Forecasting of Grain Pile Temperature from Meteorological Factors Using Machine Learning

Shanshan Duan, Weidong Yang Member, IEEE, Xuyu Wang, Member, IEEE, Shiwen Mao, Fellow, IEEE, and Yuan Zhang

Abstract—Food storage security is critical to the national economy and people's lives. The environmental parameters of a granary should be accurately monitored in order to provide a better preservation environment for food storage. In this paper, we use temperature sensors to measure and collect grain temperature data for a period of 423 days from a real world granary, and collect the corresponding meteorological data from China Meteorological Data Network. We propose to leverage meteorological data to predict the average temperature of the grain pile with machine learning algorithms: a support vector regression (SVR) approach and an adaptive boosting (AdaBoost) approach. We incorporate different kernel functions in the SVR model and choose the appropriate base-estimator and the number of estimators in the AdaBoost model. We then analyze the correlation between a large amount of historical data from the granary and the corresponding meteorological forecast data based on the Pearson correlation coefficient. We find that there are strong correlations between some meteorological factors. In order to eliminate redundant information, we reduce the dimension of data by principal components analysis (PCA), and compare the prediction models before and after PCA dimension reduction. The results show that the proposed approaches can achieve a high accuracy and the Adboost method after PCA dimension reduction achieves the best performance.

Index Terms—Food storage; Temperature sensors; Meteorological metrics; Support Vector Regression; Adaptive Boosting; Machine learning.

I. INTRODUCTION

The demand for food will be doubled by 2050 as population and social mobility increase [1]–[6]. Globally, more than 2 billion tons of food are harvested each year [7]. However, up to one third of the annual total global production of grain is lost because of poor post-harvest management. Lack of control over grain moisture content, high temperature, and insect infestation are the three most significant factors causing the loss. In fact, high grain moisture and temperature can provide favorable conditions for hot spot development, mold growth, and insect infestation [8]. Grain is still a physiologically active organism during storage, and is affected by the physical and biological environment. These internal and external factors are closely related to the safe storage of grain. With the development of science and technology, grain storage technology has been improved and food security has been better guaranteed. However, there are still many risk factors in the process of grain storage. The complex grain storage ecosystem is under the joint influence of the environment sub-ecosystem and the granary protection construction [9].

Grain temperature is an important indicator of grain conditions. Its detection and control technology are critical for the operation of grain warehouses (or, granary). In the entire grain detection system, detection of the temperature of stored grain is a relatively mature technology and has been widely used in national reserves [10]. In fact, the storage temperature is highly predictable when aggregating over thousands of granaries and storage parameters. Different from the developed countries, the present situation of grain storage in China is unique [8]. Recently, Yang et al. present a non-destructive and economic wheat moisture detection system with commodity WiFi, which can achieve high classification accuracy for both LOS and NLOS scenarios [11], [12].

To improve the accuracy of grain pile temperature forecasting, we focus on the issue of using the National Meteorological Information Center (NMIC) meteorological forecast to accurately predict grain pile temperature. In this paper, we first discuss the temperate measurement system for food storage. For a period of 423 days, we used temperature sensors to measure and collect of grain temperature data from the granary storage at the No. 1 Warehouse in the Xishan District of Kunming, Yunnan province, China. We also collect the corresponding meteorological data from China Meteorological Data Network. We provide an analysis of the correlation between a large amount of historical data from the granary and the corresponding meteorological data. We find that the surface temperature of grain pile has higher correlations with air temperature, relative humidity, and Ocm ground temperature, but a smaller correlation with air pressure. We propose to predict surface temperature of grain pile using multiple meteorological factors, aiming to achieve high predication accuracy.

