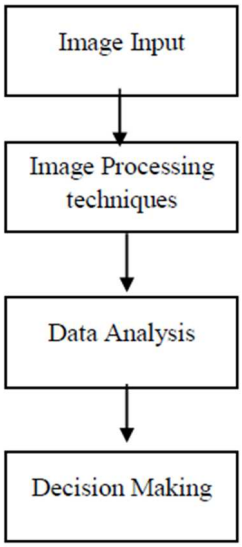

Fig 1. Block Diagram of Proposed system

A. Image Analysis

First of all, we need to classify the objects in the background with the obtained depth images. The pigs are the same as the camera; the backdrop can be shifted from depth to depth with a basic threshold followed by a structured mapping—control of the charts.

fig. 2, Past of only pig items for segmentation and calculation have been deleted successfully making it easier to detect the Objects are although the action of the pigs is erratic and unpredictable, typically they prefer to stay Due to their routines together.

Fig 2. Examples of the initial image depth and findings after the threshold procedure

For the extraction of each contour pig, it is necessary to distinguish them. The greatest challenge here is how the touch can be differentiated. Pigs, specifically. The smaller image resolution means that the practice of segmentation in deep images is not effortless reached.

1. *Random Forest Technique*
   Random Forest [1] is a randomized method of Study that a priori recognizes the target class. And a pattern. (Classification or regression) is developed to simulate possible responses. Many of the 100 decision trees are meant to train bootstrap samples. However, in the tree structure, m from the p predictors can be randomly chosen for each iteration, and only on one of those m variables can the distribution be done[19]. The aim is to increase the consistency of the solutions obtained from materials by applying a technique that has proven very successful.

$$Propbability = (p - m)/p \qquad (1)$$

A theoretically dominant predictor, It avoids the effort to hit some tree. Other predictors would get their chance by blocking the dominants, and the diversity of the tree will increase.

Algorithm1. Random Forest Algorithm [9]

1. Input: Many decision trees for training Similarity threshold d Level of accuracy 5-0077
2. Input: Modern Algorithm of Random Forest RF:
3. for classifying and forecasting the set of research samples, a decision tree is used;
4. For each treeCM decision, each tree's classification findings are counted, and the uncertainty matrix is generated.
5. Creating a metric matrix of similarities for random forestsMF: for (a, b = 1 to T) & (a < b))) Measuring the difference Matrix between decision tree I and decision tree DCMa(ab) Measuring the difference Matrix is derived from DCM (ab) Calculates DCMa(ab) at that point as the element value of MF.
6. The criterion for comparisons is d, and the threshold for grouping capacity is 5-007.
7. At present, Minab=MF is the minor non-zero factor.
8. (<E min)
9. The row and column components in Decision Tree and MF are all set to 0 if (Decision Tree I with less grouping capabilities) (Delete Decision Tree a)
10. Minab = The smallest non-zero element next
11. Otherwise, the element row's tree integration and the non-zero column will join the existing RF End random forest.

2. *Boruta*
   Borut's algorithm is a random forest envelope algorithm. The variations in the mean exactness of forest declines and the value calculated for the score by average decline precision are considered (i.e., Z score). The relation between the feature and the feature focuses primarily on the whole choice of options. The expected value is eliminated by a shadow value that is too essential to evaluate biomedical data before compilation—the feature's relation. In R and Python, man, there are ready-made kits. The following are the algorithms of Borut:

   The original sample function matrix is M; this has been randomly scrambled to establish the sample's shadow character P, then a hybrid function.

$$N = MP \qquad (2)$$

   N was produced at this moment.

B. Image Segmentation

The method of detecting standing pigs, cloud data from the stalk is gathered when the sensor Kinect is mounted on the installed roof. Threshold values filter the depth information to produce a new depth map. After the target identification and selection procedure, standing pig data can be easily collected.

Figure3. RGB image

Figure3 represents the RGB image of the pig farm taken from the above angle in the farm.

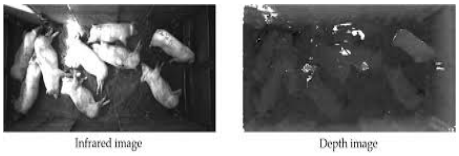

Figure4. IR image and Depth image.

