Prediction of Strawberry Yield and Farm Price Utilizing Deep Learning

Lobna Nassar
Electrical and Computer Engineering Department
University of Waterloo
Ontario, Canada
lnassar@uwaterloo.ca

Ifeanyi Emmanuel Okwuchi
Systems Design Engineering Department
University of Waterloo
Ontario, Canada
ifeanyi.okwuchi@uwaterloo.ca

Muhammad Saad
Electrical and Computer Engineering Department
University of Waterloo
Ontario, Canada
m32saad@uwaterloo.ca

Fakhri Karray
Electrical and Computer Engineering Department
University of Waterloo
Ontario, Canada
karray@uwaterloo.ca

Kumaraswamy Ponnambalam
Systems Design Engineering Department
University of Waterloo
Ontario, Canada
ponnu@uwaterloo.ca

Prarabdha Agrawal
Electrical and Computer Engineering Department
University of Waterloo
Ontario, Canada
p8agrawal@edu.uwaterloo.ca

***Abstract*— The currently deployed prediction models for strawberry fresh produce (FP) are based on either conventional machine learning (ML) or on simple deep learning (DL) models that are mostly applied for yield prediction. In this paper, we propose more comprehensive DL models that are applied for the first time to predict strawberry yield. The strawberry price is predicted as well directly from weather input parameters and yield. The strawberry price prediction is achieved using compound DL models such as Convolutional Long Short-Term Memory Recurrent Neural Network (CNN-LSTM). It is found that by adding attention, the performance of the compound models usually improves. After utilizing an aggregated performance measure to find the best model, the Attention-CNN-LSTM model proved to be the best compared to the rest of the deployed conventional ML models as well as the compound and simple DL models. The aggregated measure shows that this model is capable of precisely predicting strawberry prices five weeks ahead while maintaining the lowest prediction error and the highest model correlation.**

Keywords—Price Prediction, Yield Prediction, Deep Learning, Neural Networks, Attention, Fresh Produce, Strawberry

I. INTRODUCTION

Finding a precise farmer price prediction model is an important part of the procurement process in the fresh produce supply chain management (FSCM) since it protects farmers from price crash [1]. Farmer prices are affected by expected strawberry yield which depends on many factors related to weather, soil, irrigation technology, environment…etc. that are affected by the diverse regions of the sstrawberry. Many of these factors are becoming even more uncertain due to weather change, and are hard to predict, making decisions on strawberry prices extremely challenging task.

The lack of required data and high expense of data acquisition hinder the smooth application of complex forecasting models. Hence, it is decided to focus on strawberry price prediction, given the availability of free data for strawberry yield and prices from California strawberry Commission site [2] along with corresponding weather data from California Irrigation Management Information System (CIMIS) site [3]. Since California is one of the leading producers of strawberry as mentioned in [4], two of its largest stations are considered in this study.

Currently, the prices are determined using prediction methods that manipulate historical data. Supervised machine learning models are trained using this historical data of influential factors (independent variables) along with their corresponding prices (dependent variable) and the output is a learned function that can be used to predict future prices (output) given the values of the influential factors (input). In literature these forecasting techniques can be divided into two main categories: univariate models where only one factor is used as an independent variable and multivariate models that consider more than one independent variable.

Most of the current tools and models used for FP predictions are underperforming because: (1) They do not consider comprehensive set of influencing factors to predict crop yields since they are usually univariate. (2) They do not utilize the existing capacity in the elicitation of complex patterns and rules inherent in big data sources of price transactions and the influential factors to provide valuable information and knowledge on underlying complex processes affecting FP prices and quantities; something that has become possible by the advent of the state-of-the-art AI machine learning (ML) and deep learning (DL) techniques.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

In this paper multivariate models are used to predict farmers' offered strawberry prices using previous California stations strawberry yields and weather data. First, the weekly yield is predicted using the weekly weather data and then the daily prices are predicted using the daily yields. Second, the daily weather data is used as a direct input to predict the farmers' offered daily prices for strawberry.

The conventional machine learning models such as Artificial Neural Network (ANN) [5,6,7,8], and K-Nearest Neighbor regression (KNN) used in previous papers [9, 4, 10] as well as the simple long short term memory (LSTM) DL model used in [11] are mainly utilized for yield prediction. In this paper, more complex deep learning models are used for the first time not only for more precise yield prediction but also for price prediction.

