Crop Yield Prediction Using Deep Neural Networks

Saeed Khaki* and Lizhi Wang

Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, United States

Crop yield is a highly complex trait determined by multiple factors such as genotype, environment, and their interactions. Accurate yield prediction requires fundamental understanding of the functional relationship between yield and these interactive factors, and to reveal such relationship requires both comprehensive datasets and powerful algorithms. In the 2018 Syngenta Crop Challenge, Syngenta released several large datasets that recorded the genotype and yield performances of 2,267 maize hybrids planted in 2,247 locations between 2008 and 2016 and asked participants to predict the yield performance in 2017. As one of the winning teams, we designed a deep neural network (DNN) approach that took advantage of state-of-the-art modeling and solution techniques. Our model was found to have a superior prediction accuracy, with a root-mean-square-error (RMSE) being 12% of the average yield and 50% of the standard deviation for the validation dataset using predicted weather data. With perfect weather data, the RMSE would be reduced to 11% of the average yield and 46% of the standard deviation. We also performed feature selection based on the trained DNN model, which successfully decreased the dimension of the input space without significant drop in the prediction accuracy. Our computational results suggested that this model significantly outperformed other popular methods such as Lasso, shallow neural networks (SNN), and regression tree (RT). The results also revealed that environmental factors had a greater effect on the crop yield than genotype.

Keywords: yield prediction, machine learning, deep learning, feature selection, weather prediction

1. INTRODUCTION

Crop yield prediction is of great importance to global food production. Policy makers rely on accurate predictions to make timely import and export decisions to strengthen national food security (Horie et al., 1992). Seed companies need to predict the performances of new hybrids in various environments to breed for better varieties (Syngenta, 2018). Growers and farmers also benefit from yield prediction to make informed management and financial decisions (Horie et al., 1992). However, crop yield prediction is extremely challenging due to numerous complex factors. For example, genotype information is usually represented by high-dimensional marker data, containing many thousands to millions of makers for each plant individual. The effects of the genetic markers need to be estimated, which may be subject to interactions with multiple environmental conditions and field management practices.

Many studies have focused on explaining the phenotype (such as yield) as an explicit function of the genotype (G), environment (E), and their interactions (G×E). One of the straightforward and common methods was to consider only additive effects of G and E and treat their interactions as noise (DeLacy et al., 1996; Heslot et al., 2014). A popular approach to study the G×E effect was

to identify the effects and interactions of mega environments rather than more detailed environmental components. Several studies proposed to cluster the environments based on discovered drivers of G×E interactions (Cooper and DeLacy, 1994; Chapman et al., 2000). Crossa et al. (1995) and Crossa and Cornelius (1997) used the sites regression and the shifted multiplicative models for G×E interaction analysis by dividing environments into similar groups. Burgueño et al. (2008) proposed an integrated approach of factor analytic (FA) and linear mixed models to cluster environments and genotypes and detect their interactions. They also stated that FA model can improve predictability up to 6% when there were complex G×E patterns in the data (Burgueño et al., 2011). Linear mixed models have also been used to study both additive and interactive effects of individual genes and environments (Crossa et al., 2004; Montesinos-López et al., 2018).

More recently, machine learning techniques have been applied for crop yield prediction, including multivariate regression, decision tree, association rule mining, and artificial neural networks. A salient feature of machine learning models is that they treat the output (crop yield) as an implicit function of the input variables (genes and environmental components), which could be a highly non-linear and complex function. Liu et al. (2001) employed a neural network with one hidden layer to predict corn yield using input data on soil, weather, and management. Drummond et al. (2003) used stepwise multiple linear regression, projection pursuit regression, and neural networks to predict crop yield, and they found that their neural network model outperformed the other two methods. Marko et al. (2016) proposed weighted histograms regression to predict the yield of different soybean varieties, which demonstrated superior performances over conventional regression algorithms. Romero et al. (2013) applied decision tree and association rule mining to classify yield components of durum wheat.

In this paper, we use deep neural networks to predict yield, check yield, and yield difference of corn hybrids from genotype and environment data. Deep neural networks belong to the class of representation learning models that can find the underlying representation of data without handcrafted input of features. Deep neural networks have multiple stacked non-linear layers which transform the raw input data into higher and more abstract representation at each stacked layer (LeCun et al., 2015). As such, as the network grows deeper, more complex features are extracted which contribute to the higher accuracy of results. Given the right parameters, deep neural networks are known to be universal approximator functions, which means that they can approximate almost any function, although it may be very challenging to find the right parameters (Hornik et al., 1990; Goodfellow et al., 2016).

Compared with the aforementioned neural network models in the literature, which were shallow networks with a single hidden layer, deep neural networks with multiple hidden layers are more powerful to reveal the fundamental non-linear relationship between input and response variables (LeCun et al., 2015), but they also require more advanced hardware and optimization techniques to train. For example, the neural network's depth (number of hidden layers) has significant impact

on its performance. Increasing the number of hidden layers may reduce the classification or regression errors, but it may also cause the vanishing/exploding gradients problem that prevents the convergence of the neural networks (Bengio et al., 1994; Glorot and Bengio, 2010; He et al., 2016). Moreover, the loss function of the deep neural networks is highly non-convex due to having numerous non-linear activation functions in the network. As a result, there is no guarantee on the convergence of any gradient based optimization algorithm applied on neural networks (Goodfellow et al., 2016). There have been many attempts to solve the gradient vanishing problem, including normalization of the input data, batch normalization technique in intermediate layers, stochastic gradient descent (SGD) (LeCun et al., 1998; Ioffe and Szegedy, 2015), and using multiple loss functions for intermediate layers (Szegedy et al., 2015). However, none of these approaches would be effective for very deep networks. He et al. (2016) argued that the biggest challenge with deep neural networks was not overfitting, which can be addressed by adding regularization or dropout to the network (Srivastava et al., 2014), but it was the structure of the network. They proposed a new structure for deep neural networks using identity blocks or residual shortcuts to make the optimization of deeper networks easier (He et al., 2016). These residual shortcuts act like a gradient highway throughout the network and prevent vanishing gradient problem.