1. *Threshold the Depth*
   The sensor is mounted at a set height, as shown in Fig. 3 and 4. Using the following Equation to set a minimum and maximum deep pixel threshold value:

$$G(i,j) = \{f(x) = \begin{cases} 255, & Dmin \le d(i,j) \le Dmax \\ d(i,j)/16, & otherwise \end{cases}$$

Where g(i, j) is the current pixel value, d(i, j) is the sensor's depth value, The matter below is Dmax, and the deal below is Dmin.

2. *Proposed Image Processing*

The proposed multi-stage image dividing algorithm contains three main image processing phases, i.e., per-segment and post-processing. The vertical direction has more good localization since the profound camera is mounted on the pork farm's roof, right above the feeder.

In the first step, the depth image is converted to a 2-D image to represent the pixel image's vertically oriented value. Of the favorable depth detail, the camera size. This approach is defined as follows.

$$Pixel\ val = \frac{255 * (max - d)}{max - min}$$

D represents a vertical depth value, and max and min are the maximum depth value and minimum depth value as seen in the diagram. As the images often contain many other artifacts like the sinks on the ground, a simple quantification is carried out to settle the picture by deleting these background objects. The new concept, as seen in Fig. 2, successfully eliminated the backdrop for segmentation and calculation with only the pig artifacts left

The result of this paper's marker-checked watershed scheme is seen in Fig. 5. Apart from the algorithm are different artifacts of the pigs insufficient after the distance has been transformed.

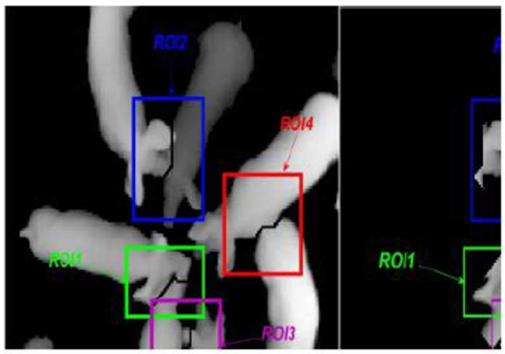

Figure5. ROIs images results of the proposed method.

New ROIs for further processing are automatically generated depending on the separation line. In these ROIs, the geodesic distance approach reinforces the segmented lines. The geodesic observations are closer to the facts of the planet. For, e.g., on the ground truth line, the geodesic line is closely overlapped in the fourth ROI. This showed the effectiveness of multi-stage segmentation in the deep images in distinguishing the moving objects of pigs.

3. *KNN*

The classification of K-NN is based upon a theorem of compactness, which requires an examination. The item in its immediate local area will have the same classmark as training items. Climate. The climate Environment

Algorithm2. KNN ALGORITHM

1. The object being evaluated is assigned to a particular class when K's value is one(1)—centered on the nearest neighbor's data.
2. Any object is allocated to the upper class of closest neighbors, whereas K > 1. Any clustering algorithm, which implies that a partition exists, should be used to fulfill the compact principle.
3. Difference cluster points in objects in the same type (intracluster) Size) are below a specific value when 5-007 > 0, and the distance from things is under 0. The various classes (cross-cluster distances) are more than 5-007.

Test datasets include the animal's average weight in 25 consecutive days, including 7, 0990 data samples shown in figure5. Our linear fitting is based on the least square Equation and matches this data set. Several questions, such as the pig population's epidemic and the pig's pathological behavior, affect the consequences of this.

IV. CONCLUSION

In this article, for measuring weight and growth estimation of the rate of pigs, the analysis method is suggested to monitor real-time circumstances and data. Despite the depth resolution, the picture is shallow, and in one image, there are many pigs. With the planned multi-phase Study of the image segmentation, the sharks' contours can be precisely derived with the random forest and borut algorithms method. We often use linear fitting for the estimated weight value of the pig contours extracted. Piece-wise to calculate overall growth rate inside the pig farm, taking data from big data to process in the neural network for quicker processes and KNN algorithm for the better predictive result to get the resulting pig weight estimate. Although the statistics are noisy and data on the ground reality data is not available, we have excluded external data and increased the robustness of the prediction.

REFERENCES

[1] Mitrović, K., Milošević, D., &Greconici, M. (2019, May). Comparison of Machine Learning Algorithms for Shelter Animal Classification. In *2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (pp. 211-216). IEEE.

