Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons

DABIN ZHANG, SHANYING CHEN, LIWEN LING, AND QIANG XIA

College of Mathematics and Informatics, South China Agricultural University, Guangdong 510642, China

Corresponding author: Liwen Ling (linglw@scau.edu.cn)

ABSTRACT The fluctuations of agricultural commodity prices have a great impact on people's daily lives as well as the inputs and outputs of agricultural production. An accurate forecast of commodity prices is therefore essential if agricultural authorities are to make scientific decisions. To forecast prices more adaptively, this study proposes a novel model selection framework which includes time series features and forecast horizons. Twenty-nine features are used to depict agricultural commodity prices and three intelligent models are specified as the candidate forecast models; namely, artificial neural network (ANN), support vector regression (SVR), and extreme learning machine (ELM). Both random forest (RF) and support vector machine (SVM) are applied to learn the underlying relationships between the features and the performances of the candidate models. Additionally, a minimum redundancy and maximum relevance approach (MRMR) is employed to reduce feature redundancy and further improve the forecast accuracy. The experimental results demonstrate that, firstly, the proposed model selection framework has a better forecast performance compared with the optimal candidate model and simple model average; secondly, feature reduction is a workable approach to further improve the performance of the model selection framework; and thirdly, for bean and pig grain products, different distributions of the time series features lead to a different selection of the optimal models.

INDEX TERMS Model selection, agricultural commodity, price forecasting, time series features, forecast horizons.

I. INTRODUCTION

Agricultural commodities are essential to people's daily lives. In recent years, the price fluctuations of agricultural commodities have become more severe and have exerted negative effects on society. For the consumer, an excessive increase in prices will impose a great burden on people's food expenditures, thus impacting their general welfare. For the agriculturalist, large price fluctuations will increase the uncertainty of production, thus adding to the number of risks that must be managed. Consequently, an accurate prediction of the price of agricultural commodities is vital for agricultural authorities to make scientific decisions and to guarantee a favorable operation of the social economy.

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Tucci.

The literature provides a number of methods to forecast the prices of agricultural commodities, including statistical methods and intelligent methods. Statistical methods are the most popular methods for forecasting a time series. For instance, Darekar and Reddy [1] predicted the cotton price of major producing states in India with auto-regressive integrated moving average model (ARIMA). Xu *et al.* [2] used an exponential smoothing model (ETS) to forecast the carrot price in China. Evans and Nalampang [3] employed a multivariate regression model to forecast the price trend of U.S. avocado. In recent years, as agricultural commodity price series become more volatile, powerful AI models with favorable self-learning capability have emerged to handle with the complex price forecasting task. For example, Wei, *et al.* [4] employed a back-propagation neural network to predict the time series for several agricultural commodity prices.

Xiong, *et al.* [5] proposed the STL-ELM method for forecasting vegetable prices in China. Liu *et al.* [6] predicted the cyclical and trend components of hog prices using a sub-series search method and SVR. All of these studies consistently report their superiority compared with statistical models. More studies over the past ten years are shown in Table 1.

Authors	Commodity	Forecast model
Xu, et al. [18]	sugar	BP Neural Network
Jha, et al. [19]	oilseed	Time-delay Neural Network
Zhang, et al. [20]	tomato	Wavelet Neural Network
Xiong, et al. [21]	cotton, corn	VECM-MSVR, SSVR
Ayankoya, et al. [22]	grain commodities	BP Neural Network
Cai, et al. [23]	pork	EMD-SVR
Adya, et al. [16]	sugar, cotton, corn, soybean, coffee	MSS-ANN
Duan, et al. [24]	mandarin(fish), metapenaeus(prawn), portunus(crab)	CA-SVR
He, et al. [25]	soybean	APSO-SVR
Li, et al. [26]	potato	Dynamic Chaotic Neural Network
Wang, et al. [27]	corn	SSA-ELM
Xiong, et al. [5]	cabbage, pepper, cucumber, green bean, and tomato	STL-ELM, TDNN, SVR, ELM
Zhou, et al. [28]	wheat	APSO-SVR
Zhong, et al. [29]	soybean	QR-RBF Neural Network
Zheng, et al. [30]	many samples	HGWO-SVR
Wu, et al. [31]	pork	RBF Neural Network

It can be seen from Table 1 that various kinds of models are widely used for different agricultural commodity forecasting tasks. According to the 'no free lunch' theory [7], there is no single model suitable for all the commodities. When facing a new type of agricultural commodity, it is not easy for people to identify which is the optimal model for this specific forecasting task. Of course, decision makers can compare the performance of several commonly used forecasting techniques and figure out the most favorable one. However, training various models is a time-consuming process. Obviously, a fast and automatic algorithm is needed to identifying the most suitable forecasting method for agricultural commodities.

In the past 30 years, the model selection approach has been used extensively for choosing the optimal model for various types of input data. That is to say, the underlying relationships between the features of the input data and the performance of a candidate algorithm will be discovered by learners through numerous training samples. Once there are some new data, the optimal model will be selected automatically by the trained learners based on the features of the new data. Although training for the learner takes much computational time, the pay-off could be a significant gain in being able to choose the optimal model for a new series more quickly. Therefore, we propose to use the model selection

method to select the optimal forecast model for a time series automatically.

Since the 1990s, feature-based model selection has been applied to time series forecasting. For instance, Prudêncio and Ludermir [8] used decision tree to select between two models to forecast stationary time series. To calculate the optimal model from four statistical forecast methods, Wang, *et al.* [9] proposed rule induction based on a decision tree that incorporated thirteen time series features. Lemke and Bogdan [10] employed general statistical features, frequency domain features, auto-correlations, and diversity features as input; base forecast models, including statistical methods, intelligent models, and their combination, were used as candidate models, and five learners were used to study the NN3 and NN5 datasets. Scholz-Reiter *et al.* [11] predicted customer demands for production planning by using six base forecast models as candidates, and by constructing a knowledge base based on a decision tree with twenty-six features, comprised of fourteen common measures and twelve RQA measures. Kück, *et al.* [12] considered landmarking as features in an empirical study on NN3 data. Talagala, *et al.* [13] proposed a model selection method based on Random Forest (RF) and thirty-six features to identify the optimal model for each time series on the M3 and M1 datasets. Ali, *et al.* [14] used three classifiers including feed-forward neural network, decision tree and support vector machine to investigate the situations in which the use of additional data can the improve performance of a meta-learning system. Matijaš, *et al.* [15] applied decision tree to select forecast model for load multivariate time series. Additionally, Adya and Lusk [16] applied model selection to an expert system to forecast complex time series and help decision-making. In order to reduce the redundancy of features, feature reduction was used before classification. Widodo and Budi [17] constructed historical database with feature selection methods (sequential floating forward selection) and verified the effectiveness of feature reduction with M1 data. Ali, *et al.* [14] exploited Random Forest based feature scoring approach to reduce features.

To the best of our knowledge, forecast models perform differently at each forecast horizon; hence horizon is an important factor in choosing the optimal forecast model. However, this factor is seldom considered in previous studies. Moreover, the datasets used in previous studies were mainly M3, NN3, and NN5, which contain few agricultural time series. Therefore, there is still a research gap in constructing a model selection framework for forecasting agricultural commodity prices.

In this study, we propose a model selection framework which involves both time series features and forecast horizons for forecasting agricultural commodity prices. Within this framework, twenty-nine features are extracted according to the periodicity, nonlinearity, and complexity of agricultural commodity price time series. Intelligent forecast models (i.e., ANN, SVR, and ELM) are specified as the candidate models. The relationships between these features and the performances of the candidate models are learned by classifiers,

which include RF and SVM. Feature reduction (the minimum redundancy and maximum relevance method) is also utilized to reduce feature redundancy and improve the forecast accuracy of the model selection framework. We test the effectiveness of considering the forecast horizon as the input feature and apply the feature reduction strategy to improve the performance of the classifier. Finally, we use principal component analysis to analyze the relationship between different commodities and the corresponding optimal forecast models.

The main contributions of this study are as follows. (a) We propose a model selection framework for forecasting agricultural commodity price time series based on time series features and forecast horizons. (b) We verify that the minimum redundancy and maximum relevance method can effectively reduce the redundancies between the features and is a workable approach to improving the performance of the classifier.

