Indian Journal of Animal Sciences **93** (4): 401–405, April 2023/Article
https://doi.org/10.56093/ijans.v93i04.131010

Principal component analysis in pig breeds identification

SANKET DAN[1]✉, SATYENDRA NATH MANDAL[1], PRITAM GHOSH[1], SUBHRANIL MUSTAFI[1] and SANTANU BANIK[2]

Kalyani Government Engineering College, Kalyani, Nadia, West Bengal 741 235 India

Received: 2 December 2022; Accepted: 24 March 2023

ABSTRACT

Maintaining the purity of pig breeds is an essential task for their economic value. The traditional breed identification methods through coat colour are prone to error due to huge intra-breed variation. This paper uses principal component Analysis (PCA) to classify the pig breeds using their images. Individual images of five different pure breeds were captured from organized farms in India under both controlled and uncontrolled environments. Three different image sets were created, containing images in the controlled, uncontrolled, and mixed environment image sets. With 80:20 training to testing datasets, 93% accuracy was found in the proposed method of principal component analysis. Finally, two performance-based comparative analyses of our method were done with PCA-based methods and other renowned techniques used for animal breed identification, wherein our PCA method outperformed others in both comparative scenarios.

Keywords: Breed identification, Confusion matrix, Euclidean distance, Image space, Principal components

Animal Husbandry and livestock sector leads a major role in the socio-economic growth of the Country (Neethirajan and Kemp 2021). According to the World Development report 2008, 70% of the rural economic growth depends on the livestock sector (Pica *et al.* 2008). Animal breed detection is one the major tools for boosting the fastest growing livestock sector. Genotype based marking procedures like whole-genome sequencing, microsatellite markers etc. have been used for recognition of animal breed. The phenotypic features such as muzzle print, body shape, coat colours and pattern have been used as the recognition trait(s) of individual animals and also the individual breed (Lahiri *et al.* 2011, Andrew *et al.* 2016, Kumar *et al.* 2018). Based on the phenotypic features, some research works were done for identification of animal breed like dog using Convolution neural network, Artificial neural network, Deep learning (Hsu 2015, Ráduly *et al.* 2018, Borwarnginn *et al.* 2019, Mandal *et al.* 2020, Fuad *et al.* 2021). Principal Component Analysis has been widely used for face detection, animal recognition, image compression etc. In this paper, pig breeds were recognized from their images using Principal Component Analysis. The individual pig from five pig breeds were captured from organized farms located at different places in India. Total 1000 images were captured to build the pig breed dataset where 500 images were captured in a controlled environment and the rest of the images are from uncontrolled environment. In both cases, the side profile of each pig was captured. The captured images were divided 80:20 ratio as training and test sets. The principal component analysis was applied on training sets and the training templates (Central mean, image space and project train image set) were developed. In test phase, the test template (projected test image) was developed for each pig image using central mean and training image space from training phase. The Euclidean distances were calculated among projected test image and each column of projected training image set. The minimum value (Minimum Euclidean distance) was calculated and corresponding column was selected in project training image set. The breed of test image was predicted as same breed at selected column in projected training set. Finally average precision, recall F1-score and accuracy were computed and a comparison was made with accuracies with other breeds from other established method.

MATERIALS AND METHODS

In this paper, the principal component analysis was applied on captured pig images and breeds of images were predicted. The details of methodology is divided into some steps as shown in Fig. 1. The steps were operated sequentially where first block was pig image capturing and final step was pig breed prediction.

Pig image capturing and image grouping: The individual pig was captured in controlled and uncontrolled environments using both mobile phones and DSLR camera. Among the registered and exotic pure breed pigs available in India, five breeds named Duroc, Ghungroo, Yorkshire,

Present address: [1]Kalyani Government Engineering College, Kalyani, Nadia, West Bengal. [2]ICAR-National Research Centre on Pig, Rani, Guwahati, Assam. ✉Corresponding author email-sanketdan@gmail.com

