VIP-STB farm: scale-up village to county/province level to support science and technology at backyard (STB) program.

YAN, Y., ZHAO, S., FANG, Y., LIU, Y., CHEN, Z. and REN, J.

2022

VIP-STB Farm: Scale-up Village to County/Province Level to support Science and Technology at Backyard (STB) Program

Yijun Yan[1], Sophia Zhao[1], Yuxi Fang[1], Yuren Liu[1], Zhongxin Chen[2], Jinchang Ren[1*]

[1] Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
[2] Institute of Agricultural Resources & Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China

*Corresponding author: Jinchang.Ren@strath.ac.uk

Abstract. In this paper, we introduce a new concept in VIP-STB, a funded project through Agri-Tech in China: Newton Network+ (ATCNN), in developing feasible solutions towards scaling-up STB from village level to upper level via some generic models and systems. There are three tasks in this project, i.e. normalized difference vegetation index (NDVI) estimation, wheat density estimation and household-based small farms (HBSF) engagement. In the first task, several machine learning models have been used to evaluate the performance of NDVI estimation. In the second task, integrated software via Python and Twilio is developed to improve communication services and engagement for HBSFs, and provides technical capabilities. In the third task, crop density/population is predicted by conventional image processing techniques. The objectives and strategy for VIP-STB are described, experimental results on each task are presented, and more details on each model that has been implemented are also provided with future development guidance.

Keywords: Precision agriculture, Machine learning, Information fusion.

1 Introduction

According to the national statistics announced in Dec 2017, ~25% of the population or ~314million people in China are working in the agriculture sector, where less than 20% is below 35-year old. Among them, over 43% have been educated up to primary school level, and only less than 8.5% have achieved the senior high-school level. Within these, ~96% are household-based small farms (HBSF). For the aging and poor-educated HBSFs, their poverty has been a persistent problem that affects the social and economic development of China.

To tackle this serious problem, the Chinese government has put strategic plans to innovate HBSF, where STB is one of the successful examples. In STB, skilled and well-educated researchers are assigned to villages to identify problems and provide solu-

tions. By closely monitoring the environmental parameters and crop growing conditions with sensors and smartphones, the production yield has been improved by 90% whilst the environmental factors have been reduced by 30-40%.

Although STB shows a successful model for innovation of HBSF, several critical drawbacks have constrained its migration from village level to the county or province level as detailed below.

1) Labour intensive: manual acquisition of data for monitoring growing conditions and estimation of yield is very labour intensive and costly thus economically inviable and non-scalable;
2) Lack of automation: Empirical guidance was adopted followed by estimated plant density and yield which seems to be not statistical sound and effective as modern HPC and machine learning can offer smart decision-making with negligible extra cost;
3) Inefficient communications in response to recommendations sent in texts;
4) Environmental issues and limited sustainability due to uncontrolled waste of water and other resources as well as potential land degradation by over-applying fertiliser and chemicals.

Taking STBs in Laoling City and Yangxin County of Shandong Province for a case study, we aim to demonstrate feasible solutions to improve the balanced, quality-ensured and sustainable innovation of rural areas of China. Some effective techniques and systems are used in the project and three contributions are summarized as follows:

1) Automatic NDVI estimation agricultural digital camera (ADC) from multispectral data acquired from Landsat satellite;
2) Effective TTS interface is built up to ensure HBSFs understand the recommendations and act timely;
3) Automation of the process to reduce the labour cost, where estimation from remote sensing data will be derived to replace the labour-intensive manual data acquisition from fields which are separately located and can be hard to access under severe weather conditions, where image processing techniques will be used to estimate crop population density based on readily available datasets;

The outline of this paper is as follows: Section 2 evaluates the performance of different regression methods for the ADC-NDVI estimation purpose. Section 3 describes the implementation of the test-to-speech (TTS) module. Automatic estimation of crop density is introduced in Section 4. Finally, some concluding remarks and future work are summarized in Section 5.