The remainder of this paper is structured as follows. The experimental framework, including feature extraction, feature selection, time series forecasting, and classification are presented in Section II. Section III describes the data, experimental design, parameter settings, and evaluation criteria. The experiment results are analyzed in Section IV, and Section V concludes.

II. EXPERIMENT FRAMEWORK AND METHODS
A. MODEL SELECTION
Meta-learning has been employed for algorithm recommendation tasks for some time and, since 2004, it has also been investigated in the area of time series forecasting [8]. In this special case of meta-learning, the aspect of interest is the relationship between data features and algorithm performance [32]; a classifier is usually applied to learn that relationship. The experimental framework for model selection using meta-learning for agricultural commodity price time series forecasting is shown in Figure 1. Three main steps are involved in this research; namely, feature extraction, feature selection, and classification.

In Step 1, twenty-nine time series features are extracted, including complexity features, linearity features, and stationarity features. The optimal forecast model for the time series is specified by comparing the forecast errors of the three candidate models at each horizon. Hence, both horizon information (horizon features) and the optimal model for the corresponding horizon will be recorded in the classification sample.

In Step 2, feature reduction is performed using an MRMR approach, with the aim of reducing feature redundancy and improving the generalization capability of the classifier. The ranking of the Mutual Information (MI) values of all the features will be obtained by the MRMR algorithm, and the ultimate features selected will be generated by the backward search method.

In Step 3, the classifiers proposed in the study are constructed by two popular machine learning approaches;
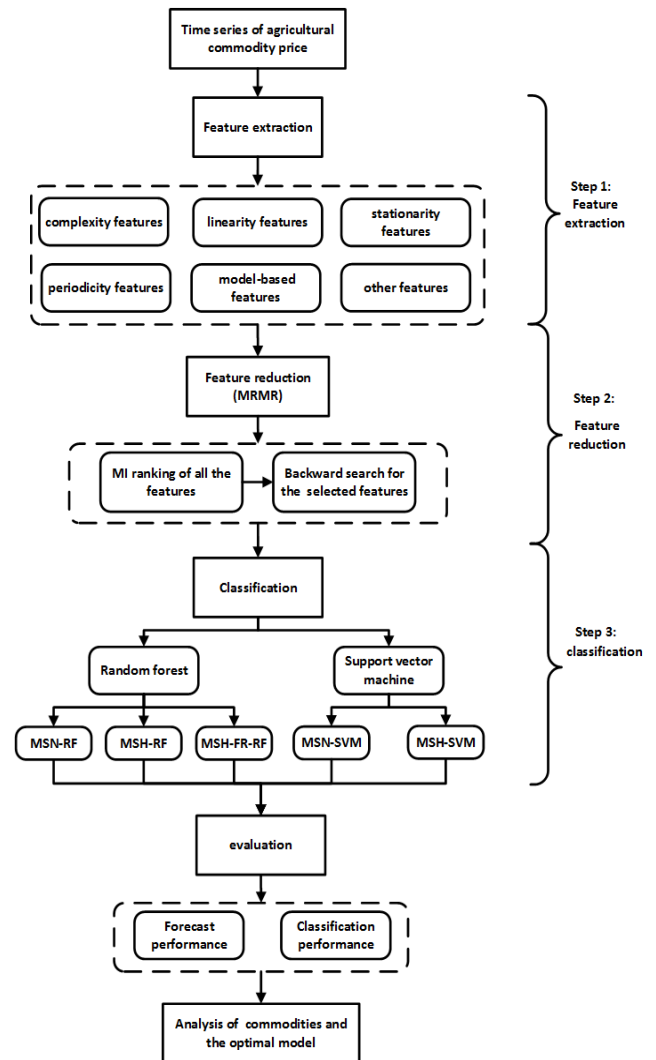

FIGURE 1. The overall experiment framework of model selection for agricultural commodity prices forecasting.

i.e., SVM and RF. Additionally, there are different schemes for developing the model selection framework, which involve a naïve classifier (abbreviated as MSN), a classifier with forecast horizon features (abbreviated as MSH), and a classifier with the reduced features (abbreviated as MSH-FR). Therefore, we have a total of five competing classifiers in this study; i.e., MSN-SVM, MSN-RF, MSH-SVM, MSH-RF, and MSH-FR-RF. Details of these classifiers (including the reason for excluding MSH-FR-SVM) are provided in Section III.

The forecast performance of the model selection framework is subsequently evaluated by two criteria; i.e., the mean absolute percent error (MAPE) and the improvement ratio (IR). The classification performance is estimated by classification accuracy (ACC). Finally, principal component analysis is applied to analyze the relationship between commodities and the optimal forecast model. Details of the analysis are provided in Section IV.

B. FEATURE EXTRACTION

Due to the complexity, nonlinearity and periodicity of an agricultural commodity price time series [30], [33], [34], the framework utilizes several categories of features to represent a time series, as shown in Table 2. Most of the features are selected according to [13] and implemented in 'tsFeatures' package in R. The implications of the selected features are shown as follows.

TABLE 2. Features used in this study.

	Feature	Description
Complexity features		
1	entropy	spectral entropy
2	hurst	hurst exponent
Linearity features		
3	nonlinearity	nonlinearity
4	curvature	curvature
5	linearity	linearity
Stationarity features		
6	stability	stability
7	lumpiness	lumpiness
8	unitroot_pp	the test stationarity based on Phillips-Person test
9	unitroot_kpss	the test stationarity based on KPSS test
Periodicity features		
10	seasonality	the strength of seasonality
11	x_acf1	the first ACF value of the original series
12	x_acf5	the sum of squares of the first 5 ACF values of original series
13	x_pacf5	the sum of squares of the first 5 PACF values of original series
14	e_acf1	the first ACF value of the remainder series
15	e_acf10	the sum of squares of the first 10 ACF values of remainder series
16	diff1_acf1	the first ACF value of the differenced series
17	diff1_acf5	the sum of squares of the first 5 ACF values of differenced series
18	diff2_acf1	the first ACF value of the twice-differenced series
19	diff2_acf5	the sum of squares of the first 5 ACF values of twice-differenced series
20	diff1x_pacf5	the sum of squares of the first 5 PACF values of differenced series
21	diff2x_pacf5	the sum of squares of the first 5 PACF values of twice-differenced series
22	seas_pacf	the partial autocorrelation coefficient at the first seasonal lag
23	seas_acf1	the autocorrelation coefficient at the first seasonal lag
Model-based features		
24	alpha	the level of suitable exponential smoothing model
25	beta	the slope of suitable exponential smoothing model
Other features		
26	spike	spikiness
27	trend	the strength of trend
28	peak	the number of peaks of the original series
29	trough	the number of troughs of the original series
Horizon features		
30	h1	mark 1 when forecast horizon =1 otherwise, mark 0
31	h3	mark 1 when forecast horizon =3 otherwise, mark 0
32	h6	mark 1 when forecast horizon =6 otherwise, mark 0
33	h12	mark 1 when forecast horizon =12 otherwise, mark 0

1) Complexity features quantify chaos and measure the long-range dependence in a time series.
2) Linearity features are important to determine the selection of models.
3) Stationarity features measure the stationarity of a time series.
4) Periodicity features provide indications on periodicity and seasonality of time series.
5) Model-based features, which characterize a time series by fitting a forecast model, are the parameters in the exponential smoothing model.
6) In other features, peak and trough capture oscillating behavior of time series. Spikiness captures the oscillating behavior of the residue of a time series by STL. Trend features characterize a time series by its degree of trend.
7) Horizon features are four binary numbers related to forecast horizons. They are marks for the corresponding optimal models at four forecast horizons.

C. FEATURE REDUCTION

Some of the features mentioned above may capture similar information on a time series, thus creating redundancies. For example, acf_1 and acf_5 capture similar information when the first five ACF values are small. These redundancies will increase the complexity of the classifiers and decrease their generalization capability [35]. Hence, feature reduction should be applied to reduce these redundancies in order to improve the model selection performance. In this study, the minimum redundancy and maximum relevance method (MRMR) is adopted for the feature reduction. Besides the redundancy between each pair of two features, this method also considers the correlation between features and class as a criterion for selecting a feature set.