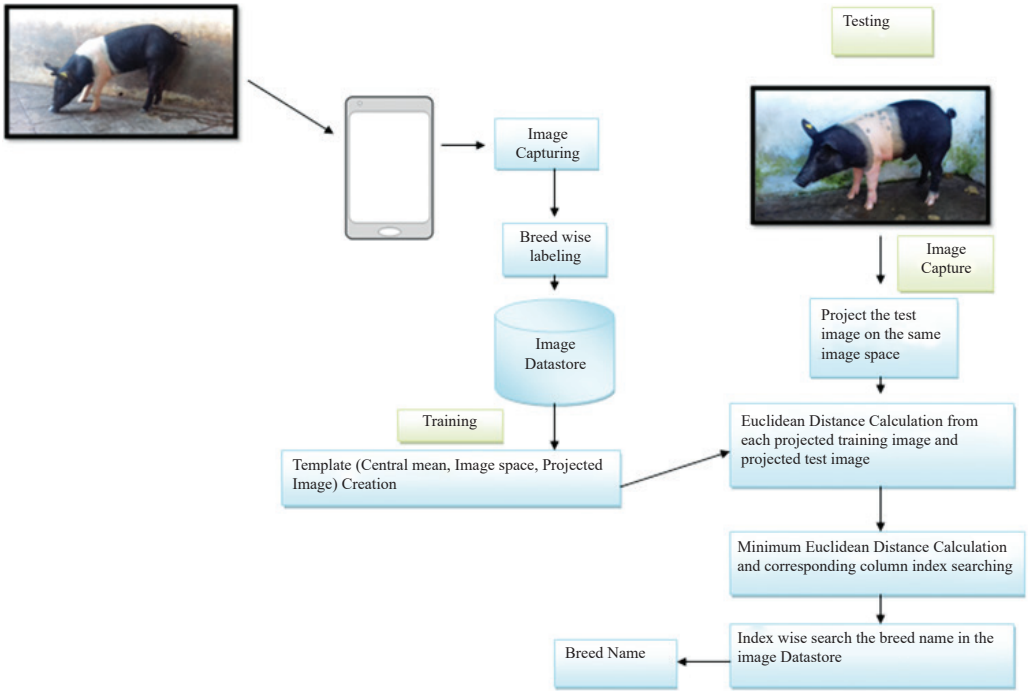

Fig. 1. Steps in pig breed prediction.

Mali, Hampshire were considered for this research. To ensure the breed's purity, the images were captured from the organized farm maintained by research institutes of India, namely ICAR National Research Centre on Pig, Rani, Assam; ICAR Research Complex for NEH Region, Umiam, Meghalaya and ICAR Research Complex for NEH Region, Tripura Centre, Tripura.

In uncontrolled image capturing, the individual pig was photographed when they were moving freely in the farm. The images were captured from almost same distances and camera was focused at centre of pig. The side profile was captured from each individual pig (Supplementary Fig. 1). In controlled environment, a green curtain was placed at the back of the pig in such a way that the captured images have uniform background and other conditions kept same as uncontrolled mage capturing steps (Supplementary Fig. 2).

The captured images were categorized into three groups; controlled, uncontrolled and mixed sets. Pictures (100) were captured from each breed in controlled as well as uncontrolled environments i.e. total 1000 images were captured out of which 500 images were captured in a controlled environment and 500 images in uncontrolled environment and they were grouped accordingly. The mixed image group was constructed from these two sets (Table 1).

Preprocessing: The captured images were different in size and 3D in nature. In preprocessing stage, images were converted into 2D grey scale images and resized into 300 * 300 pixels to have uniformity among the images.

Datastore and labelling: In this paper, images from five different pig breeds were used for breed identification (Supplementary Fig 3). Eighty images from each breed were stored in the training set and the images from each breed were placed in separate folder and each folder was assigned their breed names (Supplementary Fig. 4.).

Principal Component Analysis and model creation: Principal Component analysis is a classical classification method that project the dataset into subspace, localize and evaluate the principal features from the image Data set. The higher dimension correlated vector is converted into lower dimension uncorrelated data. The main advantage of the PCA is that it reduces the dimension of the data, increases the processing efficiency and is less prone to noise.

Main component of the PCA is Eigen vector and Eigen value of the covariance matrix (Shan 2002, Bajwa and Hyder 2005, Rodarmel and Ren *et al.* 2021). Eigen vector is a set of spatial characteristics that can be used for recognition of the object. It determines the direction of the new feature space and Eigen value imply the magnitude of the new feature space. In the training phase, an image space was developed, which was used for identification of the breed in the testing phase Fig. 2.