2 Estimation of ADC-NDVI from remote sensing data

As one of the most important index, NDVI is usually used to detect vegetation growth status and coverage, etc. Generally, the NDVI value is calculated by TM data acquired from remote sense, and we name it as TM-NDVI in this paper. For more accurate statistics, agricultural digital camera (ADC) is used to determine the NDVI which is named as ADC-NDVI in the rest of the paper, where several students are needed to work for days to serve one particular STB site. However, manual acquisition

of data is very labor intensive and costly, which is economically inviable and affect the scalability of STB to the upper level. To reduce the labour cost, the estimation from remote sensing data will be derived to replace the labour-intensive manual data acquisition from fields which are separately located and can be hard to access under severe weather conditions. In this section, we design two experiments to estimate the ADC-NDVI through TM data. As the TM-NDVI is calculated by NIR and red band which are TM3 and TM4, thus, TM3-4 is set as a baseline feature to compare. Five machine learning models (i.e. Ridge regression [1], Support vector regression (SVR)[2, 3], Cascade neural network[4], Random forest[5] and Gaussian kernel regression[6, 7]) are used in this work to evaluate the prediction performance in terms of RMSE and r^2. The reason for choosing these models is because they are all classic models and have good capabilities for many regression problems in the real world[8-10].

The TM data is acquired from Landsat, and SPAD, LAI and ADC-NDVI data is provided by a Chinese partner. There are 110 fields with one ADC-NDVI value and 6 TM bands data where TM1 is blue band ranging from 0.45-0.52 um, TM2 is green band ranging from 0.52-0.60 um, TM3 is red band ranging from 0.62-0.69um, TM4 is near infrared ray (NIR) band ranging from 0.76-0.97um, TM5 is middle-infrared band ranging from 1.55-1.75um and TM6 is thermal infrared band ranging from 10.4-12.5um. We use 50% data for training and 50% data for testing. The prediction performance is evaluated in terms of RMSE and r^2 followed by the standard deviation in the bracket which is used to show their stability.

In the first strategy, TM1-6 multispectral information is used to be training and testing features. From Table 1, it can be seen TM1-6 shows better overall performance than TM3-4, which means more spectral information from remote sense is helpful for the estimation of ADC-NDVI. The cascade neural network shows the best prediction performance on TM1-6 but it cost much computation source. This is because the initial weight of the hidden layer in the neural network is randomly selected, and some ill-

Table 1 Performance evaluation in terms of RMSE and r^2 in strategy 1

Regressors	TM3-4		TM1-6		
	RMSE	r^2	RMSE	r^2	Time (s)
Ridge	0.030 ($\pm$0.0017)	0.708 ($\pm$0.0465)	0.028 ($\pm$0.0024)	0.731 ($\pm$0.0704)	0.0089
SVR	0.106 ($\pm$0.0072)	0.698 ($\pm$0.0574)	0.101 ($\pm$0.0129)	0.720 ($\pm$0.0800)	3.7471
Random forest	0.035 ($\pm$0.0021)	0.596 ($\pm$0.0488)	0.036 ($\pm$0.0028)	0.579 ($\pm$0.0479)	0.1094
Cascade neural network	0.030 ($\pm$0.0027)	0.702 ($\pm$0.0632)	0.025 ($\pm$0.0032)	0.778 ($\pm$0.0636)	3.1944
Gaussian kernel	0.031 ($\pm$0.0022)	0.674 ($\pm$0.0615)	0.028 ($\pm$0.0019)	0.726 ($\pm$0.0542)	0.0371
Overall	0.046	0.676	0.044	0.705	

Table 2 Performance evaluation in terms of RMSE and r^2 in strategy 2

Regressors	TM3-4		TM3-4+LAI+SPAD		
	RMSE	r^2	RMSE	r^2	Time (s)
Ridge	0.030 ($\pm$0.0017)	0.708 ($\pm$0.0465)	0.028 ($\pm$0.0024)	0.742 ($\pm$0.0449)	0.0228
SVR	0.106 ($\pm$0.0072)	0.698 ($\pm$0.0574)	0.095 ($\pm$0.0129)	0.764 ($\pm$0.0576)	3.8931
Random forest	0.035 ($\pm$0.0021)	0.596 ($\pm$0.0488)	0.033 ($\pm$0.0028)	0.648 ($\pm$0.0241)	0.1859
Cascade neural network	0.030 ($\pm$0.0027)	0.702 ($\pm$0.0632)	0.025 ($\pm$0.0032)	0.793 ($\pm$0.0355)	2.4197
Gaussian kernel	0.031 ($\pm$0.0022)	0.674 ($\pm$0.0615)	0.028 ($\pm$0.0019)	0.737 ($\pm$0.0497)	0.1508
Overall	0.046	0.676	0.042	0.735	

suited initial value will cause the convergence time longer. Ridge regression has the lowest computation cost, and the second-best stability and prediction performance. This is because ridge regression is mainly used to solve the linear regression problem. Although ADC-NDVI and TM-NDVI are derived from different sources, it still has some linear connection between each other. In addition, Gaussian kernel regression and SVR have similar prediction performance, but the former has the best stability. Random forest has the worst performance. As a result, ridge regression and cascade neural network can be considered as the top two methods with either best usefulness or efficiency.