Deep learning models have recently been used for crop yield prediction. You et al. (2017) used deep learning techniques such as convolutional neural networks and recurrent neural networks to predict soybean yield in the United States based on a sequence of remotely sensed images taken before the harvest. Their model outperformed traditional remote-sensing based methods by 15% in terms of Mean Absolute Percentage Error (MAPE). Russello (2018) used convolutional neural networks for crop yield prediction based on satellite images. Their model used 3-dimensional convolution to include spatiotemporal features, and outperformed other machine learning methods.

The remainder of this paper is organized as follows. Section 2 describes the data used in this research. Section 3 provides a detailed description of our deep neural networks for yield prediction. Section 4 presents the results of our model. Section 5 describes the feature selection method. Finally, we conclude the paper in section 6.

2. DATA

In the 2018 Syngenta Crop Challenge (Syngenta, 2018), participants were asked to use real-world data to predict the performance of corn hybrids in 2017 in different locations. The dataset included 2,267 experimental hybrids planted in 2,247 of locations between 2008 and 2016 across the United States and Canada. Most of the locations were across the United States. This was one of the largest and most comprehensive datasets that were publicly available for research in yield prediction, which enabled the deployment and validation of the proposed deep neural network model. **Figure 1** shows the distribution of hybrids across the United States.
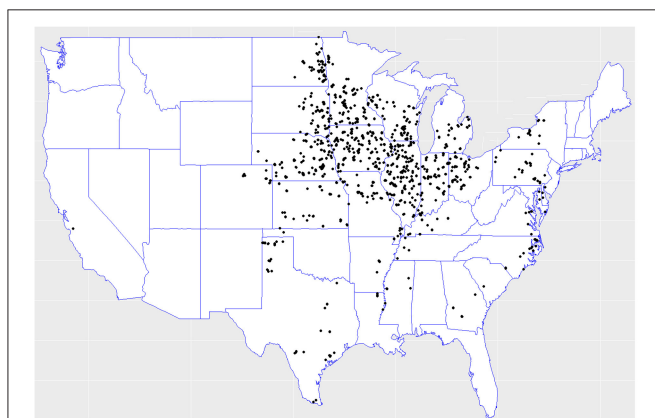
The training data included three sets: crop genotype, yield performance, and environment (weather and soil). The genotype dataset contained genetic information for all experimental hybrids, each having 19,465 genetic markers. The yield performance dataset contained the observed yield, check yield (average yield across all hybrids of the same location), and yield difference of 148,452 samples for different hybrids planted in different years and locations. Yield difference is the difference between yield and check yield, and indicates the relative performance of a hybrid against other hybrids at the same location (Marko et al., 2017). The environment dataset contained 8 soil variables and 72 weather variables (6 weather variables measured for 12 months of each year). The soil variables included percentage of clay, silt and sand, available water capacity (AWC), soil pH, organic matter (OM), cation-exchange capacity (CEC), and soil saturated hydraulic conductivity (KSAT). The weather data provided in the 2018 Syngenta Crop Challenge were normalized and anonymized. Based on the pattern of the data, we hypothesized that they included day length, precipitation, solar radiation, vapor pressure, maximum temperature, and minimum temperature. Part of the challenge was to predict the 2017 weather variables and use them for yield prediction of the same year.

The goal of the 2018 Syngenta Crop Challenge was to predict the performance of corns in 2017, but the ground truth response variables for 2017 were not released after the competition. In this paper, we used the 2001–2015 data and part of the 2016 data as the training dataset (containing 142,952 samples) and the remaining part of the 2016 data as the validation dataset (containing 5,510 samples). All validation samples were unique combinations of hybrids and locations, which did not have any overlap with training data.

3. METHODOLOGY

3.1. Data Preprocessing

The genotype data were coded in $\{-1, 0, 1\}$ values, respectively representing aa, aA, and AA alleles. Approximately 37% of the genotype data had mi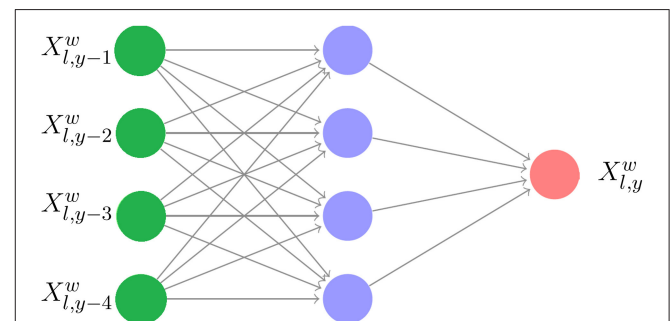ssing values. To address this issue, we used a two-step approach to preprocess the genotype data before they can be used by the neural network model. First, we used a 97% call rate to discard genetic markers whose non-missing values were below this call rate. Then we also discarded genetic markers whose lowest frequent allele's frequency were below 1%, since these markers were less heterozygous and therefore less informative. As a result, we reduced the number of genetic markers from 19,465 to 627. To impute the missing data in the remaining part of the genotype data, we tried multiple imputation techniques, including mean, median, and most frequent (Allison, 2001), and found that the median approach led to the most accurate predictions. The yield and environment datasets were complete and did not have missing data.

3.2. Weather Prediction

Weather prediction is an inevitable part of crop yield prediction, because weather plays an important role in yield prediction but it is unknown a priori. In this section, we describe our approach for weather prediction and apply it to predict the 2016 weather variables using the 2001–2015 weather data.