In particular, we develop a support vector regression (SVR) approach [13]–[16] and an adaptive boosting (AdaBoost) approach to predict surface temperature of grain pile using

DOI: 10.1109/ACCESS.2019.2940266.
multiple meteorological factors. In fact, because of some outliers recorded by temperature sensors, the raw data cannot be directly employed for the prediction model. We implement outlier detection and removal to delete bad data samples, and apply data normalization to all the sampled data to guarantee that the meteorological data and the surface temperature of grain pile have the same unit. Then, we leverage the calibrated meteorological data and grain temperature data to train the prediction model. Finally, we incorporate different kernel functions in the SVR model, and choose the appropriate base-estimator and the number of estimators in the AdaBoost model to predict the temperature of grain pile using meteorological data. We then analyze the correlation between a large amount of historical data from the granary and the corresponding meteorological forecast data based on the Pearson correlation coefficient. It is found that there is a strong correlation between some meteorological factors. In order to eliminate redundant information, we reduce the dimension of data by principal component analysis (PCA). We compare the prediction accuracy using different kernels in SVR model (such as the linear kernel function, the polynomial kernel function, and the Gaussian radial basis function (RBF) kernel) [17], [18] and Adaboost model with random forest regressor as base-estimator before and after PCA dimension reduction.

The main contributions of this paper are summarized below.

- To the best of our knowledge, this is the first work to use meteorological metrics to predict the average temperature of grain pile with machine learning.
- We employ temperature sensors to measure the grain temperature data from a real world grain storage for a period of 423 days, and collect meteorological data for the same region and time period. Then, we analyze the correlation between a large amount of historical granary data and the corresponding meteorological data based on the Pearson correlation coefficient.
- We implement outlier detection and data normalization for the raw meteorological and grain pile temperature data. We use different kernel functions with the SVR model and Adaboost model with random forest regressor to predict the average temperature of grain pile based on meteorological data. We compare the accuracy of grain surface temperature prediction using different kernel functions of SVR model and Adaboost model before and after PCA dimension reduction. The results show that the Adaboost method with random forest regressor as base-estimator after principal components analysis achieves the best performance.

The remainder of this paper is organized as follows. Section II discusses the related work. The granary temperature measurement system is presented in Section III. Section IV describes grain temperature data measurement, collection process, and data preprocessing, and Section V discusses the SVR model and the Adaboost model. Section VI validates the performance of the proposed method using real world data. Section VII summarizes this paper.
recognition model based on depth confidence network was established by using depth learning method to predict grain heap temperature [36]. Existing data-driven solutions do not take into account the characteristics of temperature gradient with time series. For temperature, increasing or decreasing temperature is a time-dependent gradient process, and its trend is very important for temperature prediction. The existing models have insufficient ability to deal with the time series correlation of temperature data, which leads to defects in temperature prediction. Many high-precision mathematical models and improved measurement systems are proposed to improve temperature monitoring and food storage management capabilities [37]–[39]. However, the work of grain temperature forecasting has been focused on time series models, which does not consider the effect of external meteorological factors. In fact, meteorological factors have been successfully utilized for accurate solar intensity forecasting [40], [41]. Our models can exploit meteorological factors for grain pile temperature forecasting, which is greatly different from the exiting works.

III. THE TEMPERATURE MEASUREMENT SYSTEM

To collect grain temperature data, we deploy a set of temperature sensors in the tall granary. Fig. 1 illustrates the tall flat granary architecture, which is divided into 10 rows from east to west, five regions from south to north, and four layers from top to bottom. Then 200 temperature sensors are deployed in this granary, and the sensors are encapsulated in cables and the cables are inserted into the grain pile at certain places. In the tall square granary, the temperature sensor layout principle is that the distance between the horizontal and horizontal temperature measuring cables should be no more than 5 m, the distance between the vertical cables should be no more than 2 m, and distance from the cables to the grain surface, granary bottom, and granary wall should within 0.3 m to 0.5 m.