In the second strategy, TM3-4, SPAD and LAI are used to be training and testing features. SPAD and LAI are two important parameters of crops, which have been already given by the Chinese partner. Therefore, how are those parameters and spectral data related to ADC-NDVI will be investigated in our experiment. From Table 2**Error! Reference source not found.**, the random forest regression still performs the worst. For the other four techniques, their performance in strategy 2 is better than that in strategy 1. Gaussian kernel and ridge have just little gap between each other but are still not as good as cascade neural network and SVR. The prediction accuracy and stability of cascade neural network are better than SVR and the computation time is lower in strategy 2 than that in strategy 1. As a result, cascade neural network works the best in strategy 2. However, it relies on the feature selection, which affects its performance and computation cost much.

From the experimental results, some findings are summarized as follows:
1) LAI and SPAD are more useful than TM 1,2,5,6 for ADC-NDVI estimation.
2) ADC-NDVI can be potentially predicted through remote sensing data.
3) Both cascade neural network and ridge are the best two model which either has best prediction performance or the most efficient.
4) With better feature selection, the prediction performance of most regression model can be well improved.

For the future work, instead of using TM1-6, SWIR data (1000-1700nm) can be also used for NDVI prediction. Then some novel band selection methods[11, 12] can be used to extract the most useful information and help to get more accurate prediction results. In addition, some novel deep learning methods such as segmented auto encoder[13] and deep neural network[14] can be also used to improve the prediction performance.

3   Text to speech module for HBSFs engagement

This section will present a TTS module which can send a voice call message to farmers and remind them to finish the farming task(s). It is very important since each type of crop has different growth condition, irrigation and cultivation strategies. Wrong cultivation strategies or wrong irrigation time may affect the yield of the crops. Although smartphones are recommended as the best way to communicate with HBSFs to provide them crop conditions and suggested operations, these may not be fulfilled as 50% HBSFs failed to receive the text message or to respond accordingly. Also, it is found 57% HBSfs have no smartphones and only 31% using social media such as WeChat, due to the aging and poorly-educated background and limited income. This

Fig. 1. Concept of TTS module.

leads to a communication barrier where effective solutions are needed, simply because a large portion of them cannot read or understand the instructions in texts. Therefore, it is necessary to develop a TTS module to help those farmers make the right move.

Fig. 1 shows the concept of the TTS module. Twilio is an open PaaS (platform as a service) platform which focuses on communication services and provides technical capabilities. It is a well-known and leading cloud computing communication company in the world, which has more than 50 million registered develops and three billion market cap. Twilio packages the complex underlying communication function into API which allows software developers to programmatically make the function of phone calls, messaging and VoIP (voice on internet protocol) on any web, desktop and mobile applications. In another word, any function can be achieved by a few lines of code. Although the Twilio service is not free, the price is still very cheap. It uses pay-as-you-go model and the price of voice and message is $0.0218/min and $0.028/min, respectively. In addition, it has not only web platform, but also mobile Twilio Client which can be used to Android and iphone platform. It means the voice and messaging functions can be also added into any mobile app, which benefits the users a lot. In the future, we can also develop a special TTS app for HBSFs.

Due to too many advantages, we employ the Twilio's voice and messaging functions and build a basic interactive GUI (Fig. 2) in Python to call its functions. Once the
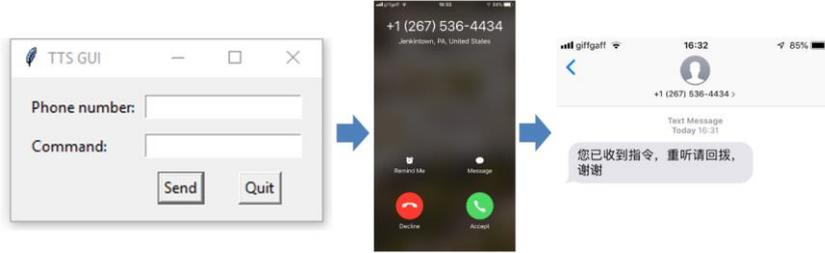
Fig. 2. Basic GUI for TTS

farmer's phone number and command are input, the farmer's phone will receive a voice call. If the farmer doesn't hear the voice call clearly, he can also call back to the server and rehear it. The workflow of the TTS module involves the following steps:

1) Input message
2) Encode message to be used in a URL and create a XML file
3) Return XML file to Twilio cloud server
4) Transfer the XML file to a MP3 file
5) Call from Twilio's number to target's number and play the MP3.