For the correlation analysis, this study adopts Mutual Information (MI), which can measure both linear and nonlinear correlations between various features [36]. MRMR with MI measure was proposed by [37] and has been applied to many feature reduction tasks [38], [39]. The main goal of this method is to identify feature sets with maximum relevance to class and with minimum redundancy within each feature. The mutual information and the objective function are described in (1) and (2). This method has been realized by the 'mRMRe' function in R.

$$\mathrm{I}(x; y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \tag{1}$$

$$\max_{x_j \in X - S_{m-1}} \left[I\left(x_j; c\right) - \frac{1}{m - 1} \sum_{x_i \in S_{m-1}} I\left(x_j; x_i\right) \right] \tag{2}$$

In (1) and (2), x and y are two variables, S is the selected feature subset, m is the number of selected features, and $I(x; y)$ is the mutual information between x and y.

After obtaining the ranks of all 29 features from MRMR, a wrapper feature selection method called backward search [37] is used to figure out the optimal feature set.

The backward search tries to exclude one redundant feature at a time from the end of feature rank and estimate the classification accuracy to evaluate the feature subset. The features with the highest classification accuracy are the optimal features.

D. FORECAST MODEL

Due to the complexity and nonlinearity features of an agricultural commodity price time series, three workable and widely used AI models in agricultural commodity price forecasting are considered as the forecast models in this paper: artificial neural network (ANN); support vector regression (SVR); and extreme learning machine (ELM). The details are as follows.

ANNs are data-driven flexible models which are capable of approximating a large class of nonlinear problems [40]. One of the classic neural networks is the back-propagation neural network (BPNN), which includes feedforward and back-propagation. It is well known for its error learning algorithm in adjusting weights and bias. In general, a BPNN with a single hidden layer can generate the desired accuracy for a time series forecasting application [41].

SVR is originally proposed by Vapnik and based on the structured risk minimization principle [42]. It performs non-linear mappings through the application of kernels, which include nonlinear and linear kernels. It has been applied to forecast complex time series in industry [41], agriculture [43] and aviation [44].

ELM is a single hidden layer feedforward neural networks proposed by [45]. Unlike traditional learning algorithms in feedforward neural network, where parameters are tuned iteratively, the Moore-Penrose generalized inverse is applied to determine the output weights in ELM [6], thus requiring little time for training. This advantage has been applied to classification tasks and regression tasks in numerous studies [27], [46], [47].

E. CLASSIFICATION MODEL

In the current study, model selection is a classification problem with the goal of selecting the candidate forecast model with the lowest MAPE on the test set of a given time series. Two widely used classifiers are employed: Random Forest (RF) and Support Vector Machine (SVM).

Random forest is an ensemble model based on decision trees [48]. An ensemble trees with random feature selection and bagging prevents the model from overfitting and provides more accurate results in prediction [49]. Hence RF has been widely used in a number of studies [50]–[52].

SVM can map nonlinear data into a higher dimension level via a kernel function in order to classify data more accurately. In its early days, SVM was only used for binary classification tasks. In 2011, however, Chang *et al.* [53] employed a one-against-all strategy in SVM for solving multiclass classification problems. Since then, it is applied to a wide range of multiclass machine learning tasks [42], [50].

Due to the difference of classification principle in these two classifiers, it can verify the universality of the proposed model selection framework.

III. EXPERIMENT

A. DATA

All the datasets used here were available on http://www.caaa.cn/market. The research sample includes 522 monthly agricultural commodity price series, covering the commodities of piglet, hog, beef, and so on, as listed in Table 3 with the corresponding quantities.

TABLE 3. Quantity of the agricultural commodity price series.

Commodity	Quantity	Commodity	Quantity
piglet	29	hog	29
pork	29	beef	29
wheat bran	29	pig feed	29
layer chick	27	egg	29
corn	29	bean	29
pig grain	19	fish meal	28
mutton	29	live chicken	28
broiler feed	29	plain chicken	27
broiler	26	layer feed	29
wool	10		

B. EXPERIMENTAL DESIGN

The model selection framework used in this study is shown in Figure 2. First, the features of the time series are used as the inputs to the classifier, and the optimal model is treated as the label of the classifier. Both input and output make up the samples for classification. Second, all of the samples are divided into a training set and test set. The classifier is then trained by the training set. Finally, the test set is used as the input to the classifier, and the trained classifier will determine the selected model.

Two experiments were conducted to verify the effectiveness of the proposed method within the overall research framework. The purpose of the Experiment I was to test whether the forecast horizon feature can improve the performance of the classifier. Based on the results of Experiment I, Experiment II investigated further the feature reduction performance, which aims to reduce the redundancies between the features.

In these two experiments, simple model averaging (SMA) is considered as a benchmark to verify the effectiveness of reducing the risk of model selection. The predicted value of SMA is the average of the predictions of three candidate forecast models.

C. EXPERIMENT I: VERIFICATION OF THE HORIZON FEATURES PERFORMANCE

To verify the horizon feature performance, two different model selection frameworks were constructed: Model Selection Naïve (MSN) and Model Selection with Horizons (MSH). MSN is the baseline of the model selection framework in this study. In MSN, the time series features are the only input to the classifier. For each time series, the average forecast error (aErr) is calculated for each candidate model across four different forecast horizons; i.e., one-step, three-steps, six-steps, and twelve-steps-ahead. Comparing the aErrs
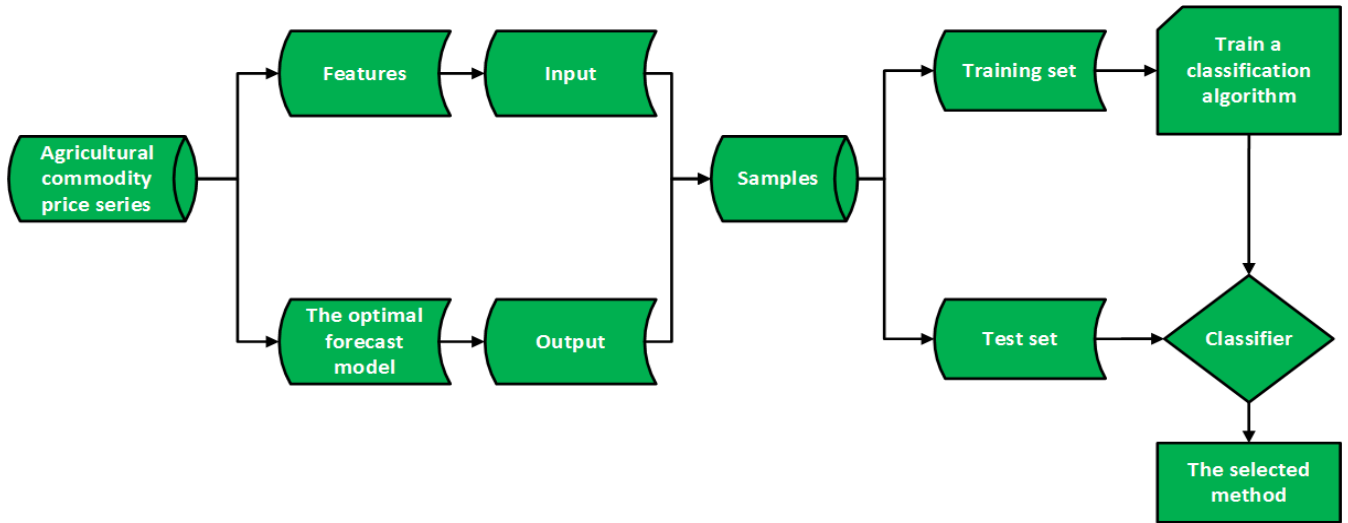
FIGURE 2. The model selection framework.

of all three candidate models, the one with the lowest forecast error is specified as the output label.