Fig. 2 presents a cross-sectional view of the granary. The temperature monitoring system generally includes temperature sensors, temperature measuring cables, and a computer monitoring terminal. Each vertical line in the figure represents a cable, and four temperature sensors are encapsulated in each cable. To monitor the abnormal change of temperature in the early stage of grain damage, the distance between the temperature measuring cables should be less than 0.5 m. Therefore, a large number of cables are needed, which would be hard to deploy (i.e., buried in the gain pile at precise locations) and lead to high measurement cost. Fig. 3 is the structural diagram of the grain condition measurement and control system in the granary. The computer sends test commands to the extension, receives test data from the extension, and then processes the receiving data. The extension receives the computer command, detects temperature data, and sends the results to the computer. The digital sensor is encapsulated inside the cable and laid inside the barn. Both digital temperature sensors and humidity sensors use a wire bus communication protocol to report sensory data.

The inspection time of grain temperature is preferably from 9 am to 10 am every day, when the temperature is close to the daily average temperature. While checking the temperature of the grain, we should also check the temperature inside the granary and the temperature outside the granary for analysis and comparison. All data is sampled once a day.

IV. DATA COLLECTING AND PREPROCESSING

A. Data Collecting

We measured and collected the grain temperature data of the grain storage in the No. 1 Warehouse at Xishan District of Kunming, Yunnan, China for a period of 423 days since January 1,
Then we downloaded the meteorological data of the corresponding region for the corresponding period of time from China Meteorological Data Network (http://data.cma.cn/). The meteorological metrics include air pressure, air temperature, relative humidity, precipitation, evaporation, wind speed, sunshine duration and 0 cm ground surface temperature. The temperature samples of grain pile are recorded by deployed sensors at 9 am every day. Then, we measure how the surface temperature of the grain pile changes with the meteorological variables and how these variables are influenced by each other.

Fig. 4 presents the surface temperature of the grain pile and the air temperature of the period. We find that the surface temperature of the grain pile and the air temperature are positively correlated. In other words, the temperature of the first layer of the stored grain pile becomes higher or smaller as the air temperature increases or decreases. However, there are other factors that also contribute to the surface temperature of the grain pile reading, since the surface temperature of the grain pile has been delayed for several months compared to the air temperature. It is noticed the highest air temperature was in June, while the highest temperature of the first layer of the grain pile was in September.

In Fig. 5, we can see that air temperature (subplot (a)), 0 cm ground surface temperature (subplot (b)), and relative humidity (subplot (c)) are all positively correlated with the surface temperature of the grain pile, especially at higher values. If the air temperature, 0 cm ground surface temperature, or relative humidity become larger, the surface temperature of the grain pile will likely increase too.

To study the correlation between the average temperature of the first layer of the stored grain pile and the meteorological metrics, we compute the Pearson correlation coefficients between pair of the factors. Table I provides the Pearson product moment correlation coefficients for all the meteorological variables and the surface temperature of grain pile. The higher the absolute value of the correlation coefficient, the higher the correlation between the two parameters. From Table I, we find that the surface temperature of grain pile has higher correlations with air temperature, relative humidity, 0 cm ground temperature, but with a smaller correlation with air pressure. Based on this study, we develop an SVR algorithm and an Adaboost algorithm to predict the surface temperature of grain pile using multiple meteorological parameters, which is discussed in the following section.
In this section, SVR with different kernels and Adaboost with random forest regressor as the base-estimator are utilized to predict the surface temperature of grain storage.

### B. Data Preprocessing

Due to some outliers recorded by the sensors, the data units are inconsistent; thus the raw data cannot be directly used by the prediction models. Therefore, the data must be processed before prediction. The data processing module includes outlier detection and data normalization, which are discussed in the following.

- **Outlier detection**: Some abnormal values are reported by temperature sensors. Outlier detection is used to recognize bad data values, which should be removed from the raw data. In this paper, we leverage the Pauta criterion method and the *linear trend at point* method to get rid of outliers. The outlier detection method is as follows.

**Step 1**: Let \( X_i, i = 1, 2, ..., n \), be the \( i \)th value of meteorological metrics or the average temperature of the first layer of grain pile. We calculate the arithmetic mean value as
\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{1}
\]

**Step 2**: We then obtain the residual \( e_i \) as in (2) and the standard deviation \( \sigma \) of the meteorological metrics or the average temperature of the first layer of grain pile as in (3).
\[
e_i = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2. \tag{2}
\]
\[
\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}. \tag{3}
\]

**Step 3**: For all \( X_i, i = 1, 2, ..., n \), if \(|e_i| > 3\sigma\), we consider \( X_i \) as an abnormal value and replace it with the arithmetic mean value \( \bar{X} \).