Table 1. No of images in image groups

Group name	Training set		Test set	
	No of images from controlled group	No of images from uncontrolled group	No of images from controlled group	No of images from uncontrolled group
Controlled	400	0	100	0
Uncontrolled	0	400	0	100
Mixed	200	200	50	50

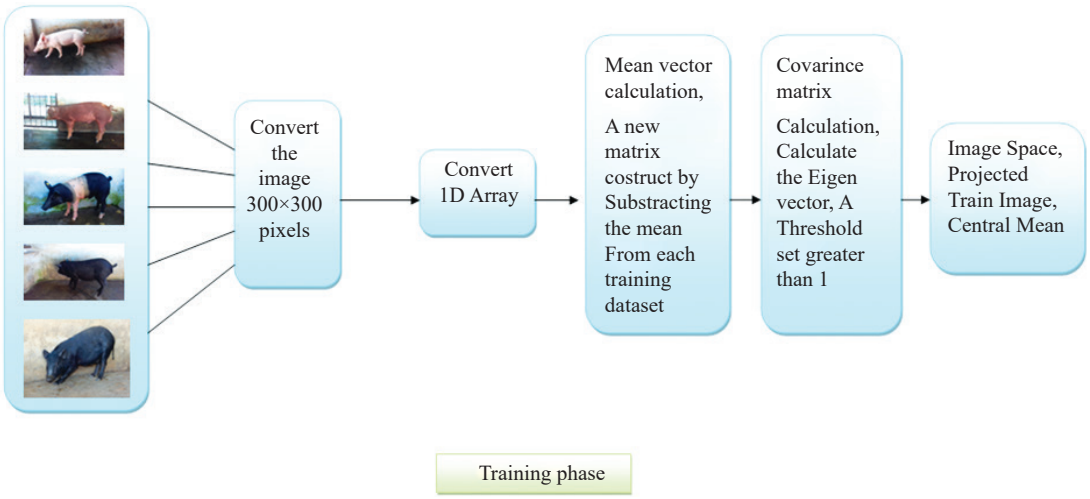

Fig. 2. Training Phase.

Pre-processed images were used in training phase. The measured eigen values and the corresponding eigen vectors were calculated from training images. Some eigen vectors were removed which is less than 1 for reducing the dimension and an image space was created. The images were then projected onto this image space, and stored in the Training Set database. The steps involved in training phase have been summarized in Algorithmic format in Supplementary Fig. 5.

Breed prediction: In the testing phase, each of the new pig image from test set has been analyzed and the principal feature was computed using the central mean and image subspace from training phase (Fig. 3).

Test Image principal feature was compared with the Train Image principal feature using the Euclidean distance. Breed label key corresponding to the minimum Euclidean distance was found and corresponding breed of test image was predicted. The details steps taken for the breed identification are given in Algorithmic format in Supplementary Fig. 6.

RESULTS AND DISCUSSION

Breed prediction in controlled, uncontrolled and mixed groups: In this paper, algorithms were implemented using Matlab 2016b. The training algorithm was applied on three different image sets (Table 1), such as controlled, uncontrolled and mixed image groups. The prediction model for each group was developed. The developed model for each image group was applied on its test image set and the confusion matrices from each set are given in Supplementary Fig. 7, Supplementary Fig. 8 and Supplementary Fig. 9.

The performance matrix is given in Table 2 and the average precision, recall, F1-score and accuracies were computed as 93.86%, 93.29%, 93.57% and 93.00% respectively (Supplementary Fig. 10).

Some algorithms were applied in identification of different animal breeds and prediction accuracies are given in Table 3. Other PCA based prediction models were compared with result produced in this paper is shown in Table 4. The pictorial representation of comparisons is