4 Automatic estimation of crop density/population from images

In this section, a threshold-based segmentation method is introduced to calculate the crop density (d) in an image. With the growth of the crops, the density of crops, as a key factor of the final yield, is increasing as well. Therefore, the precise calculation of the crop density is very helpful for yield prediction. However, the existing challenge is most official agricultural data is acquired from the satellite and the resolution of those image data is very low, which leads to density calculation difficult.

Although we don't have the satellite data, we simulate the low-resolution condition by rescaling the image data acquired from some Chinese field (Fig. 3). Leaf detection in this work is performed by color discrimination. Unlike the traditional Red-Green-Blue (RGB) color space, the Hue-Saturation-Value (HSV) approach involves parametrization including not only true color (hue) but also color depth (saturation) and color darkness (value), as can be seen in Fig. 4. As a result, the HSV color space is
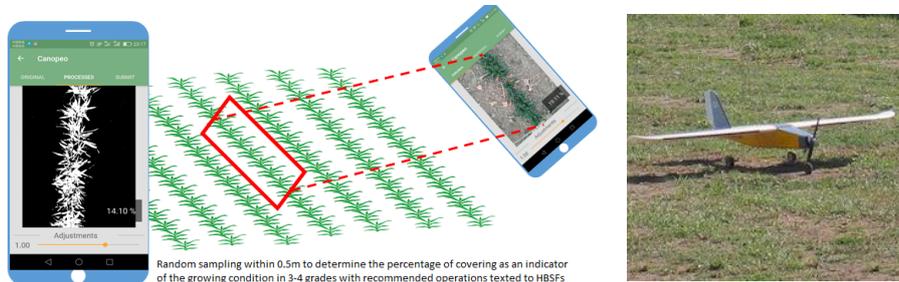

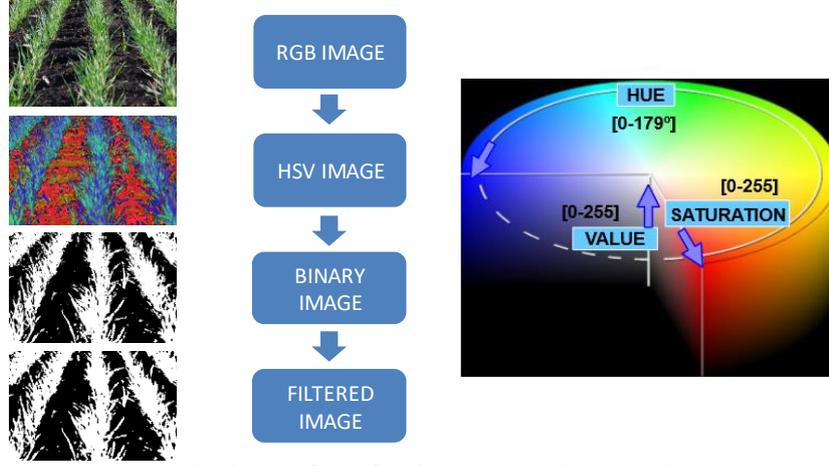Fig. 3. Manual data acquisition for crop density estimation.

Fig. 4. Workflow of leaf area calculation method.

much more suited for addressing real-world environments consisting of light reflections, shadows and darkened regions etc. Therefore, the real-time image processing workflow involves the following steps (Fig. 4): (i) transformation from RGB image to HSV image, (ii) binalization of HSV image, by means of applying the selected HSV color range thresholds, and (iii) posterior treatment of the binary image, including size filtering and morphologies, to avoid the detection of unrelated pixels. Finally, the crop density (D) is calculated by the ratio of foreground pixels (F_p) and whole pixels (W_p) in the segmented result.

Due to the color of crop leaf is mostly green include light and dark green. The selected color range thresholds of H, S and V is usually defined by $H \in [35°, 99°], S \in [43, 255], V \in [46, 255]$, respectively. With reducing the resolution of the image, the bit per pixel (bpp) is also getting lower and lower. In the experiment, we also notice that the density value will become larger when the resolution of the image decreases. To avoid too much variate, we introduce a penalty value γ so that the final density value (Eq.1) can keep constant under different image scale.

$$D = \frac{F_p}{W_p} - \gamma \qquad (1)$$

The penalty value γ is estimated by ridge regression which is expressed in Eq.2:

$$\gamma = (X^T X + \lambda I)^{-1} X^T y \qquad (2)$$

where X is the scale rate ranging from 1 to 100 with 1 interval, $y = F_p * W_p^{-1} - groundtruth, \lambda = 0.001$.