**Step 4**: Repeat the above three steps till all the \( X_i \)'s are processed.

- **Data normalization**: To guarantee that the meteorological data and the surface temperature of grain pile have the same unit, we choose the zero-mean normalization method to normalize all sampled data. The normalized value \( Z_i \) is computed as
\[
Z_i = \frac{1}{\sigma} (X_i - \bar{X}), \quad i = 1, 2, ..., n. \tag{4}
\]

After data preprocessing, the training samples are
\[
T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}, \tag{5}
\]
where \( x_i \) is a vector of eight calibrated meteorological metrics in the \( i \)th sample, and \( y_i \) represents the calibrated average surface temperature of grain pile in the \( i \)th sample.

### V. Forecasting Models for Grain Pile Surface Temperature

#### A. Support Vector Regression (SVR)

The SVR model is utilized to learn a function \( f(x) \), which is close to the gain surface temperature \( y \) as much as possible [13]. The function is defined by
\[
f(x) = w^T \cdot \phi(x) + b, \tag{6}
\]
where \( w \) and \( b \) are the parameters to be determined, and \( \phi(\cdot) \) is a generic function. A deviation \( \varepsilon \) is used to evaluate the loss between the output \( f(x) \) of the model and the true grain surface temperature \( y \). In other words, when \(|f(x) - y| < \varepsilon\), the prediction result can be considered to be accurate.

The SVR problem can be formulated as follows.
\[
\min_{w, b} \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^{n} l_\varepsilon(f(x_i) - y_i), \tag{7}
\]
where \( C \) is a regularization constant and \( l_\varepsilon(\cdot) \) is an insensitive loss function of \( \varepsilon \). Adding a slack variable to the loss metric, problem in (7) can be transformed into a minimization problem (8) as follows.
\[
\min_{w, b, \xi, \hat{\xi}} \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^{n} (\xi_i + \hat{\xi}_i), \tag{8}
\]
s.t.
\[
f(x_i) - y_i \leq \xi_i + \varepsilon, \quad i = 1, 2, ..., n \tag{9}
\]
\[
y_i - f(x_i) \leq \hat{\xi}_i + \varepsilon, \quad i = 1, 2, ..., n \tag{10}
\]
\[
\xi_i \geq 0, \hat{\xi}_i \geq 0, \quad i = 1, 2, ..., n, \tag{11}
\]
where \( \xi_i \) and \( \hat{\xi}_i \) are slack variables. To solve problem (8), we first obtain the following Lagrange function using the Lagrange multiplier method, defined as follows.
Then, we compute the maximum sample error on the training set:

\[
L(w, b, \alpha, \hat{\alpha}, \xi, \hat{\xi}, \mu, \hat{\mu}) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \left( \xi_i + \hat{\xi}_i \right) - \sum_{i=1}^{n} \mu_i \cdot \xi_i - \sum_{i=1}^{n} \hat{\mu}_i \cdot \hat{\xi}_i + \sum_{i=1}^{n} \alpha_i \cdot (f(x_i) - y_i - \varepsilon - \hat{\xi}_i) + \sum_{i=1}^{n} \hat{\alpha}_i \cdot (y_i - f(x_i) - \varepsilon - \xi_i),
\]

where \(\mu, \hat{\mu}, \alpha,\) and \(\hat{\alpha}\) are Lagrange multipliers. Applying the duality theory, the average temperature and humidity of the grain surface can be estimated as

\[
f(x) = \sum_{i=1}^{n} (\alpha_i - \hat{\alpha}_i) \cdot \phi(x_i) + b.
\]