To choose the most appropriate prediction model, three performance evaluation measures are utilized: the mean absolute error (MAE), mean square error (MSE) [12] and R^2 as well as an aggregated measure that summarizes the results of the three measures to decide the winning model. It is found that the compound DL models outperform the ML and simple DL models in the price prediction using yield application. Additionally, adding attention to the compound models helps in improving the prediction performance. It is also found that the ATT-CNN-LSTM model outperforms the rest of the ML and DL models across both applications of yield and price prediction using weather.

The rest of this paper is structured as follows: In Section 2, the utilized datasets are described along with the preprocessing needed for interpolating the missing data. Section 3, includes the details of the strawberry price prediction models including the conventional ML as well as the simple and compound DL models. In Section 4, details of the applied performance measures are presented. In Section 5, the conducted experiments are illustrated along with their result analysis. Finally, Section 6 concludes the work by short discussion of drawn conclusions followed by future work.

II. Data Sources and Preprocessing

The utilized dataset for testing the deployed models contain California's weather data as well as strawberries yield and price information from Oxnard and Santa Maria stations. The Salinas/Watsonville and Orange/San Diego/Coachella district region are used as well. The data is downloaded from two publicly available websites [2, 3] from 2006 to 2019.

A. Daily and Weekly Data

The daily yield and price data is extracted from the California strawberry Commission website [2]. For the weekly data, since the weather CIMIS site [3] only produces hourly, daily or yearly reports; no weekly data is available. The daily extracted weather data is aggregated to get weekly weather data for the prediction of weekly yield and price from weekly weather. The aggregation of the parameters values is found by averaging the daily values available in that week for each of the parameters to represent the weekly value of these parameters; evapotranspiration rate (ETo), precipitation, solar radiation, dew point, vapor pressure, wind speed, soil temp, the max, min, avg of air temp. & relative humidity parameters. The maximum of the daily maximum relative humidity values and air temperatures for the week is considered as the value of the maximum air temperature and humidity of that week similarly the minimum of the minimum daily air temperatures and relative humidity values available for the week is calculated to be that week minimum daily air temperature and humidity. The weekly yields and prices, matching the weekly weather, are extracted directly from the California Strawberry Commission site that offers weekly reports; no aggregation.

B. Dealing with Missing Data

The missing weekly weather data is interpolated using the linear interpolation function as used in [13] which is a function in Python language. The function of interpolation works by ignoring the index and treat the values as equally spaced. This way the function looks at the entire dataset as a trend. The trend is fitted as per the data without missing values and then on the basis of the trend the missing values are predicted and placed. [13]. Records with missing yield or price data are entirely dismissed.

III. Various Prediction Models for Strawberry

For all deployed models, tuning starts with learning rate scheduler to determine the optimal learning rate. Based on the validation results, either overfitting or underfitting, hyper-parameters like dropouts, batch-normalization are adjusted. Model weights are also saved as training proceeds keeping the weights corresponding to the epoch with best validation loss to prevent overfitting when the number of epochs is too high.

A. Conventional Machine Learning (ML) Models

1) Gradient Boosting (GB): is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak decision trees [14].

2) eXtreme Gradient Boosting or X-Gboost (XGB): is an implementation of gradient boosted decision trees designed for speed and performance. It is an optimized distributed gradient boosting framework designed to be highly efficient, flexible for most kinds of problems while being robust to overfitting. It implements parallel tree boosting (Gradient Boosting Machine) in fast and accurate way [15]. The model is built using the regression squared error function as the loss function while other hyper-parameters are set to default values. This gives better results than randomized search cross validation [16]. Like other traditional ML algorithms, its performance doesn't improve as much with increase in data.

B. Deep Learning Models

1) Simple models.

a) Long Short-Term Memory (LSTM): A special architecture of Recurrent Neural Networks (RNN) [17]. This is suitable for this type of applications, given their inherent time series and dynamic behavior. Basic RNNs are networks of neuron-like nodes designed into successive layers. Each node in a given layer is connected with a directed (one-way) connection to every other node in the next successive layer.