From Table 3, we can see the density value under different scale is very close to the that of the ground-truth, and the MAE is 0.02. However, due to limited sample, the performance of this model needs to further validate on more samples in the future. To further improve the segmentation method, saliency detection[15], image segmentation[16] and deep learning method[17] can be employed to get better segmentation performance and also get more accurate density/population estimation.

Table 3 Visualization of low resolution image and density value under different scale.

Image	Legend	X (%)	Bpp	D (%)
	50% scale	Ground-truth 100	17.03	45.26
		50	14.94	45.29
		10	7.17	45.39
	8% scale	9	6.71	45.12
		8	6.21	45.40
		7	5.68	45.11
	5% scale	6	5.41	45.43
		5	4.81	45.08
		4	4.23	45.46
	1% scale	3	3.81	45.02
		2	3	45.60
		1	1.34	44.75

5   Conclusion

In this paper, by introducing data fusion and AI-driven machine learning techniques, three solutions are derived towards three different challenges of Chinese agriculture. The performance still has many rooms to improve, but the STB programme for economic growth in precision agriculture is scaled up with such concepts, especially in promoting the largest and most vulnerable groups, i.e. HBSFs, and has a significant impact to improve the balanced, quality-ensured and sustainable innovation of rural areas of China. Our future work mainly focuses on the improvement of current three solutions which are summarised as follows:

1) Fusion of multi-modal and multi-source data for accurate modeling and prediction including field measurements and remotely sensed.
2) AI-driven method for more accurate estimation of crop population.
3) Improvement of TTS module where a more interactive function will be included for more effective communication with HBSFs to fulfill the recommended operations.

References

1. A. E. Hoerl, R. W. Kannard, and K. F. Baldwin, "Ridge regression: some simulations," *Communications in Statistics-Theory and Methods,* vol. 4, pp. 105-123, 1975.
2. A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing,* vol. 14, pp. 199-222, 2004.
3. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM transactions on intelligent systems and technology (TIST),* vol. 2, p. 27, 2011.
4. B. Warsito, R. Santoso, and H. Yasin, "Cascade Forward Neural Network for Time Series Prediction," in *Journal of Physics: Conference Series*, 2018, p. 012097.
5. L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.
6. P.-S. Huang, H. Avron, T. N. Sainath, V. Sindhwani, and B. Ramabhadran, "Kernel methods match deep neural networks on timit," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 205-209.
7. A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in neural information processing systems*, 2008, pp. 1177-1184.
8. J. Tschannerl, J. Ren, F. Jack, J. Krause, H. Zhao, W. Huang*, et al.*, "Potential of UV and SWIR hyperspectral imaging for determination of levels of phenolic flavour compounds in peated barley malt," *Food chemistry,* vol. 270, pp. 105-112, 2019.
9. N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: techniques and challenges," *Sensors,* vol. 19, p. 1423, 2019.
10. W. Zhang and A. T. Goh, "Multivariate adaptive regression splines and neural network models for prediction of pile drivability," *Geoscience Frontiers,* vol. 7, pp. 45-52, 2016.
11. J. Tschannerl, J. Ren, P. Yuen, G. Sun, H. Zhao, Z. Yang*, et al.*, "MIMR-DGSA: Unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm," *Information Fusion,* vol. 51, pp. 189-200, 2019.
12. Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE transactions on neural networks and learning systems,* vol. 27, pp. 1279-1289, 2016.
13. J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang*, et al.*, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing,* vol. 185, pp. 1-10, 2016.
14. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche*, et al.*, "Mastering the game of Go with deep neural networks and tree search," *nature,* vol. 529, p. 484, 2016.
15. Y. Yan, J. Ren, G. Sun, H. Zhao, J. Han, X. Li*, et al.*, "Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement," *Pattern Recognition,* vol. 79, pp. 65-78, 2018.
16. X. Xie, G. Xie, X. Xu, L. Cui, and J. Ren, "Automatic Image Segmentation With Superpixels and Image-Level Labels," *IEEE Access,* vol. 7, pp. 10999-11009, 2019.
17. X. Xu, G. Li, G. Xie, J. Ren, and X. Xie, "Weakly Supervised Deep Semantic Segmentation Using CNN and ELM with Semantic Candidate Regions," *Complexity,* vol. 2019, 2019.