### B. Adaptive Boosting Regression (AdaBoost)

In addition to SVM, we also consider AdaBoost model for grain pile temperature forecasting from meteorological factors, which is one of the most famous algorithms in the boosting family. The working mechanism of boosting algorithm is as follows: firstly, a basic learner is trained from the initial training set and the distribution of training samples is adjusted according to the performance of the basic learner, and then the next basic learner is trained based on the adjusted sample distribution. The above steps are repeated until the number of basic learners reaches the predetermined value. All the basic learners are combined according to the combination strategy to obtain the final strong learner. Basic learners are weak learners which are only better than random guesses, such as a simple decision tree. The reason for using weak learners instead of strong learners is that it is often much easier to find a weak learner than a strong learner. Boosting mechanism is to start from the weak learner and build a strong learner through repeated learning [42]–[44]. In particular, the core idea of AdaBoost is to select a base estimator, fit a series of weak estimators on a series of data, and according to the accuracy of these estimators, provide each estimator a weight, then multiply all estimators by their respective weights and add them together to obtain the final prediction. The basic principle of the AdaBoost algorithm is shown in Fig. 6.

The input of Adaboost model is the training samples \(T\), base learning algorithms \(L\) and the number of base learners \(M\). The process is as follows:

**Step 1:** The weight distribution of initial training samples is provided,

\[
D_1 = (w_{11}, \ldots, w_{1i}, \ldots, w_{1n}), \quad w_{1i} = \frac{1}{n}.
\]

**Step 2:** For iteration round \(t = 1, 2, \ldots, M\), we train a base learner using training data set with current distribution \(D_t\)

\[
h_t = L(D, D_t).
\]

Then, we compute the maximum sample error on the training set

\[
E_t = \max |y_i - h_t(x_i)|, \quad i = 1, 2, \ldots, n.
\]

We can then obtain the relative error of each sample \(e_{ti}\), i.e.,

\[
e_{ti} = \frac{|y_i - h_t(x_i)|}{E_t}.
\]

In addition, the regression error rate of the base learner \(h_t\) on training set is

\[
\varepsilon_t = \sum_{i=1}^{n} w_{ti} \cdot e_{ti}.
\]

The weight coefficient \(\alpha_t\) of the base learner \(h_t\) is

\[
\alpha_t = \frac{\varepsilon_t}{1 - \varepsilon_t}.
\]

Then, we update the sample distribution \(D_{t+1} = (w_{t+1,1}, \ldots, w_{t+1,i}, \ldots, w_{t+1,n})\) of training set, that is,

\[
w_{t+1,i} = \frac{w_{ti} \cdot \alpha_t^{1-e_{ti}}}{z_t},
\]

where \(z_t\) is the normalization factor, which is defined by

\[
z_t = \sum_{i=1}^{n} (w_{ti} \cdot \alpha_t^{1-e_{ti}}).
\]

**Step 3:** Finally, the basic learners are combined to obtain the final strong learner

\[
H(x) = \sum_{i=1}^{M} \ln \left( \frac{1}{\alpha_t} \right) \cdot g(x),
\]

where \(g(x)\) is the median of all \(\alpha_t \cdot h_t(x)\).

### VI. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed approach using real measurement data: the temperature data of grain storage at the No. 1 Warehouse in the Xishan District of Kunming, Yunnan, China for a period of 423 days. The meteorological data of the corresponding region and period are obtained from the China Meteorological Data Network (http://data.cma.cn/).

#### A. Performance Evaluation of SVR with Different Kernel Functions

Firstly, we leverage SVR to predict the surface temperature of grain piles based on meteorological metric data. Because of the poor thermal conductivity of the grain kernel itself and the thermal insulation of the silo wall of the granary, some samples cannot be linearly divided in the two dimensional space. Thus, we consider a kernel function to map the samples to a higher dimensional space, which can achieve a better separability performance. Under Python 2.7 environment, the training process of regression model is as follows: 8 meteorological factors are taken as independent variables, average temperature of grain stack surface as dependent variables, and independent variables and dependent variables are as input of SVR. The mapping relationship between X and Y is constructed by machine learning self-learning. In order to ensure that the same training set and test set are segmented in each run, the same random number seeds are set [13].
In this paper, the linear kernel function, the polynomial kernel function, and the Gaussian RBF kernel function are used to predict the average temperature of the first layer of the stored grain pile [45]. We discuss the detailed experimental results in the following.