The work is supported by Grant's from NSERC CRD Funding and Loblaws Research Chair in Artificial Intelligence.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Each node (neoron) has a time-varying real-valued activation function. Each connection (synapse) has a modifiable real-valued weight. Nodes are either input nodes (receiving data from outside the network), output nodes (yielding results), or hidden nodes (modify data on the way from input to output) [17]. Recurrent neural networks (RNN) address the issue of sequential data by including a feedback which serves as a kind of memory. So the past inputs to the model leave a footprint. LSTM extends this idea by creating both a short-term and a long-term memory component. Hence, LSTM is a great tool for anything that has a sequence. Sequence-to-Sequence LSTM models have a wide array of applications such as time series forecasting. Furthermore, the inclusion of the gates enables the system not to experience the vanishing or exploding gradient issue [18]. Two LSTM structures are tried one with one hidden layer and 60 nodes and another deep LSTM with two hidden layers refered to as LSTM Deep in Fig. 8, 14 and 15. The LSTM Deep has two dropout layers 2 of 0.2, one activation layer of relu, one flattening layer, two dense layers with one layer of 30 nodes and last layer of one node, 2000 epochs, Adam with MSE as a the loss function.

b) Convolutional Neural Network (CNN): CNNs are widely used for image recognition but have also been used with text and time series as well as sequence data [19]. CNNs reduce dimensionality by using shared weights and they also have the ability to take into account, the spatial relationships that exist in the data [20,]. CNNs however are difficult to tune, require very large datasets and do not extract temporal features.

c) Gated Recurrent Unit (GRUs): GRUs is a class of RNNs that are similar to LSTMs with forget gates and fewer parameters. They do not have output gates and this streamlined architecture leads to faster convergence. GRUs face the problem of not being able to perform unbounded counting unlike LSTMs. LSTMs also perform better than GRUs in most problems [21].

2) Compound models:

a) CNN-LSTM: Due to the inability of CNNs to extract temporal features and failure of LSTMs to extract spatial features, researchers have considered stacking both CNNs and LSTMs so as to take advantage of their strengths [20, 22]. This combination has been applied in text classification, inventory time series analysis, etc. However, not much work has been done in fresh produce (FP) price prediction. Therefore, the use of CNN-LSTM in the price prediction application is introduced in this work. To get the best model version for this application, the hyper-parameter is tuned then the best set of hyper-parameters are found; The data goes into a Conv1D layer with 60 filters, kernel size of 5, stride 1, causal padding and a ReLu activation function. This is followed by batch normalization to speed up training and add regularization. The output is fed to an LSTM with 60 LSTM units and a ReLu activation. Another similar LSTM follows and feeds into a dense layer with 30 neurons. A series of dense layers with 10 and 1 neuron respectively follow. Our optimal loss function is MSE with Adam optimizer having a learning rate of 1e-4.

b) CNN-LSTM-GRU: Another explored architecture is stacking CNN, LSTM along with GRU. The Convolutional layer extracts the spatial features while LSTM and GRU extract the temporal features. This composition has never been applied to fresh food price prediction. The detailed architecture and best set of parameters are found to be as in Fig. 1. The input goes into a 1D convolutional layer with 60 filters, kernel size 5, causal padding and a ReLu activation function. The output then goes into an LSTM layer with 60 units followed by a GRU layer with 60 units. This goes into a sequence of a 30 neuron dense layer, 0.15 dropout, 10 neuron then single neuron dense layers. Adam optimizer with a learning rate of 1e-4 and MSE loss function are used.

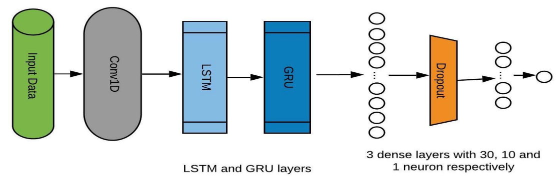

Fig. 1. CNN-LSTM-GRU.

c) ConvLSTM: The ConvLSTM model in [23] is used to take advantage of the internal convolutions as opposed to stacking up convolutional layers on LSTM layers. There has been very little work done using ConvLSTMs on time series data because they are originally built for images. The model is utilized here for fresh produce yield and price predictions. The data is reshaped into 5 dimensions [sample size, time steps, rows, columns, channels]. Our data has 68 weather input features get broken down into 2-time steps of 34 and 34 columns which results in a reshape of [sample size, 2, 1, 34, 1]. The input goes into ConvLSTM2D with 60 filters, kernel size of (1,3), stride of 1 and a ReLu activation function. This then undergoes batch normalization. This sequence of ConvLSTM and batch normalization is repeated 2 more times. Later, this goes into a dropout of 0.4, gets flattened, 32 neuron dense layer, dropout of 0.4, 10 neuron dense layer, dropout of 0.3 and finally a single neuron dense output as in Fig 2.