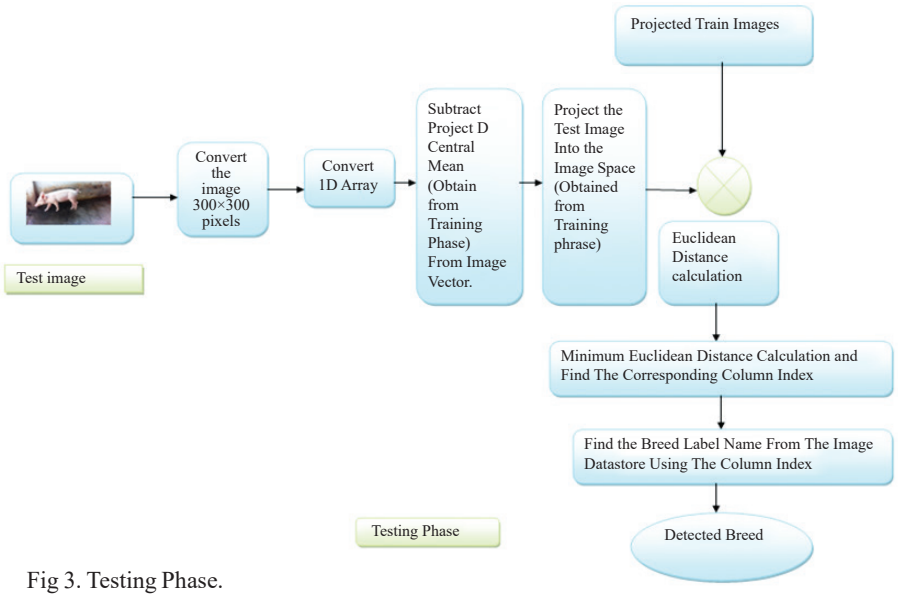

Fig 3. Testing Phase.

Table 2. Performance matrix

Group	Precession	Recall	F1-Score	Accuracy(%)
Controlled	98.18%	98.00%	98.08%	98.00%
Uncontrolled	90.87%	89.89%	90.38%	89.00%
Mixed	92.55%	92.00%	92.27%	92.00%
Average	93.86%	93.29%	93.57%	93.00%

given in Supplementary Fig. 11 and Supplementary Fig. 12. It is observed from Table 4, the prediction accuracies of the order controlled (98%)> mixed (92%) > and uncontrolled (89%). This is due to the non-uniform background of the captured images in uncontrolled images.

The average accuracy of predicted model is 93% i.e. the model accurately predicted breeds of 93 individual out of 100 test sample. The false positive and false negative does not influence greatly. The precision is 93.86% i.e. model correctly predicted breed of almost 94 out of 100 pig images, the value is pretty good. The recall is 93.29% i.e. the model predicts almost 94 pig images correctly from 100 test images. the value is good as it is >0.5. F1-score is weighted average of precision and recall (Huilgol, n.d.). The F1-score value of 93.57% and establishes that our model is good. Finally, in this paper, the model was given high accuracy values with high precision and recall values.

Comparison of different breed prediction algorithms: Three different breed prediction algorithms were compared with our algorithm. Among the three algorithms, the prediction accuracies are of the order: Convolutional Neural Network (CNN) (87.42%) > Local Binary Pattern Descriptor (LBP) (70.35%) > linear Discriminant Analysis (LDA) (67.33%) and our PCA based approach showed better result than the rest (Table 3).

Table 3. Comparison of different breed prediction algorithms

Algorithm	Database	Accuracy
Local Binary Pattern Descriptor (LBP) (Cai and Li 2013)	Cow	70.35%
Linear Discriminant Analysis (LDA) (Kamencay *et al.* 2016)	Bear, Hog, Dear, Fox and Wolf	67.33%
Convolutional Neural Network (CNN) (Rishita and Harris 2018)	Dog	87.42%
Principal Component Analysis (PCA) (Present Paper)	Pig	93.00%

Comparison of different pca-based algorithms used in breed prediction: Principal component analysis was applied to different animal databases containing different animal breed images. The databases and the prediction accuracies are given in Table 4. It was observed that the PCA based model in this paper has outperformed the others.

In this paper, principal component analysis was applied on three groups of pig images captured in controlled, uncontrolled and mixed sets. The confusion matrixes were generated. The average precession, recall, F1-score and accuracies were computed as 93.86%, 93.29%, 93.57% and 93.00%. The average accuracy was compared with other applied algorithms in prediction of different animals. The average accuracy was also compared with other results predicted by PCA. It is observed that the PCA based pig breeds prediction had given better result not only with prediction results in different applications but also given better result in breed prediction using algorithms. As PCA based prediction model has outperformed the rest of the applied methods, the PCA based model can be used for prediction of pig breeds from their images.

Table 4. Comparison of different PCA-Based algorithms used in breed prediction

Species	Accuracy
Dog (Prasong and Chamnongthai 2012)	70.35%
Dog (Kumar and Singh 2018)	81.89%
Dog (Dandil and Polattimur 2018)	89.50%
Cow (Dandil and Polattimur 2018)	91.00%
Cat (Dandil and Polattimur 2018)	83.50%
Goat (Dandil and Polattimur 2018)	87.50
Rabbit (Dandil and Polattimur 2018)	86.00
Pig (Present paper)	93.00%

ACKNOWLEDGEMENTS

The authors would like to thank UGC (University Grants Commission) for granting us to pursue the research work. Authors are also grateful to Dr A. Bandopadhyay, Senior Consultant, ITRA Ag & Food; Dr. Saurabh Kumar Das, Principal, Kalyani Government Engineering College; Dr. Binay Singh, Scientist, ICAR-RC for NEH Region, Tripura Center, Agartala; Dr. Dilip Kumar Hajra, Assistant Professor, Department of Agronomy, UBKV, Pundibari, Cooch Behar; Dr. Arnab Sen, Head, Animal Health, ICAR Research Complex for NEH, Barapani to carry out this research work.