MSH is a model selection framework which contains the features of different forecast horizons. For each time series, the optimal model is specified by comparing the forecast errors of the candidate models at each horizon, as shown in Figure 3(b). That is to say, we have a total of four optimal models according to four different forecast horizons. Therefore, the features of both the time series and the forecast horizons are used as inputs to the classifier, whereas the optimal model of the corresponding horizon is regarded as the output label.
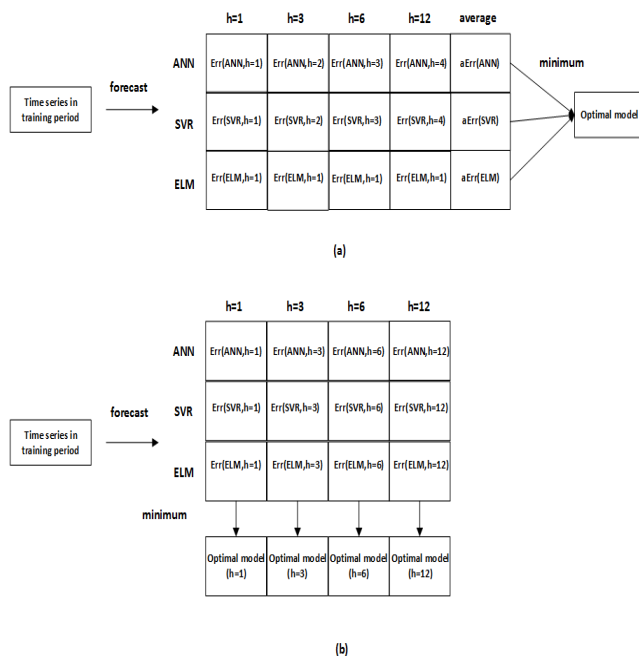

FIGURE 3. Specification of the optimal models for (a) MSN and (b) MSH.

D. EXPERIMENT II: VERIFICATION OF THE FEATURE REDUCTION PERFORMANCE

To evaluate the feature reduction performance, a model selection framework with horizon features and feature reduction (abbreviated as MSH-FR) was developed as a competing model for MSH. The main difference between MSH-FR and MSH is that the input features of MSH-FR have been reduced, whereas MSH uses the original time series and horizon features.

Both RF and SVM are employed in MSN and MSH as classifiers except for MSH-FR. The reason of exclusion is that the training of MSN and MSH produced a much better performance from RF than SVM; thus only RF was used as the classifier in MSH-FR. This means that a total of five classifiers were constructed in this study: MSN-RF, MSH-RF, MSH-FR-RF, MSN-SVM, and MSH-SVM.

E. PARAMETER SETTINGS

Parameters can make effect on the model performance. In this study, the artificial intelligent models are executed within a certain parameter range, which is shown as followed.

ANN, SVR, and ELM were used as the candidate forecast models in this study. A single hidden-layer neural network was applied by using the 'nnet' function of the 'nnet' package in R. This is an automatic function that uses a quasi-Newton method for optimization purposes; specifically, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. The neural network parameters were tuned within a maximum of 300 iterations with twice time-series-average-period neural units in the hidden layer and a weight decay of 0.05. The optimal pair parameters of SVR with RBF kernel were found within the range $[10^{-3}, 10^{+3}]$ by means of a grid search using the 'tune.svm' function of the 'e1071' package. The ELM was tuned within a number of units between 1 and 50 in the hidden layer using the 'elm_train'

function of the 'elmNNRcpp' package. As for the candidate forecast models, each time series was divided into a fitted period (80%) and a forecast period (20%). The fitted period was used to fit the candidate forecast models whereas the forecast period was used to identify the optimal model from the range of candidates. Each candidate model ran ten times in order to obtain an average performance.

As for classification learners, Random Forest (RF) was employed using the 'randomForest' package. As for random forest, the number of trees is set to 1000 and the number of randomly seleted features is set to be one third of the total number of features available [13]. In order to search the optimal cost and gamma for the SVM classifier, a grid-search with the range $[10^{-3}, 10^{+3}]$ and a ten-cross-validation were used in the training set. Both RF and SVM are trained with a five-fold cross-validation method and ran twenty times.

F. EVALUATE CRITERION

In this study, two criteria were used for evaluating the prediction accuracy: Mean Absolute Percent Error (MAPE) and Improvement Ratio (IR). Classification accuracy (ACC) was used to estimate the classification performance. MAPE is a popular accuracy measure in the forecast community, the definition is shown as follows:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i - \widehat{Y_i}}{Y_i} \right| \times 100\% \quad (3)$$

In (3), N is the number of observations in the testing period, $\widehat{Y_i}$ is the predicted value, and Y_i is the corresponding actual value.

IR is a percentage comparison that measures the improvement of model A compared to model B, as follows:

$$\text{IR}_{MAPE} = \frac{MAPE_B - MAPE_A}{MAPE_B} \times 100\% \quad (4)$$

In (4), $MAPE_B$ is the average MAPE of the predicted value forecast by model B and $MAPE_A$ is the average MAPE of the predicted value forecast by model A.

The classification accuracy is defined as follows.

$$\text{ACC} = \frac{N_c}{N_t} \quad (5)$$

In (5), N_c is the number of correct classified instances and N_t is the number of total instances.

IV. RESULTS

A. FEATURE DESCRIPTION AND ANALYSIS

Statistical descriptions of all the features are listed in Table 4. These statistical values indicate that the features have different magnitudes; thus normalization should be employed before classification.

The correlation diagram based on mutual information(MI) is shown in Figure 4. The dark point at the top right-hand corner represents the maximum MI value of all the twenty-nine features. The greater the correlation, the deeper the color.

TABLE 4. Statistical description of features of time series.

Feature	Mean	Standard deviation	Minimum	Maximum
entropy	0.5166	0.0957	0.4119	0.9278
trend	0.9541	0.0671	0.4798	0.9996
spike	<0.0001	<0.0001	<0.0001	0.0006
linearity	11.9004	3.3175	-5.5354	14.5444
curvature	-1.2945	2.5414	-7.3815	9.9994
e_acf1	0.5132	0.2032	-0.1416	0.8261
e_acf10	0.4758	0.3017	0.0097	1.1692
seasonality	0.2865	0.1564	0.0651	0.7934
peak	6.2414	3.5058	1.0000	12.0000
trough	5.4502	2.9288	1.0000	12.0000
lumpiness	0.0465	0.3184	<0.0001	6.2433
stability	0.9519	0.1177	0.3253	1.0567
hurst	0.9978	0.0130	0.7774	1.0000
unitroot_kpss	3.4473	1.0446	0.1328	4.4699
unitroot_pp	-7.1267	11.7038	-150.8649	0.9937
nonlinearity	0.5168	0.8131	0.0006	6.0814
x_acf1	0.9607	0.0598	0.3962	0.9940
x_acf5	4.1279	0.7502	0.6259	4.7992
diff1_acf1	0.0687	0.2885	-0.5432	0.6454
diff1_acf5	0.1384	0.1258	0.0018	0.9829
diff2_acf1	-0.4133	0.1779	-0.7716	0.0658
diff2_acf5	0.2601	0.1400	0.0593	0.9266
seas_acf1	0.7122	0.2070	-0.2120	0.9091
x_pacf5	0.9940	0.1088	0.2805	1.2651
diff1x_pacf5	0.1386	0.1128	0.0018	0.5100
diff2x_pacf5	0.4705	0.2228	0.0932	1.1279
seas_pacf	-0.0604	0.0874	-0.3274	0.1467
alpha	0.8819	0.1851	0.0793	0.9999
beta	0.0651	0.1801	0.0001	0.9999

Figure 4(a) shows the correlation among features before feature reduction. It can be seen that most of the correlations are light colored, which reveals that these features contain diverse information on the time series. A few points are dark colored, which implies that the information contained in these features is redundant. These redundancies may have negative effects on the generalization performance of the classifiers. Thus feature reduction should be employed to eliminate the redundancies.

Figure 4(b) shows the correlation of the time series features after feature reduction. Compared to Figure 4(a), the numbers of the dark colored points have been reduced. This result shows that MRMR approach is a workable approach to feature reduction.

After feature reduction, twenty-five features including twenty-one time series features and four horizon features remained. In general, the average MI of each pair of two features has been reduced by 7.45%. The details of the selected features are listed in Table 5. Four horizon features have been retained, which demonstrates that the forecast horizon features are important for the performance of the classifier.