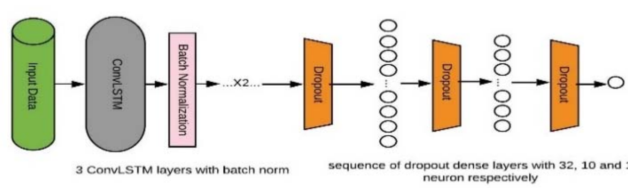

Fig. 2. ConvLSTM Model.

3) Adding Attention

a) Self-Attention: The aim of attention is to help the model focus on important parts of the input data as opposed to all the information. An attention function can be described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors [24]. Self-attention is an attention mechanism relating different positions of a single sequence in order to compute a representation of the same sequence. Attention is applied to the unit of each sequence and the unit of all sequences. Specifically, for our case, additive attention is utilized.

Additive attention computes the compatibility function using a feed-forward network with a single hidden layer [22]. The attention is computed as follows:

$$h_{t,t'} = \tanh(x_t^T W_t + x_{t'}^T W_x + b_t) \tag{1}$$

$$e_{t,t'} = \sigma(W_a h_{t,t'} + b_a) \tag{2}$$

$$a_t = \mathrm{softmax}(e_t) \tag{3}$$

$$l_t = \sum_{t'} a_{t,t'} x_{t'} \tag{4}$$

Where σ is the element wise sigmoid function, Wt and Wx are weight matrices corresponding to x_t^T and $x_{t'}^T$, Wa is the weight matrix corresponding to their non-linear combination and b_t, b_a are bias vectors [24]. Equation (4) shows how the attention l_t is calculated. To calculate this attention, the probability distribution a_t in (3) of the compatibility score $e_{t,t'}$ in (2) should be found first. This compatibility score is computed based on $h_{t,t'}$, the hidden representation of x_t and $x_{t'}$ computed in (1).

The use of attention has proved to enhance the performance of deep learning models [22-27]. This manifests in text sentiment analysis [22], vegetable yield prediction [26], healthcare question-answering, etc., but it is yet to be applied to FP price prediction [27]. The attention is added in the following subsections to three of the compound models previously presented.

b) CNN-LSTM Model with Attention: The proposed model is similar to the CNN-LSTM mentioned earlier but with self-attention after the LSTM layers. This attention layer uses additive attention with a sigmoid activation function as illustrated in Fig. 3.

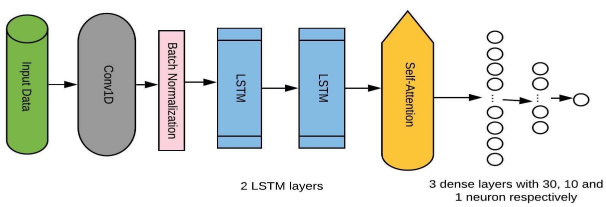

Fig. 3. CNN-LSTM with Attention.

c) ConvLSTM Model with Attention: Researchers have used temporal attention with a ConvLSTM for unsupervised anomaly detection [28]. The use of self-attention is proposed here to aid the ConvLSTM model focus on more relevant parts of the input data. In this model, the attention type is additive with a sigmoid activation function; see Fig. 4.

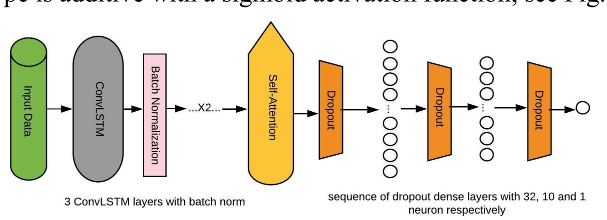

Fig. 4. ConvLSTM with Attention.

d) CNN-LSTM-GRU with Attention: In literature, there is almost no mention of CNN-LSTM-GRU with attention being used for fresh produce yield and price modelling. Adding an attention mechanism to the original CNN-LSTM-GRU is proposed here where the additive type self-attention with a sigmoid activation function is deployed; see Fig. 5.