The linear kernel function is defined as follows.

\[ k(x, x_i) = x \cdot x_i. \] (23)

We use a linear kernel for the SVR model, where the dimension of the feature space is the same as the input space. It requires fewer parameters and also achieves a faster computational speed. The collected meteorological data of 423 days and the corresponding average temperature of the first layer of grain pile are used as data samples. The temperature of the first layer of grain pile per day corresponds to eight metrics of meteorology at the same time. We randomly select 80% of the samples as the training set, and the remaining 20% of the samples as the test set. The results of predicting the average temperature of the first layer of grain pile using the linear kernel function are presented in Fig. 7.

The polynomial kernel function is defined as follows.

\[ k(x, x_i) = ((x \cdot x_i) + 1)^d, \] (24)

where \( d \) represents the order of the polynomial. In our experiment, we set \( d = 2 \), which achieves a good performance. Due to the ventilation and food turning operations during grain storage, the sample size collected within a time period is limited and the feature dimension is relatively small. The polynomial kernel function can map the low-dimensional input space to a high-dimensional feature space, but the corresponding computational complexity is higher. The results of predicting the average temperature of the first layer of grain pile obtained by the polynomial kernel function are shown in Fig. 8.
The Gaussian RBF kernel function is defined as follows.

\[
k(x, x_i) = \exp \left( -\frac{\|x - x_i\|^2}{\delta^2} \right),
\]

where \(\delta\) is the parameter of the Gaussian RBF. This is a locally strong kernel function that maps samples into a higher dimensional space. It can achieve a good performance for both large and small samples, and requires fewer parameters than the polynomial kernel function. The results of predicting the average temperature of the first layer of grain pile using the Gaussian RBF kernel function are shown in Fig. 9.

\section*{B. Performance Evaluation of Adaboost with Random Forest Regressor as Base-Estimator}

The advantage of Adaboost is that various regression models can be used to construct weak learners, and it is not easy to appear over-fitting phenomenon. Its disadvantage is that it is sensitive to abnormal samples. Abnormal samples may have larger weight in iteration, which affects the prediction accuracy of strong learners. Fortunately, we preprocess the abnormal samples using outlier detection and data normalization, which just overcomes the shortcomings of Adaboost. Adaboost algorithm itself is a lifting algorithm. It can use any learner as a base learner. Generally speaking, the most widely used weak Adaboost learners are decision tree and neural network. For decision tree, Adaboost classification uses CART classification tree, while Adaboost regression uses CART regression tree. According to the predictive performance of SVR in section VI-A, we first select the SVR as the base learner. Under the same number of basic learners and learning rate, the linear kernel function, polynomial kernel function and Gauss radial basis function of SVR are trained respectively. It is found that when the polynomial kernel function of SVR and the Gauss radial basis function are used as the base learners, the calculation time is too long to be applicable. Then we choose the linear kernel function of SVR and the random forest regressor as the base learner. The results show that the prediction error of using random forest regressor as the base learner is small. In addition, because the high accuracy and good anti-noise ability of random forest, it is used as the basic learner of AdaBoost regression in this paper. In order to prevent over-fitting, a regularization term \(v\) is usually added to the Adaboost algorithm, which is often referred as a learning rate. For the same training set, smaller \(v\) means more iterations of weak learners. The step size and the maximum number of iterations are usually used together to determine the fitting effect of the algorithm.