Let $X_{l,y}^w$ denote the weather variable w at location l in year y, for all $w \in \{1, ..., 72\}$, $l \in \{1, ..., 2247\}$, and $y \in \{2001, ..., 2016\}$. To predict the 2016 weather variables using historical data from 2001 to 2015, we trained 72 shallow neural networks for the 72 weather variables, which were used across all locations. There were two reasons for the aggregation of 2,247 locations: (1) the majority of the locations were in the middle west region, so it was reasonable to make the simplifying assumption that the prediction models were uniform across locations, (2) combining historical data for all locations allows sufficient data to train the 72 neural networks more accurately.

For each weather variable w, the neural network model explains the weather variable $X_{l,y}^w$ at location l in year y as a response of four previous years at the same location: $\{X_{l,y-1}^w, X_{l,y-2}^w, X_{l,y-3}^w, X_{l,y-4}^w\}$. We have tried other parameters for the periodic lag and found 4 years to yield the best results. As such, there were 24,717 samples of training data for each weather variable. The resulting parameters of the networks were then used to predict $X_{l,y=2016}^w$ using historical data of $X_{l,y=2012}^w$ to $X_{l,y=2015}^w$ for all l and w. The structure of a shallow neural network is given in **Figure 2**.


FIGURE 1 | Hybrids locations across the United States. Data collected from the 2018 Syngenta Crop Challenge Syngenta (2018).

FIGURE 2 | Neural network structure for weather prediction with a 4-year lag.

The reason for using neural networks for weather prediction is that neural networks can capture the nonlinearities, which exist in the nature of weather data, and they learn these nonlinearities from data without requiring the nonlinear model to be specified before estimation (Abhishek et al., 2012). Similar neural network approaches have also been used for other weather prediction studies (Maqsood et al., 2004; Bustami et al., 2007; Baboo and Shereef, 2010; Kaur et al., 2011; Abhishek et al., 2012; Bou-Rabee et al., 2017).

3.3. Yield Prediction Using Deep Neural Networks

We trained two deep neural networks, one for yield and the other for check yield, and then used the difference of their outputs as the prediction for yield difference. These models are illustrated in **Figure 3**. This model structure was found to be more effective than using one single neural network for yield difference, because the genotype and environment effects are more directly related to the yield and check yield than their difference.

The following hyperparameters were used in the training process. Each neural network has 21 hidden layers and 50 neurons in each layer. After trying deeper network structures, these dimensions were found to provide the best balance between prediction accuracy and limited overfitting. We initialized all weights with the Xavier initialization method (Glorot and Bengio, 2010). We used SGD with a mini-batch size of 64. The Adam optimizer was used with a learning rate of 0.03%, which was divided by 2 every 50,000 iterations (Kingma and Ba, 2014). Batch normalization was used before activation for all hidden layers except the first hidden layer. Models were trained for 300,000 maximum iterations. Residual shortcuts were used for every two stacked hidden layers (He et al., 2016). We used maxout activation (Goodfellow et al., 2016) function for all neurons in the networks except for the output layer, which did not have any activation function. In order to avoid overfitting, we used the L_2 regularization (Ng, 2004) for all hidden layers. We also added L_1 regularization (Ng, 2004) to the first layer to decrease the effect of redundant features, as in Lasso (Tibshirani, 1996). **Figure 4** depicts the detailed structure of the deep neural network, which was the same for yield and check yield prediction.

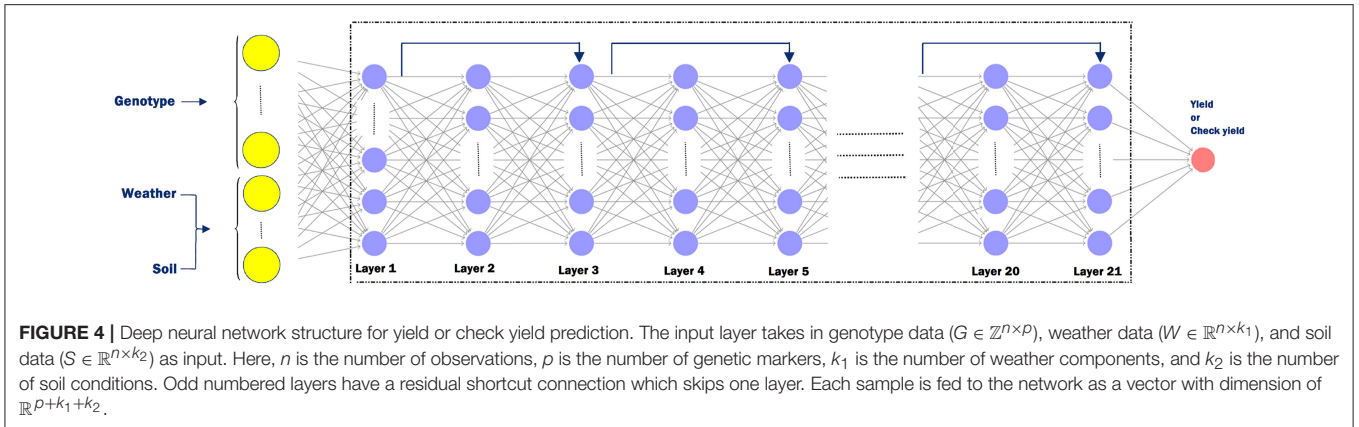**FIGURE 3** | Neural networks designed for predicting yield difference.
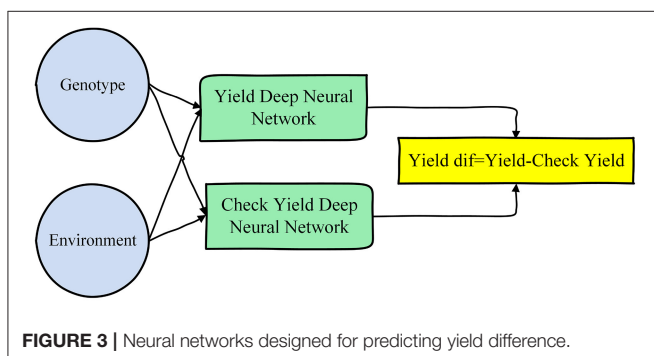
4. RESULTS