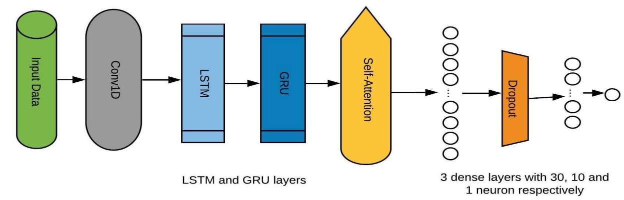

Fig. 5. CNN-LSTM-GRU with Attention.

IV. Performance Measures

To choose the most appropriate prediction model for strawberry prices, performance evaluation measures should be utilized for measuring the accuracy level of each of the deployed prediction models. The most frequently used performance measures in literature as in [29,30,31] are the Mean Absolute Percentage Error (MAPE), the Mean absolute error (MAE) and the Root Mean Square Error (RMSE) [32]. The R^2 measure is also used to measure the degree of correlation between the predicted and actual values [33]. Therefore, three of these measures are used plus a proposed aggregated measure to decide the winning model in terms of the overall performance. The two error metrics, MAE and RMSE are both negatively-oriented scores; hence it is decided to aggregate them by taking their average error. Since R^2 is a positively-oriented score, $1 - R^2$ has to be calculated before incorporation with the mentioned two measures to convert it to a negatively-oriented one to match MAE and RMSE. The aggregated measure is simply the average of the MAE and RMSE measures multiplied by the $1\text{-}R^2$. The resulting measure is a negatively-oriented one summarizing the overall performance; the lower the aggregated score the higher the accuracy.

V. Experiments and Result analysis

Three experiments are conducted: for indirect price prediction there are two experiments; predicting weekly yield using weekly weather experiment then daily price prediction using daily yield one. These two are followed by one experiment for direct weekly price prediction using weekly weather parameters. The details of the three experiments along with their results are summarized in this section.

A. Weekly Yield Prediction from Weekly Weather-W2Y

Since the main factor that affects yield is weather, we focused on California's yield prediction using weather data of California [2]. The aim is predicting average weekly yield (pounds/acre) using weather parameters as features as deployed in some other studies [34].

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

The California strawberry dataset is used in this experiment where 13 weather parameters are extracted in total to predict the yield; see Fig. 6. Missing weather values are interpolated as mentioned in Section II. On the other hand, the weekly yield data is collected from the California Strawberry Commission website [2] for Oxnard and Santa Maria stations. The weeks (rows) on which the yield is zero are discarded.

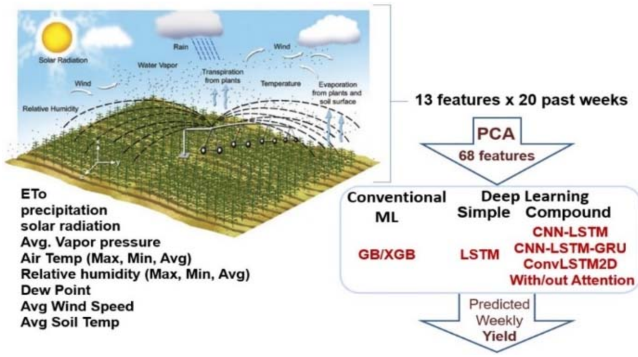

Fig. 6. Weather to yield prediction.

The yield data consists of the total yield volume in pounds and the average Pounds/Acre produced per week by each region in addition to the price per pound; see Fig. 7.

Yield Date (DD/MM/YYYY)	Price ($/Pound)	Yield (Pound/Acre)	Weather Date (DD/MM/YYYY)
02/09/2006	$1.34	10.00	06/03/2006
09/09/2006	$1.34	15.00	13/03/2006
16/09/2006	$1.34	31.00	20/03/2006
23/09/2006	$1.34	52.00	27/03/2006
30/09/2006	$1.34	149.00	03/04/2006
07/10/2006	$1.34	315.00	10/04/2006

Fig. 7. Screen shot of the Excel sheet showing the mapping of weather data to yield [starting date of five months window of weather data is associated with the following month yield data].