B. PERFORMANCE OF MODEL SELECTION

The model selection experiments for forecasting agricultural commodity prices were conducted using the research design described above. Accordingly, the forecast performances of all the candidate models and the model selection frameworks were evaluated using the two accuracy mea-
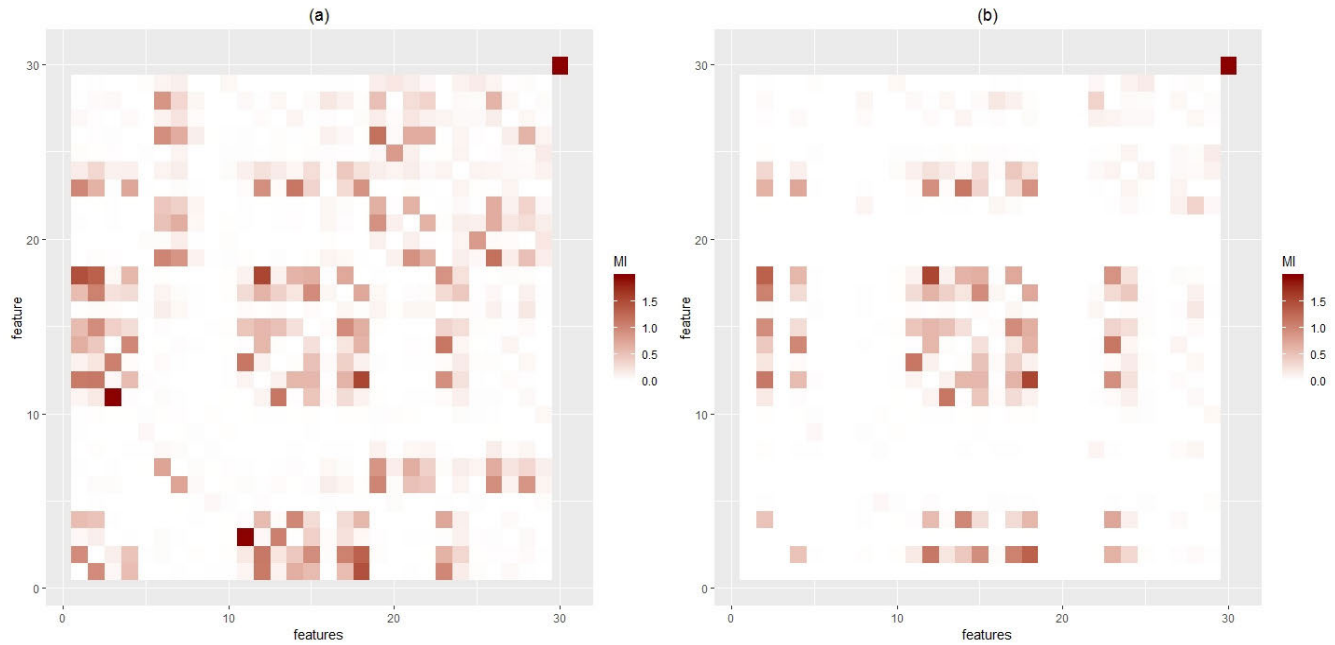
FIGURE 4. Correlation matrix plots based on MI (a) before and (b) after feature reduction.

TABLE 5. The reserved features after feature reduction.

Category	Features
Periodicity features	x_acf5, seas_pacf, x_acf1, diff1x_pacf5, x_pacf5, seas_pacf, seasonality, diff2_acf5
Stability features	unitroot_pp, stability, unitroot_kpss, lumpiness
Linearity features	linearity, curvature, nonlinearity
Complexity features	hurst
Model-based features	alpha, beta
Horizon features	h1, h3, h6, h12
Other features	trend, peak, trough

TABLE 6. Forecast performance of the MSN in terms of MAPE.

	h=1	h=3	h=6	h=12	average
ANN	3.4988	7.7901	10.5496	12.8591	8.6744
SVR	4.1948	10.0278	12.9542	14.6353	10.4530
ELM	4.2058	8.1399	10.6749	14.6674	9.4220
SMA	3.6617	8.0175	10.4539	12.6093	8.6856
MSN-RF	3.7114	7.8813	10.3578	12.7189	**8.6673**
MSN-SVM	3.7969	7.8901	10.3106	13.0068	8.7511

TABLE 7. Forecast performance of the MSH and MSH-FR in terms of MAPE.

	h=1	h=3	h=6	h=12	average
ANN	3.4985	7.7772	10.5583	12.8583	8.6731
SVR	4.1918	10.0144	12.9646	14.6332	10.451
ELM	4.2061	8.1293	10.6854	14.6664	9.4218
SMA	3.6610	8.0038	10.4634	12.6104	8.6847
MSH- RF	3.4901	7.4729	9.8113	12.6254	8.3499
MSH- SVM	3.5738	7.7456	10.1239	12.5543	8.4994
MSH-FR-RF	3.4877	7.4733	9.8148	12.5562	**8.3330**

sures MAPE and IR, and the classification performance was estimated using ACC. Table 6 and Table 7 show the forecast performances in terms of MAPE. The last column labeled "average" shows the average performances of the models across all four forecast horizons. In order to illustrate intuitively the advantage of the model selection framework, we compare the performance of each selection framework to the optimal single model ANN. The results are shown in Table 8. Table 9 shows the classification performances of the three model selection frameworks in terms of ACC.

C. RESULTS OF EXPERIMENT I

Comparing the three single forecast models (ANN, SVR, and ELM), it can be seen from Table 6 and Table 7 that ANN is the most powerful as it has the smallest average forecast error. Consequently, ANN is specified as the optimal single model in our study. Therefore, in order to verify the effectiveness of

the model selection framework, the performances of the forecast models selected by different classifiers will be compared with that of ANN.

As for the two classifiers used in this study (RF and SVM), we find that the forecast error of SVM is on average larger than that of RF in all the cases, as can be seen in Table 6 and Table 7, which indicates that RF is much more suitable for this classification task. Moreover, the classification accuracy of RF is higher than that of SVM in both MSN and MSH, as shown in Table 8. The reason may be that RF has advantages over SVM when dealing with unbalanced data. The repeated random sub-sampling in RF has been
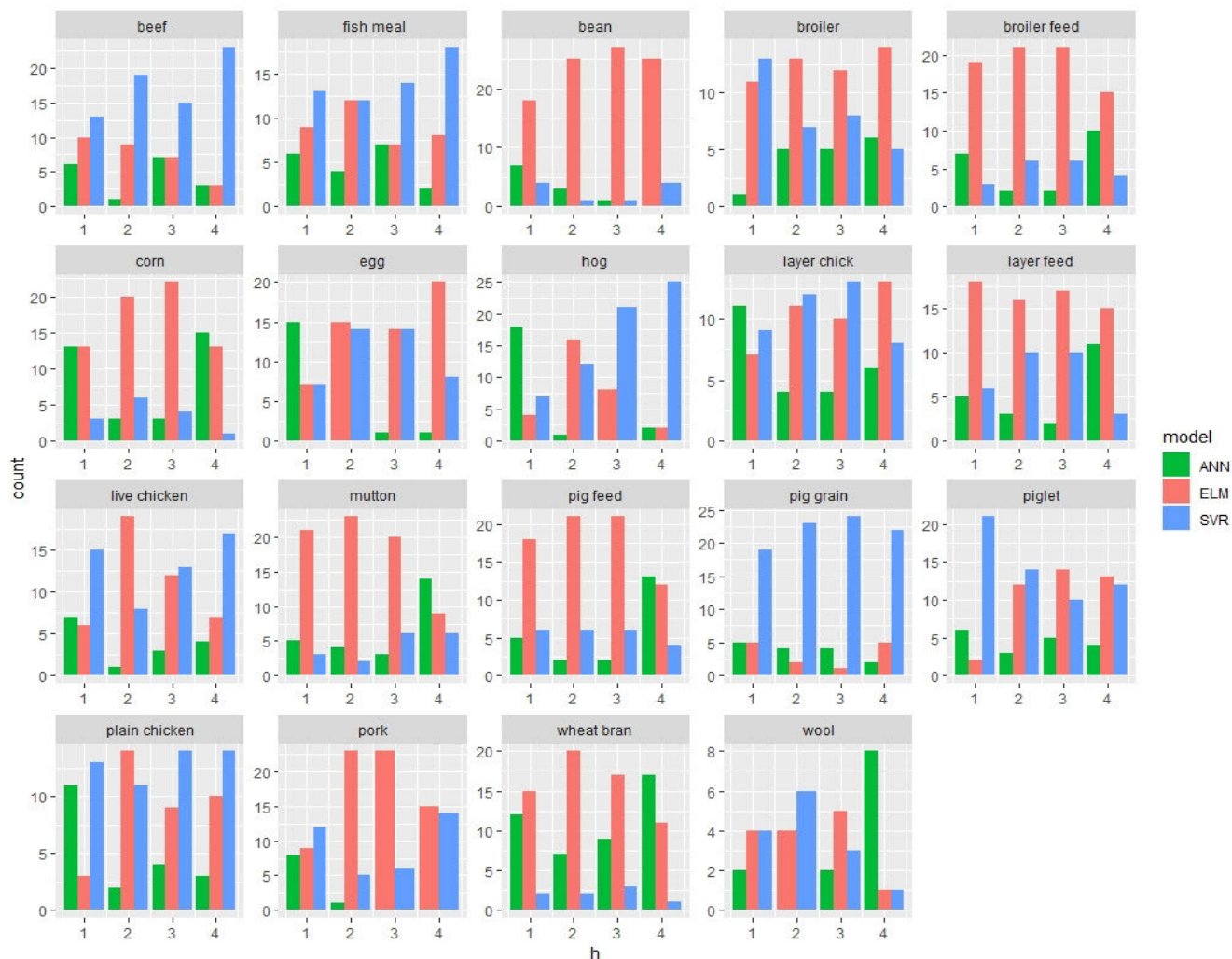
FIGURE 5. Facet plots of the optimal model selected at different horizons for different commodities.