Under Python 2.7 environment, the training process of regression model is as follows: 8 meteorological factors are taken as independent variables, average temperature of grain stack surface as dependent variables, and independent variables and dependent variables are as input of Adaboost regression. In order to ensure that the same training set and test set are segmented in each run, the same random number seeds are set [13]. If the number of base learners \(M\) is too small, it is easy to cause under-fitting, but if it is too large, it will cause a large amount of calculation. When \(M\) reaches a certain number, the model upgrade obtained by increasing the value of \(M\) will be very small, so a moderate \(M\) value is generally chosen. In our Adaboost model, the number of base learners is 11, and the learning rate is 0.001. The collected meteorological data of 423 days and the corresponding average temperature of the first layer of grain pile are used as data samples. The temperature of the first layer of grain pile per day corresponds to eight metrics of meteorology at the same time. We randomly select 80% of the samples as the training set, and the remaining 20% of the samples as the test set. The results of predicting the average temperature of the first layer of grain pile using the Adaboost algorithm with random forest regressor as base estimator is shown in Fig. 10. We can notice that the proposed Adaboost algorithm has a better regression performance compared with SVR based methods.

\section*{C. Eliminating Redundant Information}

As we show in Table I, the correlation coefficients between some meteorological factors are relatively large, which indicates that there are strong correlations among these factors such as air temperature and ground temperature, relative
humidity and evaporation. In Fig. 11, we can see that air temperature and air pressure (subplot (a)), 0 cm ground surface temperature and air pressure (subplot (b)) show negative correlation with each other. 0 cm ground surface temperature and air temperature (subplot (c)) show strong positive correlation with each other. As a result, our regression models contain redundant information, which often decreases the prediction accuracy of each model. PCA is a popular method for removing redundant information from an input dataset, thereby reducing its dimensionality [46]–[48]. Thus, we use the PCA method to remove redundant information from our feature dataset. The PCA algorithm uses an orthogonal transformation to convert a set of potentially correlated input variables into a set of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables.

We choose the first seven (highest) eigenvalues and run the regression method on the reduced feature set. The results show that the prediction model is better than the full feature set when PCA is reduced to 7 dimensions. In Fig. 15, the Adaboost algorithm with random forest as base learner can achieve a minimum RMS-Error of 1.26 after PCA and of 1.79 before PCA dimensionality reduction, respectively. In Fig. 14, the RBF kernel can obtain a RMS-Error of 3.94 after PCA and of 4.45 before PCA dimensionality reduction, respectively. In Fig. 13, the polynomial kernel can have a RMS-Error of 4.48 after PCA and of 4.69 before PCA dimensionality reduction, respectively. In Fig. 12, the linear kernel can get a RMS-Error
of 4.873 after PCA and of 5.249 before PCA dimensionality reduction, respectively. We also ran experiments for reducing the dimensionality of the feature set from 7 to 6. However, we found that all three SVM regression techniques performed worse compared to the 7-dimensional feature set. The performance degradation is that the information of the additional reduction in dimensionality is not redundant.

D. Comparison

In order to quantitatively measure the prediction performance of the SVR method using different kernel functions and the Adaboost method with random forest regressor as base estimator, the root mean square error (RMSE) is used as the evaluation criteria. In addition, the execution time of the algorithm is taken as another criterion for evaluation. Using the different kernel functions in the SVR model and random forest regressor as base estimator in the Adaboost model to predict the average temperature of the first layer of grain pile, the RMSE results are presented in Fig. 16. It can be seen that all the three schemes are quite accurate, while Adaboost method with random forest regressor as base estimator achieves the smallest RMSE result. The polynomial kernel function achieves a 10.50% reduction over the linear kernel function. The Gaussian RBF kernel function achieves a reduction of 15.08% and 5.12% over the linear and polynomial kernel function, respectively. The Adaboost method with random forest regressor as base estimator achieves a reduction of 65%, 61%, and 59% over the linear kernel function, polynomial kernel function and Gaussian RBF kernel function in SVR, respectively. The Adaboost method with random forest regressor as base estimator achieves a reduction of 65%, 61%, and 59% over the linear kernel function, polynomial kernel function and Gaussian RBF kernel function in SVR, respectively. Although the complexity of the model has not changed, the number of features is reduced after PCA dimensionality reduction. As a result, the execution time is reduced and it reaches the minimum value when RBF kernel function is used. The RMSEs and the execution time of different prediction models are shown in Table II.