The two deep neural networks were implemented in Python using the Tensorflow open-source software library (Abadi et al., 2016). The training process took approximately 1.4 h on a Tesla K20m GPU for each neural network. We also implemented three other popular prediction models for comparison: The least absolute shrinkage and selection operator (Lasso), shallow neural network (having a single hidden layer with 300 neurons), and regression tree (Breiman, 2017). To ensure fair comparisons, two sets of these three models were built to predict yield and check yield separately, and the differences of their outputs were used as the prediction for the yield difference. All of these models were implemented in Python in the most efficient manner that we were capable of and tested under the same software and hardware environments to ensure fair comparisons.

The following hyperparameters were used for the regression tree. The maximum depth of the tree was set to 10 to avoid overfitting. We set the minimum number of samples required to split an internal node of tree to be 2. All features were used to train the regression tree. We tried different values for the coefficient of L_1 term (Ng, 2004) in the Lasso model, and found that values between 0.1 and 0.3 led to the most accurate predictions.
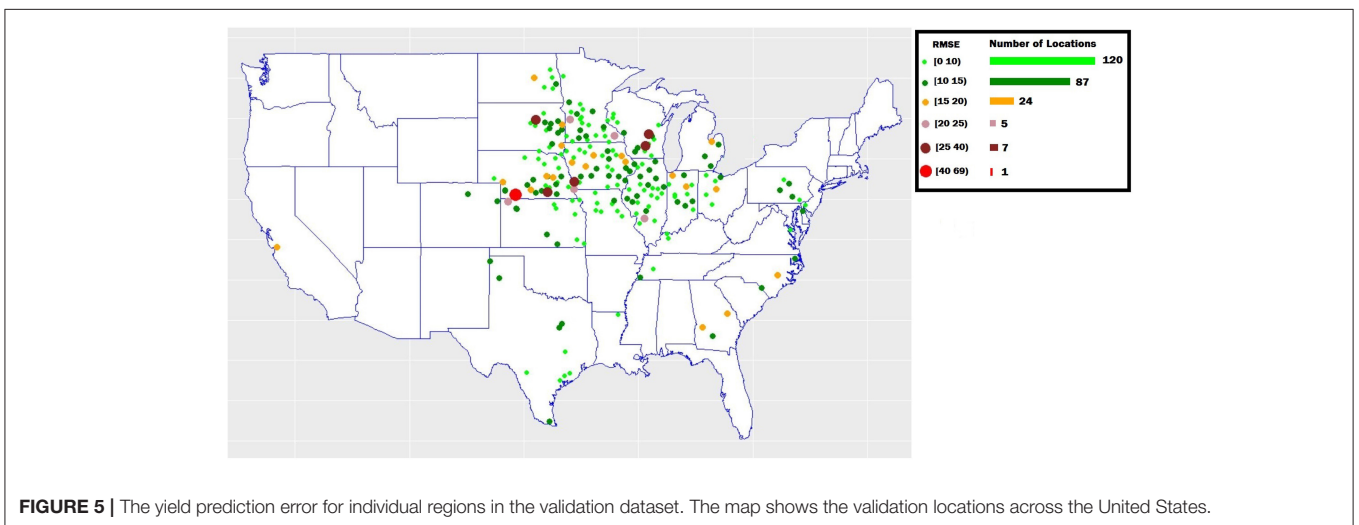
Table 1 compares the performances of the four models on both training and validation datasets with respect to the RMSE and correlation coefficient. These results suggest that the deep neural networks outperformed the other three models to varying extents. The weak performance of Lasso was mainly due to its linear model structure, which ignored epistatic or G×E interactions and the apparent nonlinear effects of environmental variables. SNN outperformed Lasso on all the performance measures except validation RMSE of the yield difference, since it was able to capture nonlinear effects. As a non-parametric model, RT demonstrated comparable performance with SNN with respect to yield and check yield but was much worse with respect to the yield difference. DNN outperformed all of the three benchmark models with respect to almost all measures; the only exception was that SNN had a better performance for the training dataset but worse for the validation dataset, which was a sign of overfitting. The DNN model was particularly effective in predicting yield and check yield, with RMSE for the validation dataset being approximately 11% of their respective average values. The accuracy for the check yield was a little higher than that for the yield because the former is the average yield across all hybrids and all years for the same location, which is easier to predict than the yield for individual hybrid at a specific location in a specific year. The model struggled to achieve the same prediction accuracy for yield difference as for the other two measures, although it was still significantly better than the other three benchmark models. The Lasso's performance seemed good for the yield difference with respect to RMSE, but it had a low correlation coefficient. This happened because Lasso's prediction was much centralized around the mean which may increase the risk of getting high prediction error on other test data. Let y, y_c, and y_d denote yield, check yield, and yield difference, respectively. Then, the variance of yield difference can be defined as
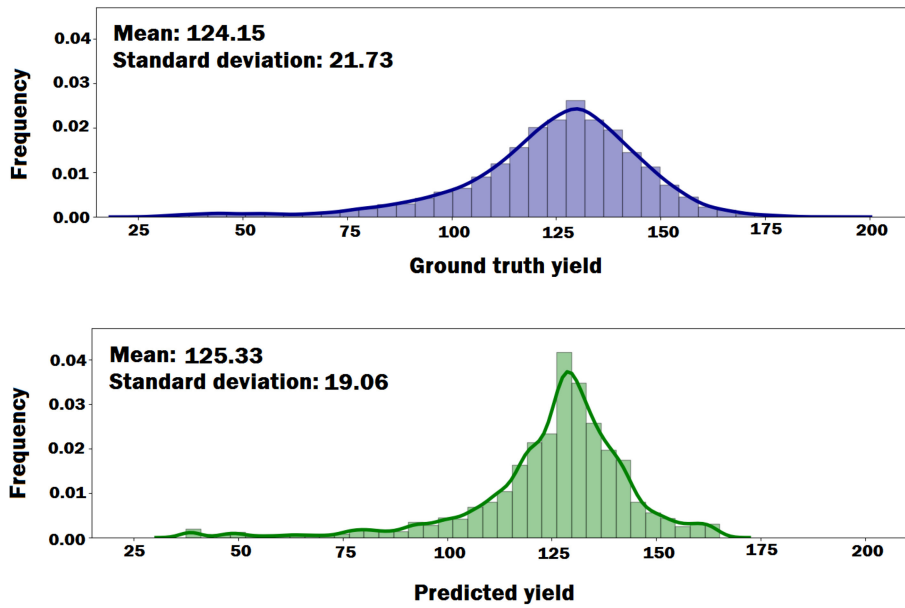
FIGURE 4 | Deep neural network structure for yield or check yield prediction. The input layer takes in genotype data ($G \in \mathbb{Z}^{n \times p}$), weather data ($W \in \mathbb{R}^{n \times k_1}$), and soil data ($S \in \mathbb{R}^{n \times k_2}$) as input. Here, n is the number of observations, p is the number of genetic markers, k_1 is the number of weather components, and k_2 is the number of soil conditions. Odd numbered layers have a residual shortcut connection which skips one layer. Each sample is fed to the network as a vector with dimension of $\mathbb{R}^{p+k_1+k_2}$.