TABLE 8. Forecast performance of MSN, MSH and MSH-FR in terms of IR.

		h=1	h=3	h=6	h=12	average
MSN	RF	-6.0749	-1.1702	1.8187	1.0905	0.0818
	SVM	-8.5193	-1.2833	2.2655	-1.1482	-1.1482
MSH	RF	0.2423	3.9126	7.0746	1.8111	3.7259
	SVM	-2.1518	0.4066	4.1141	2.3644	2.0026
MSH-FR	RF	0.3089	3.9078	7.0417	2.3499	**3.9212**

TABLE 9. Classification performance of the MSN, MSH and MSH-FR in terms of ACC.

	RF	SVM
MSN	55.78%	53.90%
MSH	61.39%	56.49%
MSH-FR	61.85%	——

found to be very effective in dealing with an imbalanced dataset [54] whereas SVM assumes that the class distribution in the dataset is uniform [55].

Focusing on the model selection framework, Table 6 shows that the average forecast error of MSN-RF is 8.6673

compared to 8.6744 for ANN. This result demonstrates the superiority of the model selection framework, which can reduce effectively the risk in model selection, thus yielding a smaller forecast error. Regarding the two strategies used for improving the performance of MSN, Table 7 shows the performance of MSH and MSH-FR. Both MSH-RF and MSH-SVM perform well across four forecast horizons compared to ANN. This may indicate that the performance of MSH is better than that of MSN. As for MSH-RF, the average forecast error is 8.3499, yielding a smaller forecast error compared with MSH-SVM. It can be seen from Table 8 that the average IR of MSH-RF is 3.7259, which is greater than that of MSN. Moreover, it can also be seen from Table 8 that the classification accuracy of MSH-RF is higher than that of MSN. These results verify the superiority of using different forecast horizons as the input features of the classifier. This method can not only improve the forecast accuracy of model selection by using the data on forecast model performance at different forecast horizons, but can also improve the classification performance of the model selection.
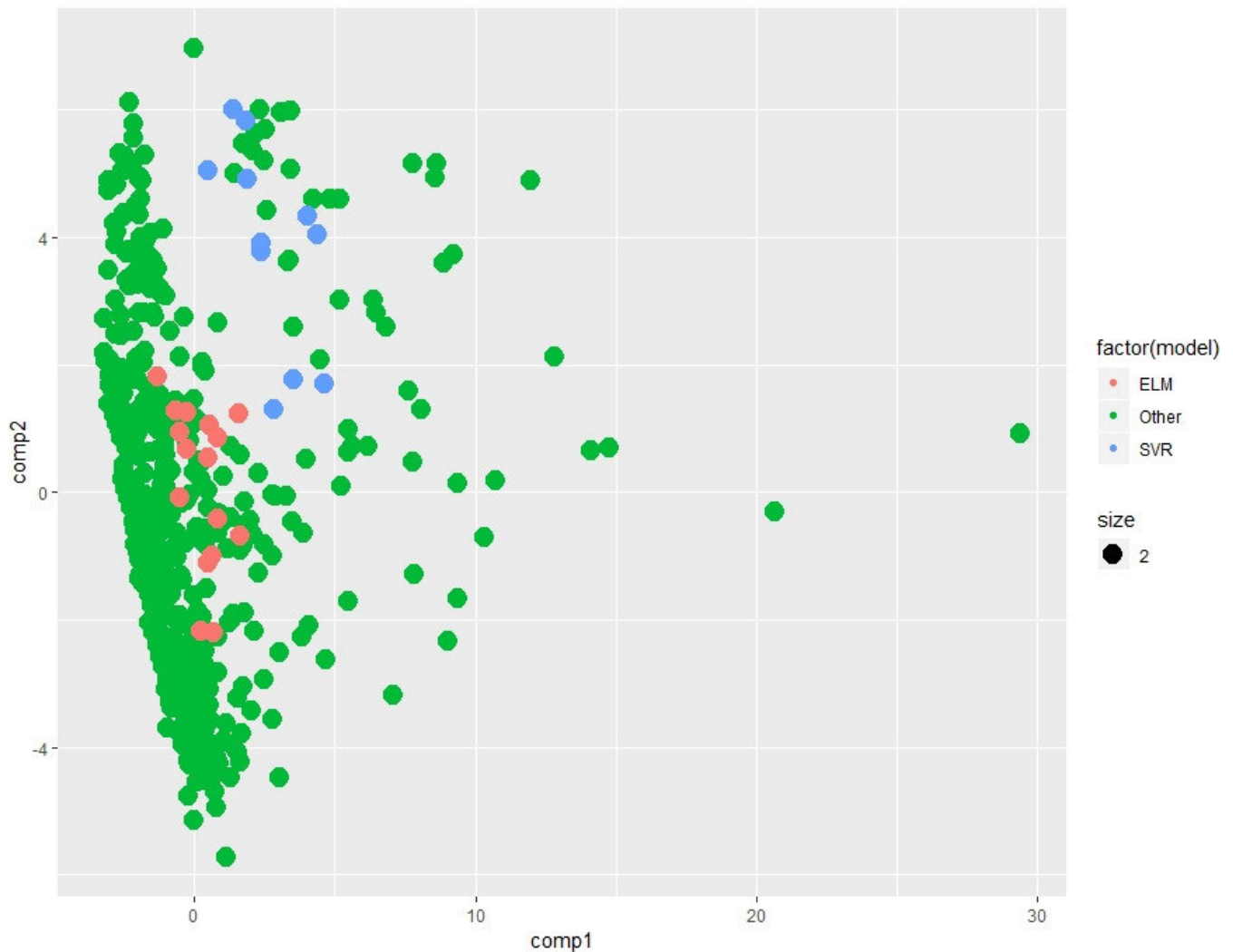
FIGURE 6. Distribution of different commodities in PCA space.

It can be seen from Table 6 that the average MAPE of SMA is 8.6856, which is only on average larger than the optimal candidate model (ANN, 8.6744). That is to say, SMA can avoid performing the worst result of forecasting and reduce the risk of model selection. Compared to SMA, MAPEs of MSN-RF and MSN-SVM are lower at $h = 3$ and $h = 6$, which indicates that the model selection framework is competitive for SMA. It can also be seen from Table 7 that the average MPAE of SMA is 8.6847 which is only larger than ANN. The MAPEs of MSH-RF and MSH-SVM are almost lower than SMA at each forecast step. It demonstrates the superiority of the model selection framework, which is more effective than SMA in reducing the risk of model selection.

D. RESULTS OF EXPERIMENT II

Based on the performance of MSH-RF, we employ a feature reduction strategy to further improve the performance of the classifier. It can be seen that the average forecast error of MSH-FR-RF is 8.3330, which is the optimal score among all the classifiers. The reason may be that feature reduction is

effective in removing the redundant features of a time series, and thus improves the performance of the classifier. It can be seen in Figure 4 that features with high correlations are minimized after feature reduction, which again demonstrates the effectiveness of feature reduction. The average IR of MSH-FR-RF is 3.9212, which is the best score of all the classifiers. Moreover, the ACC of MSH-FR-RF is 61.85%, which is also the best score for classification accuracy. Thus the effectiveness of feature reduction is fully verified in this study.