To predict the strawberry yield of the following month a lag of 20 weeks (approximately five months) is used. This lag value is chosen based on the fact mentioned in [34] that the strawberry yield is associated with the weather data with a lead time of at least 2 to 5 months. Therefore, a sliding window of weather data of 20 weeks (previous 5 months) is used to predict the yield at the end of the 5th week ahead (nearly one month ahead). For example, as illustrated in Fig. 7, the weather parameters of the 20 weeks (approximately 5 months) starting from the 6th of March till the 23rd of July, 2006 are used to predict the yield at the end of 5th week ahead (on the 2nd of September,2006).

The weather window of 20 weeks as well as the corresponding yield continuously slide a week at a time in each row to cover all years considered. The weather data of all the 20 weeks are considered as independent variables and are used to predict the yield which is the dependent variable. Therefore, the total number of independent variables are 13 (weather features) x 20 (weeks) therefore 260 parameters. Hence, using the Principle Component Analysis (PCA) is necessary in this case to reduce the number of features. After performing PCA, 68 features are left retaining 95% of the total variance in the dataset, see Fig. 6.

Out of all the examined ML and DL models detailed in Section III, the CNN-LSTM model with Attention proves to have the highest performance. The results are summarized in Fig. 8. The visual zoom displayed in Fig. 9 shows how the actual values of the yield are very close to the ones predicted using the ATT-CNN-LSTM model.

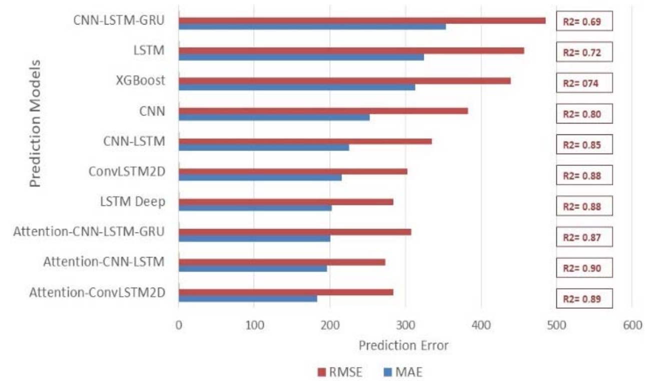

Fig. 8. Performance of yield prediction from weather measured using MAE, RMSE, R^2 for Machiene & Deep Learning models.

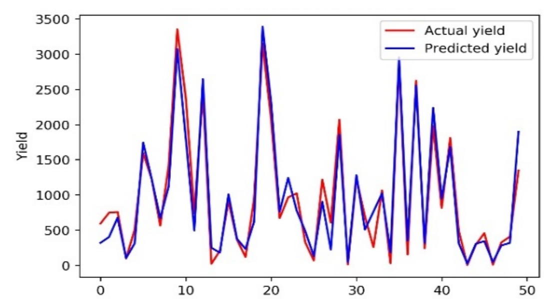

Fig. 9. Actual vs predicted yield by ATT-CNN-LSTM from weather.

B. Price Prediction

1) Daily Price Prediction from Daily Yield-Y2P: In this set of experiments, as in Fig. 10, the California dataset of strawberry yield and prices is used to train and test the ML XGB model along with the simple LSTM DL model and the compound CNN-LSTM DL model for predicting the daily prices of strawberry.

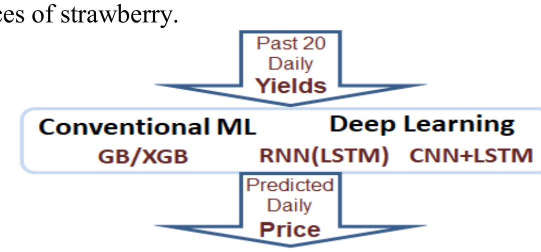

Fig. 10. Daily price prediction (One day ahead) using the 20 most recent daily yields.