VII. CONCLUSIONS

In this paper, we leveraged an SVR approach and an Adaboost approach to predict the average temperature of the first layer of the stored grain pile using meteorological metrics. Due to the poor thermal conductivity of the grain kernel itself and the thermal insulation properties of the granary wall, the average temperature of the first layer is usually delayed by a certain amount of time than the outside air temperature. Among eight factors of meteorology, there are three factors, including air temperature, 0cm ground temperature, and relative humidity, that have a greater impact on the average temperature of the first layer. We used different kernel functions in the SVR model and chose the appropriate base-estimator and the number of estimators in the AdaBoost model. Grain temperature data measured from a real granary and the corresponding meteorological data were used in our study. The results demonstrated that the Adaboost method after PCA dimension reduction can achieve a minimum RMS-Error of 1.26, which is better than kernel based SVM methods.

REFERENCES

<table>
<thead>
<tr>
<th></th>
<th>Dimensions</th>
<th>SVR-linear</th>
<th>SVR-polynomial</th>
<th>SVR-RBF</th>
<th>Adaboost</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>Full Feature Set</td>
<td>5.249</td>
<td>4.69</td>
<td>4.45</td>
<td>1.79</td>
</tr>
<tr>
<td>RMSE</td>
<td>PCA 7 Dimensions</td>
<td>4.873</td>
<td>4.48</td>
<td>3.94</td>
<td>1.26</td>
</tr>
<tr>
<td>Execution Time (s)</td>
<td>Full Feature Set</td>
<td>4.91</td>
<td>2.63</td>
<td>0.023</td>
<td>1.73</td>
</tr>
<tr>
<td>Execution Time (s)</td>
<td>PCA 7 Dimensions</td>
<td>4.83</td>
<td>2.12</td>
<td>0.021</td>
<td>1.55</td>
</tr>
</tbody>
</table>


Shanshan Duan received the B.S degree in information engineering from Henan Institute of Science and Technology, Henan, China, in 2011, and the M.S. degree in signal processing from Henan University of Technology, Henan, China, in 2016, where she is currently pursuing the Ph.D. degree in grain information processing. Her current research interests include bio-photon signal processing, grain signal processing, deep learning and temperature field simulation, analysis and distribution.

Weidong Yang received his B.S. in industrial automation, and M.S. and Ph.D. degree in Computer Science from Xidian University in China in 1999, 2005, and 2008, respectively. He is now a professor in Henan University of Technology, deputy chair of Key Laboratory of Grain Information Processing and Control (Henan University of Technology) ministry of Education. He is also a senior member of China Computer Federation (CCF). His research focuses on wireless networks security, privacy protection, and vehicular ad hoc networks, and so on.

Xuyu Wang [S’13-M’19] received the M.S. in Signal and Information Processing in 2012 and B.S. in Electronic Information Engineering in 2009, both from Xidian University, Xi’an, China. He received a Ph.D. in Electrical and Computer Engineering from Auburn University, Auburn, AL, USA in Aug. 2018. He is an Assistant Professor in the Department of Computer Science, California State University, Sacramento, CA. His research interests include indoor localization, deep learning, and IoT. He received a Woltolsz Fellowship at Auburn University, and is a co-recipient of the Second Prize of Natural Scientific Award of Ministry of Education, China in 2013, the Best Demo Award of IEEE SECON 2017, the Best Student Paper Award of IEEE PIMRC 2017, and the IEEE ComSoc MMTC 2018 Best Journal Paper Award.

Yuan Zhang received his B.S. in automatic control from University of Science and Technology of China in 1983, and Ph.D. degree in signal processing from Nanjing University of Science and Technology in China in 1996. He is now a professor in Henan University of Technology, chair of Key Laboratory of Grain Information Processing and Control (Henan University of Technology) ministry of Education. His research focuses on intelligent signal processing, electromagnetic wave detection technology and the theory and implementation technology of electromagnetic wave detection technology, and so on.