E. RELATIONS BETWEEN OPTIMAL MODEL AND AGRICULTURAL COMMODITY PRICE SERIES

In order to determine the relationships between the selected forecast model and various agricultural commodities, we plot a facet wrap as Figure 5 shows. The 1,2,3 and 4 of the *x*-axis refer to the four different forecast horizons, i. e. one-step, three-steps, six-steps and twelve-steps-ahead. The value of the *y*-axis indicates the number of times that the forecast model is selected. Green, red and blue bars represent the three

candidate forecast models (ANN, SVR and ELM). It can be seen that the optimal model for a certain category varies for different forecast horizons, and the superior model is also different for different categories.

As for different categories, Figure 5 shows that the red bar is prominent for the bean facet wrap while the blue bar is prominent for the pig grain facet wrap. That is to say, ELM is the optimal model for most of the bean price time series wherea SVR is the optimal model for most of the pig grain price time series. The reason might be that these two categories have different features which lead to the different model selection results.

In order to verify this assumption, we perform a principal component analysis (PCA), following the method proposed by Kang [56]. The first two principal components of the bean and pig grain price time series are plotted into a feature space as shown in Figure 6. The *x*-axis refers to the first principal component and the *y*-axis refers to the second principal component. The red points represent the bean price time series which take ELM as the optimal model across all the forecast horizons. The blue points represent the pig grain price time series which identifies SVR as the optimal model across all the forecast horizons. It can be seen that the zone of red points is separated from the zoo of blue points. This phenomenon indicates that the features of those two categories are quite different from each other. Therefore, different distributions of the time series features can be regarded as the main reason for the different model selection results.

V. CONCLUSION

In this paper, we proposed a model selection framework for forecasting agricultural commodity prices using both time series features and forecast horizons. Generally, three main steps were involved in the proposed model selection framework, i.e. feature extraction, feature reduction and classification. Firstly, we extracted twenty-nine time series features of agricultural commodity prices. Secondly, we used the minimum redundancy and maximum relevance method to reduce feature redundancy and improve the performance of the model selection framework. Finally, five classifiers were constructed to verify the performances of different model selection strategies. Additionally, the relation between different commodities and the optimal model was evaluated by principal component analysis. Relative to existing studies, this study verifies the effectiveness of the model selection framework in choosing the most suitable forecasting models. With agricultural commodity price series as research samples, several interesting conclusions can be made based on the empirical results. Firstly, considering the forecast horizon as one of the features can improve the performance of both classification and forecast, which demonstrates the forecast horizon should be considered as an important factor in model selection task. Secondly, MRMR can further improve the performance of the model selection framework, which indicates a workable feature reduction method should be exploited in model selection for increasing the generalization capability of classifiers. Thirdly, different distributions of time series features may lead to different optimal forecast models. It verifies the necessity of model selection based on time series features.

This study makes contributions to the literature by (1) proposing a model selection framework for agricultural commodity price prediction, by involving forecast horizon as one of the features; (2) using a workable feature reduction method to reduce the redundancy of features and improve the classification performance; and (3) discovering that different distributions of time series features may lead to different model selection results.

The experimental results demonstrate the superiority of the proposed model selection framework in terms of prediction accuracy. In particular, the proposed method almost outperforms all of the candidate models and verifies its effectiveness in model selection. Second, the proposed method outperforms the simple model averaging method and further shows that the method is effective in reducing the risk in model selection for a new time series. Third, the use of forecast horizon features and the approach to feature reduction are both feasible methods for further improving the performance of the model selection framework. The analysis shows that the optimal model for a certain category varies for different forecast horizons, and the optimal model is also different for different categories. The main reason is that the different distributions of time series features lead to different model selection results.

The proposed model selection framework could be improved from the following perspectives. First, the proposed method could be employed as an effective model selection tool for other forecast objects. Second, some powerful classifiers such as AdaBoost and Bayesian networks could be utilized to further improve the classification capability. Third, this study only considers three popular forecast models in the area of forecasting agricultural commodity prices; however, other techniques could also be introduced to make the framework more workable.

REFERENCES

[1] A. Darekar and A. A. Reddy, "Cotton price forecasting in major producing states," *Econ. Affairs*, vol. 62, no. 3, p. 373, Nov. 2017.

[2] Y. Q. Xu, Y. H. Wei, and X. G. Li, "Establishment of agricultural products,price prediction," *Statist. Decis.*, vol. 480, no. 12, pp. 75–77, 2017.

[3] E. A. Evans and S. Nalampang, "Forecasting price trends in the US avocado (Persea Americana Mill.) market," *J. Food Distrib. Res.*, vol. 40, no. 2, pp. 37–46, 2011.

[4] M. H. Wei, Q. L. Zhou, Z. J. Yang, and J. G. Zheng, "Prediction model of agricultural product's price based on the improved bp neural network," in *Proc. Int. Conf. Comput. Sci. Educ.*, 2012, pp. 613–617.

[5] T. Xiong, C. Li, and Y. Bao, "Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China," *Neurocomputing*, vol. 275, pp. 2831–2844, Jan. 2018.

[6] Y. Liu, Q. Duan, D. Wang, Z. Zhang, and C. Liu, "Prediction for hog prices based on similar sub-series search and support vector regression," *Comput. Electron. Agricult.*, vol. 157, pp. 581–588, Feb. 2019.

[7] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.

[8] R. B. Prudêncio and T. B. Ludermir, "Meta-learning approaches to selecting time series models," *Neurocomputing*, vol. 61, pp. 121–137, Oct. 2004.

[9] X. Wang, K. Smith-Miles, and R. Hyndman, "Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series," *Neurocomputing*, vol. 72, nos. 10–12, pp. 2581–2594, Jun. 2009.

[10] C. Lemke and B. Gabrys, "Meta-learning for time series forecasting and forecast combination," *Neurocomputing*, vol. 73, nos. 10–12, pp. 2006–2016, Jun. 2010.

[11] B. Scholz-Reiter, M. Kück, and D. Lappe, "Prediction of customer demands for production planning–Automated selection and configuration of suitable prediction methods," *CIRP Ann.*, vol. 63, no. 1, pp. 417–420, 2014.

[12] M. Kück, S. F. Crone, and M. Freitag, "Meta-learning with neural networks and landmarking for forecasting model selection an empirical evaluation of different feature sets applied to industry data," in *Proc. Int. Joint Conf. Neural Netw.*, 2016, pp. 1499–1506.

[13] T. S. Talagala, R. J. Hyndman, and G. Athanasopoulos. (2018). *Meta-Learning How to Forecast Time Series*. [Online]. Available: https://www.monash.edu/business/ebs/research/publications/ebs/wp06-2018.pdf

[14] A. R. Ali, B. Gabrys, and M. Budka, "Cross-domain meta-learning for time-series forecasting," *Procedia Comput. Sci.*, vol. 126, pp. 9–18, 2018.

[15] M. Matijaš, J. A. Suykens, and S. Krajcar, "Load forecasting using a multivariate meta-learning system," *Expert Syst. Appl.*, vol. 40, no. 11, pp. 4427–4437, Sep. 2013.

[16] M. Adya and E. J. Lusk, "Development and validation of a rule-based time series complexity scoring technique to support design of adaptive forecasting DSS," *Decis. Support Syst.*, vol. 83, pp. 70–82, Mar. 2016.

[17] A. Widodo and I. Budi, "Feature enhancement for model selection in time series forecasting," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst.*, 2013, pp. 367–373, doi: 10.1109/icacsis.2013.6761603.

[18] Y. Xu, S. Shen, and Z. Chen, "Research on forecast of sugar price based on improved neural network," in *Proc. 2nd Int. Symp. Intell. Inf. Technol. Secur. Inform.*, 2009, pp. 23–25.

[19] G. K. Jha and K. Sinha, "Time-delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India," *Neural Comput. Appl.*, vol. 24, nos. 3–4, pp. 563–571, Mar. 2014.

[20] J. H. Zhang, F. T. Kong, J. Z. Wu, M. S. Zhu, K. Xu, and J. J. Liu, "Tomato prices time series prediction model based on wavelet neural network," *Appl. Mech. Mater.*, vols. 644–650, pp. 2636–2640, Sep. 2014.

[21] T. Xiong, C. Li, Y. Bao, Z. Hu, and L. Zhang, "A combination method for interval forecasting of agricultural commodity futures prices," *Knowl.-Based Syst.*, vol. 77, pp. 92–102, Mar. 2015.