For predicting the price using yield data, the dataset for all daily yields (Pounds per Acre) and prices (Dollar per Pound) of California stations from four different regions (Oxnard, Santa Maria, Salinas/Watsonville and Orange/San Diego/Coachella district) is compiled from year 2006 till 2018 then sorted. The records with empty yield are discarded. To predict the strawberry price on the same day a lag of 3 weeks is used.

This lag value is chosen based on a conducted correlation analysis between yield and price where the correlation is found to go below 50% after a lag period of three weeks. As illustrated in Fig. 10, the daily yield data of all the 21 days are considered as independent variables and used to predict the price of the 21st day; therefore, a sliding window of previous yield data of 21 days is used to predict the strawberry price on the 21st day. A file is prepared that maps the previous 21 days of yield (independent variables) to the 21st day price (dependent variable) with a total of 10915 records. Fig. 11 compares the deployed ML and the simple and compound DL models performances. It is found that the compound DL-based CNN-LSTM model outperforms the other models since it has the least prediction error. The visual zoom in Fig. 12 shows how the actual values of price are very close to the ones predicted using the CNN-LSTM model.

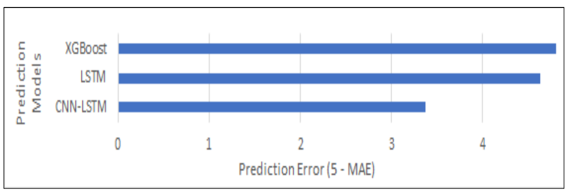

Fig. 11. Yield to Price prediction performance (MAE for ML & DL models).

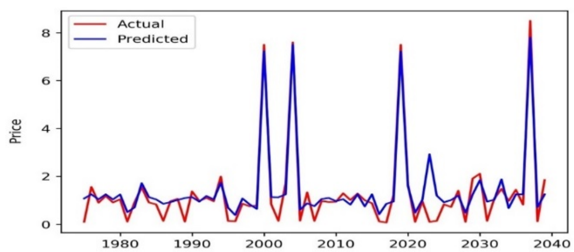

Fig. 12. Actual vs predicted price by CNN-LSTM from yield.

2) *Weekly Price Prediction from Weekly Weather-W2P:* In this experiment for the first time the price is predicted directly using the weather parameters without having to predict the yield first. Similar to the first experiment in section A, data of strawberry prices for Oxnard and Santa Maria stations is extracted and mapped to the corresponding weather parameters. To predict prices 5 weeks ahead a lag of 20 weeks of weather data is used. The weather of 20 weeks window and the corresponding price continuously slide a week at a time in each row to cover all considered years; see Fig. 7. All ML and DL models detailed in Section III and examined in the first experiment are applied. The CNN-LSTM model with attention persists to have the highest performance. The zoom in Fig. 13 shows how the actual strawberry prices are close to predicted ones using the ATT-CNN-LSTM model. Results of all deployed models are summarized in Fig. 14 as well.

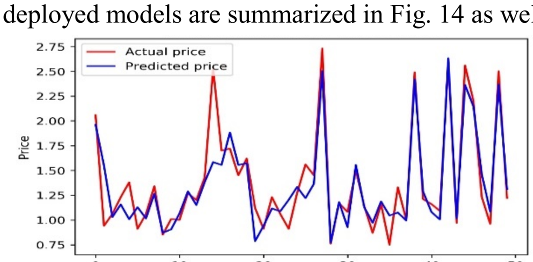

Fig. 13. Actual vs predicted price by ATT-CNN-LSTM from weather.

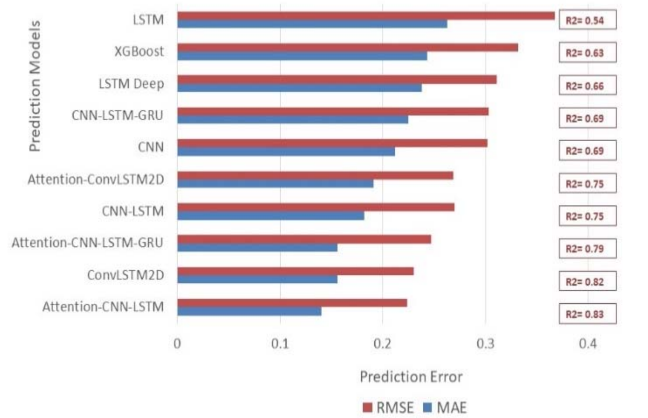

Fig. 14. Weather to Price prediction performance (MAE, RMSE, R^2 for Machiene & Deep Learning models.