TABLE 1 | Prediction performance with ground truth weather variables.

Model	Response variable	Training RMSE	Training correlation coefficient (%)	Validation RMSE	Validation correlation coefficient (%)
DNN	Yield	10.55	88.3	12.79	81.91
	Check yield	8.21	91.00	11.38	85.46
	Yield difference	11.79	45.87	12.40	29.28
Lasso	Yield	20.28	36.68	21.40	27.56
	Check yield	18.85	28.49	19.87	23.00
	Yield difference	15.32	19.78	13.11	6.84
SNN	Yield	12.96	80.21	18.04	60.11
	Check yield	10.24	71.18	15.18	60.48
	Yield difference	9.92	58.74	15.19	11.39
RT	Yield	14.31	76.7	15.03	73.8
	Check yield	14.55	82.00	14.87	69.95
	Yield difference	17.62	21.12	15.92	5.1

The average ± standard deviation for yield, check yield, and yield difference are, respectively, 116.51 ± 27.7, 128.27 ± 25.34, and −11.76 ± 14.27. DNN, Lasso, SNN, and RT stand for deep neural networks, least absolute shrinkage and selection operator, shallow neural network, and regression tree, respectively.

FIGURE 5 | The yield prediction error for individual regions in the validation dataset. The map shows the validation locations across the United States.

FIGURE 6 | The probability density functions of the ground truth yield and the predicted yield by DNN model. The plots indicate that DNN model can approximately preserve the distributional properties of the ground truth yield.

TABLE 2 | Prediction performance with predicted weather variables.

Model	Response variable	Training RMSE	Training correlation coefficient (%)	Validation RMSE	Validation correlation coefficient (%)
DNN	Yield	11.64	85.66	13.94	78.65
	Check yield	9.49	78.35	12.51	75.04
	Yield difference	12.80	37.64	15.54	19.86

TABLE 3 | Yield prediction performances of DNN(G), DNN(S), DNN(W), and Average model.

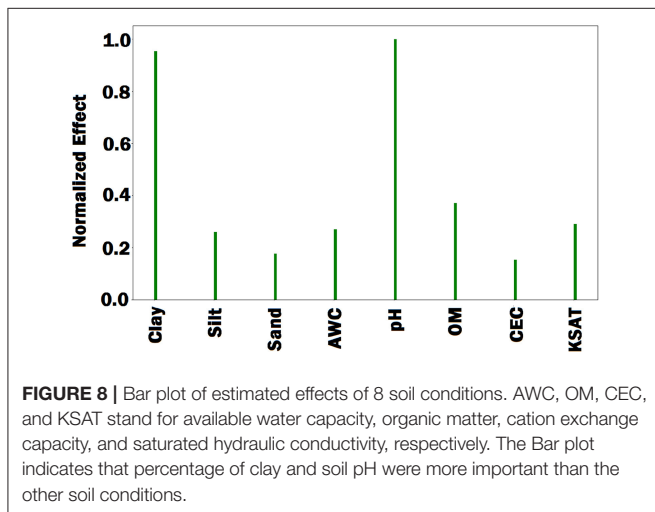| Model	Training RMSE	Training correlation coefficient (%)	Validation RMSE	Validation correlation coefficient (%)
DNN(G)	21.74	20.26	21.72	15.09
DNN(S)	15.28	73.37	15.49	72.04
DNN(W)	14.26	76.98	14.96	72.60
Average	24.40	0.0	23.14	0.0
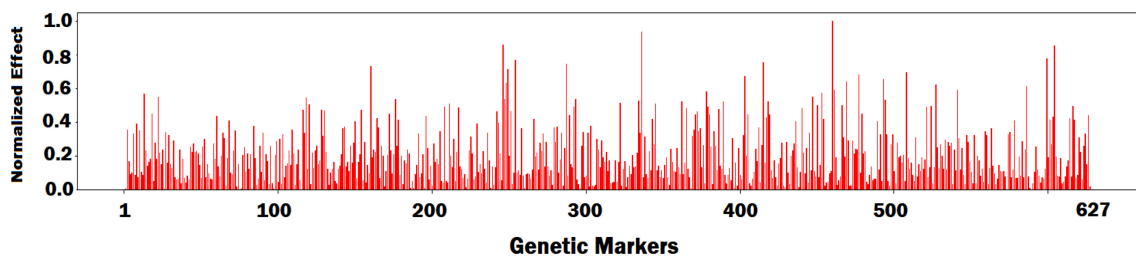
FIGURE 7 | Bar plot of estimated effects of 627 genetic markers.