[22] K. Ayankoya, A. P. Calitz, and J. H. Greyling, "Real-time grain commodities price predictions in South Africa: A big data and neural networks approach," *Agrekon*, vol. 55, no. 4, pp. 483–508, Oct. 2016.

[23] C. M. Cai, L. W. Ling, C. Niu, and D. B. Zhang, "Integration prediction of domestic pork market price based on empirical mode decomposition and support vector machine," *Chin. J. Manage. Sci.*, vol. 24, pp. 845–851, 2016.

[24] Q. Duan, L. Zhang, F. Wei, X. Xiao, and L. Wang, "Forecasting model and validation for aquatic product based on time series GA-SVR," *Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 1, pp. 308–314, Jan. 2017.

[25] P. F. He, J. Li, and D. Q. Zhang, "Predicting chinese soybean price based on APSO-SVR," *Soybean Sci.*, vol. 36, no. 4, pp. 632–638, 2017.

[26] L. Zhemin, X. Shiwei, and C. Liguo, "Prediction study based on dynamic chaotic neural network: Taking potato timeseries prices as an example," *Syst. Eng.-Theory Pract.*, vol. 35, no. 8, pp. 2083–2091, 2015.

[27] J. Wang, C. Qi, and M. F. Li, "Prediction of commodity prices based on SSA-ELM," *Syst. Eng.-Theory Pract.*, vol. 37, no. 8, pp. 2004–2014, 2017.

[28] Y. T. Zhou, F. W. Jia, and K. Wen, "Study on short-term wheat price prediction in China," *Price, Theory Pract.*, no. 3, pp. 107–110, Jul. 2018.

[29] D. Zhang, G. Zang, J. Li, K. Ma, and H. Liu, "Prediction of soybean price in China using QR-RBF neural network model," *Comput. Electron. Agricult.*, vol. 154, pp. 10–17, Nov. 2018.

[30] W. Zheng, W. C. Qiang, and L. W. De, "Forecasting model of agricultural products prices based on seasonal index adjustment and HGWO-SVR algorithm," *Statist. Decis.*, vol. 34, no. 19, pp. 33–36, 2018.

[31] P. Wu and Z. M. Li, "Forecast researchon China's pork price based on the ARIMA-GM-RBF hybrid model," *Price, Theory Pract.*, no. 1, pp. 75–78, Apr. 2019.

[32] C. Lemke, M. Budka, and B. Gabrys, "Metalearning: A survey of trends and technologies," *Artif. Intell. Rev.*, vol. 44, no. 1, pp. 117–130, Jun. 2015.

[33] C. Cui, W. M. Li, and S. Z. Zong, "Analysis and forecast of long memory characteristic of chinese livestock product price," *J. Huazhong Agricult. Univ., Social Sci. Ed.*, no. 2, pp. 1–7, 2017.

[34] Z. Y. Dong, D. Chen, and Y. H. Liang, "The fluctuation and causes of agricultural product price–taking beijing as an example," *Technoeconomics Manage. Res.*, vol. 267, no. 10, pp. 124–127, 2018.

[35] X. Q. Zeng and G. Z. Li, "A supervised solution for redundant feature detection depending on instances," in *Proc. IEEE Int. Conf. Bioinf. Biomed. Workshops*, Oct. 2012, pp. 299–306.

[36] C. S. Hong and B. J. Kim, "Mutual information and redundancy for categorical data," *Stat. Papers*, vol. 52, no. 1, pp. 17–31, Feb. 2011.

[37] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[38] A. Pattanateepapon, W. Suwansantisuk, and P. Kumhom, "Feature selection based on correlations of gene-expression values for cancer prediction," in *Proc. Biomed. Eng. Sci.*, 2016, pp. 716–721.

[39] Y. Fan, X. Li, and P. Zhang, "Real-Time static voltage stability assessment in large-scale power systems based on maximum-relevance minimum-redundancy ensemble approach," *IEEE Access*, vol. 5, pp. 27281–27291, 2017.

[40] S. Panigrahi and H. Behera, "A hybrid ETS–ANN model for time series forecasting," *Eng. Appl. Artif. Intell.*, vol. 66, pp. 49–59, Nov. 2017.

[41] J. Wang, X. Li, T. Hong, and S. Wang, "A semi-heterogeneous approach to combining crude oil price forecasts," *Inf. Sci.*, vols. 460–461, pp. 279–292, Sep. 2018.

[42] Y. Zhang, S. Wei, L. Zhang, and C. Liu, "Comparing the performance of random forest, SVM and their variants for ecg quality assessment combined with nonlinear features," *J. Med. Biol. Eng.*, vol. 39, no. 3, pp. 381–392, Jun. 2019.

[43] S. M. Guzmán, J. O. Paz, M. L. M. Tagert, A. E. Mercer, and J. W. Pote, "An integrated SVR and crop model to estimate the impacts of irrigation on daily groundwater levels," *Agricult. Syst.*, vol. 159, pp. 248–259, Jan. 2018.

[44] D. S. De O. Santos Júnior, J. F. De Oliveira, and P. S. De Mattos Neto, "An intelligent hybridization of ARIMA with machine learning models for time series forecasting," *Knowl.-Based Syst.*, vol. 175, pp. 72–86, Jul. 2019.

[45] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[46] S. Chai, Z. Xu, and Y. Jia, "Conditional density forecast of electricity price based on ensemble ELM and logistic EMOS," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3031–3043, May 2019.

[47] A. Naz, M. Javed, N. Javaid, T. Saba, M. Alhussein, and K. Aurangzeb, "Short-term electric load and price forecasting using enhanced extreme learning machine optimization in smart grids," *Energies*, vol. 12, no. 5, p. 866, Mar. 2019.

[48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[49] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst. Appl.*, vol. 134, pp. 93–101, Nov. 2019.

[50] Z. Qu, Q. Liu, and C. Liu, "Classification of congestive heart failure with different New York Heart Association functional classes based on heart rate variability indices and machine learning," *Expert Syst.*, vol. 36, no. 3, Jun. 2019, Art. no. e12396.

[51] F. B. De Santana, W. Borges Neto, and R. J. Poppi, "Random forest as one-class classifier and infrared spectroscopy for food adulteration detection," *Food Chem.*, vol. 293, pp. 323–332, Sep. 2019.

[52] B. He, T. Ji, H. Zhang, Y. Zhu, R. Shu, W. Zhao, and K. H. Wang, "MRI-based radiomics signature for tumor grading of rectal carcinoma using random forest model," *Cellular Physiol.*, vol. 234, no. 11, pp. 20501–20509, 2019.

[53] C.-C. Chang, L.-J. Chien, and Y.-J. Lee, "A novel framework for multiclass classification via ternary smooth support vector machine," *Pattern Recognit.*, vol. 44, no. 6, pp. 1235–1244, Jun. 2011.

[54] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC Med. Inform. Decis. Making*, vol. 11, no. 1, p. 51, 2011.

[55] L. V. Utkin, "An imprecise extension of SVM-based machine learning models," *Neurocomputing*, vol. 331, pp. 18–32, Feb. 2019.

[56] Y. Kang, R. J. Hyndman, and K. Smith-Miles, "Visualising forecasting algorithm performance using time series instance spaces," *Int. J. Forecasting*, vol. 33, no. 2, pp. 345–358, Apr. 2017.

DABIN ZHANG received the Ph.D. degree in computer software and theory from Wuhan University, Wuhan, China, in 2007. He is currently a Professor with the College of Mathematics and Informatics, South China Agriculture University (SCAU), Guangzhou, China. He has published four books and over 80 journal articles. His research interests include big data analysis, data mining, artificial intelligence, and forecast modeling.

LIWEN LING is currently pursuing the Ph.D. degree with South China Agriculture University (SCAU). She is also a Lecturer with the College of Mathematics and Informatics, SCAU, Guangzhou, China. Her main research interests include agricultural market analysis, price forecasting, and forecast combination.

SHANYING CHEN is currently pursuing the master's degree with the College of Mathematics and Informatics, South China Agriculture University (SCAU), Guangzhou, China. Her main research interests include model selection and price forecasting.

QIANG XIA received the Ph.D. degree from the Renmin University of China, Beijing, China, in 2015. He is currently a Professor with the College of Mathematics and Informatics, South China Agricultural University (SCAU), Guangzhou, China. He has published two books and over 15 journal articles. His research interests include time series analysis, Bayesian statistics, and financial statistics.

● ● ●