REFERENCES

Andrew, William, Sion Hannuna, Neill Campbell and Tilo Burghardt. 2016. Automatic individual Holstein Friesian cattle identification via selective local coat pattern matching in Rgb-d Imagery. *2016 IEEE International Conference on Image Processing (ICIP)*, 484–88.

Bajwa, Imran S and S Irfan Hyder. 2005. PCA based image classification of single-layered cloud types. *Proceedings of the IEEE Symposium on Emerging Technologies, 2005.*, 365–69.

Borwarnginn, Punyanuch, Kittikhun Thongkanchorn, Sarattha Kanchanapreechakorn and Worapan Kusakunniran. 2019. Breakthrough conventional based approach for dog breed classification Using CNN with transfer learning. *11th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 1–5.

Cai, Cheng and Jianqiao Li. 2013. Cattle face recognition using local binary pattern descriptor. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 1–4.

Dandil, Emre and Rukiye Polattimur. 2018. PCA-Based Animal Classification System. *2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 1–5.

Fuad, Md Tahmid Hasan, Awal Ahmed Fime, Delowar Sikder, Md Akil Raihan Iftee, Jakaria Rabbi, Mabrook S Al-Rakhami, Abdu Gumaei, Ovishake Sen, Mohtasim Fuad and Md Nazrul Islam. 2021. Recent advances in deep learning techniques for face recognition. *IEEE Access* **9**: 99112–42.

Hsu, David. 2015. Using convolutional neural networks to classify dog breeds. *CS231n: Convolutional Neural Networks for Visual Recognition [Course Webpage]* 2.

Huilgol, Purva. n.d. "Accuracy vs. F1-Score."

Kamencay, Patrik, Tibor Trnovszky, Miroslav Benco, Robert Hudec, Peter Sykora and Andrej Satnik. 2016. Accurate wild animal recognition using PCA, LDA and LBPH. *ELEKTRO*, 62–67.

Kumar, Santosh, Amit Pandey, K Sai Ram Satwik, Sunil Kumar, Sanjay Kumar Singh, Amit Kumar Singh, and Anand Mohan. 2018. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* **116**: 1–17.

Kumar, Santosh, and Sanjay Kumar Singh. 2018. Monitoring of pet animal in smart cities using animal biometrics. *Future Generation Computer Systems* **83**: 553–63.

Lahiri, Mayank, Chayant Tantipathananandh, Rosemary Warungu, Daniel I Rubenstein and Tanya Y Berger-Wolf. 2011. Biometric animal databases from field photographs: identification of individual zebra in the wild. *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, 1–8.

Mandal, Satyendra Nath, Pritam Ghosh, Kaushik Mukherjee, Sanket Dan, Subhranil Mustafi, Kunal Roy, Dilip Kumar Hajra, and Santanu Banik. 2020. InceptGI: A convnet-based classification model for identifying goat breeds in India. *Journal of The Institution of Engineers (India): Series B*, 1–12.

Neethirajan, Suresh and Bas Kemp. 2021. Digital Livestock Farming. *Sensing and Bio-Sensing Research* **32**: 100408.

Pica, G, U Pica-Ciamarra, J Otte, and others. 2008. The Livestock Sector in the World Development Report 2008: Re-Assessing the Policy Priorities.

Prasong, Pusit and Kosin Chamnongthai. 2012. Face-recognition-based dog-breed classification using size and position of each local part, and Pca." *9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 1–5.

Ráduly, Zalán, Csaba Sulyok, Zsolt Vadászi and Attila Zölde. 2018. Dog Breed identification using deep learning. *IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, 271–76.

Ren, Yanli, Xiao Xu, Guorui Feng and Xinpeng Zhang. 2021. Non-interactive and secure outsourcing of PCA-based face recognition. *Computers & Security* **110**: 102416.

Rishita, Middi Venkata Sai and Tanvir Ahmed Harris. 2018. Dog breed classifier using convolutional neural networks. *International Conference on Networking, Embedded and Wireless Systems (ICNEWS)*, 1–7.

Rodarmel, Craig and Jie Shan. 2002. Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science* **62**(2): 115–22.