Equation1. As shown in Equation 1, the yield difference was more difficult to predict since its variation depends on not only the individual variances of yield and check yield but also their covariance.

$$Var(y_d) = Var(y - y_c) = Var(y) + Var(y_c) - 2\,Cov(y, y_c) \quad (1)$$

To examine the yield prediction error for individual regions, we obtained prediction error across 244 locations existed in the validation dataset. As shown in **Figure 5**, the prediction error was consistently low (RMSE below 15) for most of locations (207 locations).

FIGURE 8 | Bar plot of estimated effects of 8 soil conditions. AWC, OM, CEC, and KSAT stand for available water capacity, organic matter, cation exchange capacity, and saturated hydraulic conductivity, respectively. The Bar plot indicates that percentage of clay and soil pH were more important than the other soil conditions.

We plotted the probability density functions of the ground truth yield and the predicted yield by the DNN model to see if the DNN model can preserve the distributional properties of the ground truth yield. As shown in **Figure 6**, the DNN model can approximately preserve the distributional properties of the ground truth yield. However, the variance of the predicted yield is less than the variance of the ground truth yield, which indicates DNN model's prediction was more centralized around mean.

To evaluate the effects of weather prediction on the performance of the DNN model, we obtained prediction results using the predicted weather data rather than the ground truth weather data. As shown in **Table 2**, the prediction accuracy of DNN deteriorated compared to the corresponding results in **Table 1**, which suggested how sensitive yield prediction is to weather prediction and the extent to which a perfect weather prediction model would improve the yield prediction results.
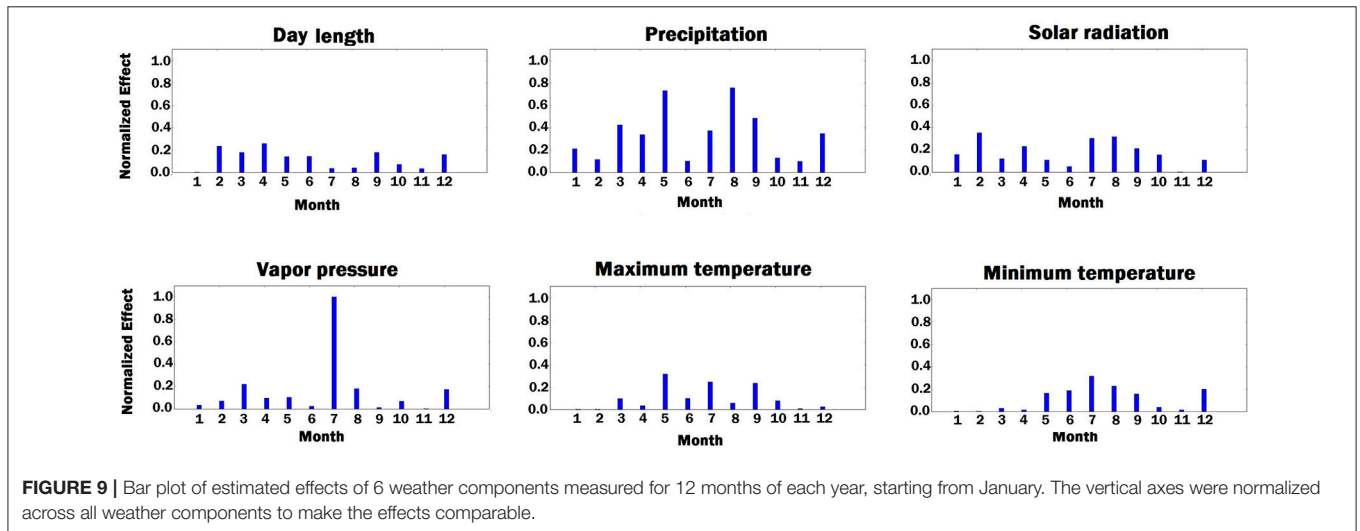
5. ANALYSIS

5.1. Importance Comparison Between Genotype and Environment

To compare the individual importance of genotype, soil and weather components in the yield prediction, we obtained the yield prediction results using following models:

DNN(G): This model uses the DNN model to predict the phenotype based on the genotype data (without using the environment data), which is able to capture linear and nonlinear effects of genetic markers.

DNN(S): This model uses the DNN model to predict the phenotype based on the soil data (without using the genotype

FIGURE 9 | Bar plot of estimated effects of 6 weather components measured for 12 months of each year, starting from January. The vertical axes were normalized across all weather components to make the effects comparable.

TABLE 4 | Yield prediction performance of DNN on the subset of features.

Model	Training RMSE	Training correlation coefficient (%)	Validation RMSE	Validation correlation coefficient (%)
DNN	12.01	84.01	12.81	81.44

The DNN model used 50 genetic markers and 20 environmental components selected by feature selection method.

and weather data), which is able to capture linear and nonlinear effects of soil conditions.