[2] Vaughan, J., Green, P. M., Salter, M., Grieve, B., &Ozanyan, K. B. (2017). Floor sensors of animal weight and gait for precision livestock farming. In *2017 IEEE SENSORS* (pp. 1-3). IEEE.

[3] Liu, T., Teng, G., & Fu, W. (2011, September). Research and development of pig weight estimation system based on image. In *2011 International Conference on Electronics, Communications and Control (ICECC)* (pp. 2774-2777). IEEE.

[4] Shipu, X., Yunsheng, W., Wenwen, H., Yingjing, W., & Yong, L. (2020, August). Design and Implementation of Pig Growth

Monitoring System Based on Wireless Communication. In *2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology (CCET)* (pp. 239-243). IEEE.

[5] Zhang, L., Gray, H., Ye, X., Collins, L., &Allinson, N. (2019). Automatic individual pig detection and tracking in pig farms. *Sensors*, *19*(5), 1188.

[6] Seo, J., Ahn, H., Kim, D., Lee, S., Chung, Y., & Park, D. (2020). EmbeddedPigDet—Fast and accurate pig detection for embedded board implementations. *Applied Sciences*, *10*(8), 2878.

[7] Cang, Y., He, H., &Qiao, Y. (2019). An Intelligent Pig Weights Estimate Method Based on Deep Learning in Sow Stall Environments. *IEEE Access*, *7*, 164867-164875.

[8] Baek, H., Chung, Y., Ju, M., Chung, Y., & Park, D. (2017, February). Segmentation of group-housed pigs using concave points and edge information. In *2017 19th International Conference on Advanced Communication Technology (ICACT)* (pp. 563-565). IEEE.

[9] Tang, R., & Zhang, X. (2020, May). CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification. In *2020 5th IEEE International Conference on Big Data Analytics (ICBDA)* (pp. 80-84). IEEE.

[10] Bosch, A., Zisserman, A., & Munoz, X. (2007, October). Image classification using random forests and ferns. In *2007 IEEE 11th international conference on computer vision* (pp. 1-8). Ieee.

[11] Rahman, A., Smith, D., Hills, J., Bishop-Hurley, G., Henry, D., & Rawnsley, R. (2016, July). A comparison of autoencoder and statistical features for cattle behaviour classification. In *2016 international joint conference on neural networks (IJCNN)* (pp. 2954-2960). IEEE.

[12] Alsahaf, A., Azzopardi, G., Ducro, B., Hanenberg, E., Veerkamp, R. F., & Petkov, N. (2019). Estimation of muscle scores of live pigs using a Kinect camera. *IEEE Access*, *7*, 52238-52245.

[13] Andrew, W., Hannuna, S., Campbell, N., & Burghardt, T. (2016, September). Automatic individual holstein friesian cattle identification via selective local coat pattern matching in rgb-d imagery. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 484-488). IEEE.

[14] Bakoev, S., Getmantseva, L., Kolosova, M., Kostyunina, O., Chartier, D. R., &Tatarinova, T. V. (2020). PigLeg: prediction of swine phenotype using machine learning. *PeerJ*, *8*, e8764.

[15] Thanapongtharm, W., Linard, C., Chinson, P., Kasemsuwan, S., Visser, M., Gaughan, A. E., ... & Gilbert, M. (2016). Spatial analysis and characteristics of pig farming in Thailand. *BMC veterinary research*, *12*(1), 1-15.

[16] Feng, Q., & Qian, S. (2019, October). Research on Management of Pig Farms. In *2019 3rd International Conference on Data Science and Business Analytics (ICDSBA)* (pp. 57-60). IEEE.

[17] Oczak, M., Viazzi, S., Ismayilova, G., Sonoda, L. T., Roulston, N., Fels, M., ... &Vranken, E. (2014). Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network. *Biosystems Engineering*, *119*, 89-97.

[18] Bingzhen, Z., Xiaoming, Q., Hemeng, Y., &Zhubo, Z. (2020, August). A Random Forest Classification Model for Transmission Line Image Processing. In *2020 15th International Conference on Computer Science & Education (ICCSE)* (pp. 613-617). IEEE.

[19] Suseendran, G., & Balaganesh, D. (2021). Smart cattle health monitoring system using IoT sensors. *Materials Today: Proceedings.*