DNN(W): This model uses the DNN model to predict the phenotype based on the weather data (without using the genotype and soil data), which is able to capture linear and nonlinear effects of weather components.

Average: This model provides a baseline using only the average of phenotype for prediction.

Table 3 compares the performances of the above 4 models in the yield prediction. The results suggested that DNN(W) and DNN(S) had approximately the same performance and their prediction accuracies were significantly higher than DNN(G), which revealed that the environmental (weather and soil) components explained more of the variation within the crop yield compared to genotype.

5.2. Feature Selection

Genotype and environment data are often represented by many variables, which do not have equal effect or importance in yield prediction. As such, it is vital to find important variables and omit the other redundant ones which may decrease the accuracy of predictive models. In this paper, we used guided backpropagation method which backpropagates the positive gradients to find input variables which maximize the activation of our interested neurons (Springenberg et al., 2014). As such, it is not important if an input variable suppresses a neuron with negative gradient somewhere along the path to our interested neurons.

First, we fed all validation samples to the DNN model and computed the average activation of all neurons in the last hidden layer of the network. We set the gradient of activated neurons to be 1 and the other neurons to be 0. Then, the gradients of the activated neurons were backpropagated to the input space to find the associated input variables based on the magnitude of the gradient (the bigger, the more important). **Figures 7–9** illustrate the estimated effects of genetic markers, soil conditions, and weather components, respectively. The estimated effects indicate the relative importance of each feature compared to the other features. The effects were normalized within each group namely, genetic markers, soil conditions, and weather components to make the effects comparable.

As shown in **Figure 9**, solar radiation and temperature have considerable effects on the variation in corn yield across different environments. High corn yield is associated with low temperature and high solar radiation since lower temperature increases growth duration, thus crops can intercept more radiation (Muchow et al., 1990). Precipitation is an important factor. Hu and Buyanovsky (2003) found that high corn yield was associated with less rainfall in the planting period, and above-average rainfall throughout May, when seed germination and emergence happened. More rainfall with cooler temperatures were also necessary from June through August, followed by less rainfall and higher temperatures in the September early October period. The amount of vapor pressure during growing season has impact on the variation in the potential corn

yield since high vapor pressure can cause yield loss in corns (Zhang et al., 2017).

To evaluate the performance of the feature selection method, we obtained prediction results based on a subset of features. As such, we sorted the all features based on their estimated effects, and selected 50 most important genetic markers and 20 most important environmental components. **Table 4** shows the yield prediction performance of DNN model using these selected features. The prediction accuracy of DNN did not drop significantly compared to the corresponding results in **Table 1**, which suggested the feature selection method can successfully find the important features.

6. CONCLUSION

We presented a machine learning approach for crop yield prediction, which demonstrated superior performance in the 2018 Syngenta Crop Challenge using large datasets of corn hybrids. The approach used deep neural networks to make yield predictions (including yield, check yield, and yield difference) based on genotype and environment data. The carefully designed deep neural networks were able to learn nonlinear and complex relationships between genes, environmental conditions, as well as their interactions from historical data and make reasonably accurate predictions of yields for new hybrids planted in new locations with known weather conditions. Performance of the model was found to be relatively sensitive to the quality of weather prediction, which suggested the importance of weather prediction techniques.

A major limitation of the proposed model is its black box property, which is shared by many machine learning methods. Although the model captures G×E interactions, its complex model structure makes it hard to produce testable hypotheses that could potentially provide biological insights. To make the model less of a black box, we performed feature selection based on the trained DNN model using backpropagation method. The feature selection approach successfully found important features, and revealed that environmental factors had a greater effect on the crop yield than genotype. Our future research is to overcome this limitation by looking for more advanced models that are not only more accurate but also more explainable.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "TensorFlow: a system for large scale machine learning," in *OSDI'16 Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, Vol. 16, (Savannah, GA), 265–283.

Abhishek, K., Singh, M., Ghosh, S., and Anand, A. (2012). Weather forecasting model using artificial neural network. *Procedia Technol.* 4, 311–318. doi: 10.1016/j.protcy.2012.05.047

Allison, P. D. (2001). *Missing Data*, Vol. 136. Thousand Oaks, CA: Sage Publications.

Baboo, S. S., and Shereef, I. K. (2010). An efficient weather forecasting system using artificial neural network. *Int. J. Environ. Sci. Dev.* 1, 321. doi: 10.7763/IJESD.2010.V1.63

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* 5, 157–166. doi: 10.1109/72.279181

Bou-Rabee, M., Sulaiman, S. A., Saleh, M. S., and Marafi, S. (2017). Using artificial neural networks to estimate solar radiation in kuwait. *Renewable Sustain. Energy Rev.* 72, 434–438. doi: 10.1016/j.rser.2017.01.013

Breiman, L. (2017). *Classification and Regression Trees*. Abingdon: Routledge.

Burgueño, J., Crossa, J., Cornelius, P. L., and Yang, R.-C. (2008). Using factor analytic models for joining environments and genotypes without crossover genotype× environment interaction. *Crop Sci.* 48, 1291–1305. doi: 10.2135/cropsci2007.11.0632

Burgueño, J., Crossa, J., Cotes, J. M., Vicente, F. S., and Das, B. (2011). Prediction assessment of linear mixed models for multienvironment trials. *Crop Sci.* 51, 944–954. doi: 10.2135/cropsci2010.07.0403

Bustami, R., Bessaih, N., Bong, C., and Suhaili, S. (2007). Artificial neural network for precipitation and water level predictions of bedup river. *IAENG Int. J. Comput. Sci.* 34, 228–233.

Chapman, S. C., Cooper, M., Hammer, G. L., and Butler, D. G. (2000). Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Aust. J. Agric. Res.* 51, 209–222. doi: 10.1071/AR99021

Cooper, M., and DeLacy, I. H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theor. Appl. Genet.* 88, 561–572. doi: 10.1007/BF01240919

Crossa, J., and Cornelius, P. L. (1997). Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. *Crop Sci.* 37, 406–415. doi: 10.2135/cropsci1997.0011183X003700020017x

Crossa, J., Cornelius, P. L., Sayre, K., Ortiz-Monasterio, R., and Iván, J. (1995). A shifted multiplicative model fusion method for grouping environments without cultivar rank change. *Crop Sci.* 35, 54–62. doi: 10.2135/cropsci1995.0011183X003500010010x

Crossa, J., Yang, R.-C., and Cornelius, P. L. (2004). Studying crossover genotype× environment interaction using linear-bilinear models and mixed models. *J. Agric. Biol. Environ. Stat.* 9, 362–380. doi: 10.1198/108571104X4423

DeLacy, I. H., Basford, K. E., Cooper, M., Bull, J. K., and McLaren, C. G. (1996). "Analysis of multi-environment trials, an historical perspective," in *Plant Adaptation and Crop Improvement*, eds M. Cooper and G. L. Hammer (Wallingford: CABI), 39124.

Drummond, S. T., Sudduth, K. A., Joshi, A., Birrell, S. J., and Kitchen, N. R. (2003). Statistical and neural methods for site specific yield prediction. *Trans. ASAE* 46, 5. doi: 10.13031/2013.12541

Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Sardinia), 249–256.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*, Vol. 1. Cambridge: MIT Press Cambridge.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas), 770–778.

Heslot, N., Akdemir, D., Sorrells, M. E., and Jannink, J.-L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127, 463–480. doi: 10.1007/s00122-013-2231-5

Horie, T., Yajima, M., and Nakagawa, H. (1992). Yield forecasting. *Agric. Syst.* 40, 211–236. doi: 10.1016/0308-521X(92)90022-G

Hornik, K., Stinchcombe, M., and White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Netw.* 3, 551–560. doi: 10.1016/0893-6080(90)90005-6

Hu, Q., and Buyanovsky, G. (2003). Climate effects on corn yield in missouri. *J. Appl. Meteorol.* 42, 1626–1635. doi: 10.1175/1520-0450(2003)042<1626:CEOCYI>2.0.CO;2

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv [preprint] arXiv:1502.03167*.

Kaur, A., Sharma, J., and Agrawal, S. (2011). Artificial neural networks in forecasting maximum and minimum relative humidity. *Int. J. Comput. Sci. Netw. Secur.* 11, 197–199.

Khaki, S. (2019). *Source Code*. Available online at: https://github.com/saeedkhaki92/Yield-Prediction-DNN.

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv [preprint] arXiv:1412.6980*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436. doi: 10.1038/nature14539

LeCun, Y., Bottou, L., Orr, G. B., and Müller, K.-R. (1998). "Efficient backprop," in *Neural Networks: Tricks of the Trade*, eds G. Montavon, G. B. Orr, and K. R. Müller (Berlin; Heidelberg: Springer), 9–50.

Liu, J., Goering, C. E., and Tian, L. (2001). A neural network for setting target corn yields. *Trans. ASAE* 44, 705. doi: 10.13031/2013.6097

Maqsood, I., Khan, M. R., and Abraham, A. (2004). An ensemble of neural networks for weather forecasting. *Neural Comput. Appl.* 13, 112–122. doi: 10.1007/s00521-004-0413-4

Marko, O., Brdar, S., Panic, M., Lugonja, P., and Crnojevic, V. (2016). Soybean varieties portfolio optimisation based on yield prediction. *Comput. Electron. Agric.* 127, 467–474. doi: 10.1016/j.compag.2016.07.009

Marko, O., Brdar, S., Panić, M., Šašić, I., Despotović, D., Knežević, M., et al. (2017). Portfolio optimization for seed selection in diverse weather scenarios. *PLOS ONE* 12:e0184198. doi: 10.1371/journal.pone.0184198

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Montesinos-López, J. C., Mota-Sanchez, D., Estrada-González, F., et al. (2018). Prediction of multiple trait and multiple environment genomic data using recommender systems. *G3 (Bethesda)* 8, 131–147. doi: 10.1534/g3.117.300309

Muchow, R. C., Sinclair, T. R., and Bennett, J. M. (1990). Temperature and solar radiation effects on potential maize yield across locations. *Agron. J.* 82, 338–343. doi: 10.2134/agronj1990.00021962008200020033x

Ng, A. Y. (2004). "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the Twenty-first International Conference on Machine Learning* (Bnaff: ACM), 78.

Romero, J. R., Roncallo, P. F., Akkiraju, P. C., Ponzoni, I., Echenique, V. C., and Carballido, J. A. (2013). Using classification algorithms for predicting durum wheat yield in the province of buenos aires. *Comput. Electron. Agric.* 96, 173–179. doi: 10.1016/j.compag.2013.05.006

Russello, H. (2018). Convolutional neural networks for crop yield prediction using satellite images. IBM Center for Advanced Studies.

Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for simplicity: the all convolutional net. *arXiv [preprint] arXiv:1412.6806*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.

Syngenta (2018). *Syngenta Crop Challenge In Analytics*. Available online at: https://www.ideaconnection.com/syngenta-crop-challenge/challenge.php/.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

You, J., Li, X., Low, M., Lobell, D., and Ermon, S. (2017). "Deep gaussian process for crop yield prediction based on remote sensing data," in *Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, CA), 4559–4566.

Zhang, S., Tao, F., and Zhang, Z. (2017). Spatial and temporal changes in vapor pressure deficit and their impacts on crop yields in China during 1980–2008. *J. Meteorol. Res.* 31, 800–808. doi: 10.1007/s13351-017-6137-z