The AGM metric explained in Section IV is finally applied to decide the winning model. The performances of the tested models are compared across the two main applications; yield and price prediction using weather. As evident in Fig. 15, due to getting the least value of aggregated score, the ATT-CNN-LSTM manages to preserve its place as the best performing model among all tested ML and DL models across both yield and price prediction applications.

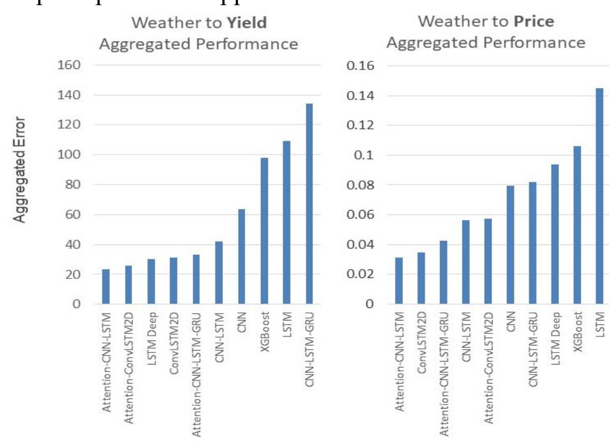

Fig. 15. ML & DL prediction models AGM accoss W2Y&W2P applications.

The contribution of adding the CNN simple DL model as well as adding attention to the AGM of the LSTM across all tested applications are summarized in Table I. These percentage improvement rates are calculated as in (5).

TABLE I. AGM REDUCTION IN PERCENTAGE DUE TO ADDING: (1) THE CNN TO THE SIMPLE LSTM DL MODEL. (2) ATTENTION TO THE RESULTING COMPOUND DL MODEL

Applications		DL Models		
		LSTM	*CNN-LSTM*	*Attention-CNN-LSTM*
AGM	*W2P*	113.43	44.88	23.53
	W2Y	0.14	0.06	0.03
	Y2P	4.64	3.38	
Modification		**+ CNN**		**+ Attention**
Improvement	*W2P*		**60%**	**48%**
	W2Y		**57%**	**50%**
	Y2P		**27%**	

$$Improvement\% = \frac{AGM_{before} - AGM_{After}}{AGM_{before}} \text{x}\,100 \qquad (5)$$

Where AGM_{before} & AGM_{after} are the prediction model *AGM* before architecture modification and after. *Improvement%* is the percentage reduction in AGM; if negative then it indicates increase in AGM which means deterioration in performance.

The *Improvement%* values listed in Table I indicate that the AGM reduces or the prediction performance improves after each addition across all tested applications.

VI. Conclusion and Future work

In conclusion, it can be deduced that in the price prediction using yield application the compound DL models such as CNN-LSTM outperform the other ML models, such as the Gradient Boosting, as well as the simple DL models, such as the LSTM by having the least value for prediction error (MAE). It is also found that all models improve after adding attention, therefore it is recommended to add attention to the compound DL models to improve prediction accuracy. After adding attention and based on the score of the aggregated performance measure, the ATT-CNN-LSTM model is found to be the winning model since it outperforms the rest of the deployed ML and DL models across both applications of yield and price prediction using weather.

Acknowledgment

We acknowledge the generous contribution of Loblaws which funded the Research Chair in AI at the University of Waterloo. Warm thanks are addressed to managers and officials of Loblaws for their support and help in this project.

References

[1] R. Waldick, L. Bizikova, D. White, and K. Lindsay, "An integrated decision-support process for adaptation planning: climate change as impetus for scenario planning in an agricultural region of Canada," Regional Environmental Change, vol. 17(1), pp. 187–200, 2017.

[2] The California Strawberry Commission website (https://www.calstrawberry.com/en-us/)

[3] The California Irrigation Management Information System (http://www.cimis.water.ca.gov/)

[4] T.B. Pathak, S.K. Dara, and A. Biscaro, "Evaluating correlations and development of meteorology-based yield forecasting model for strawberry," ADV METEOROL., vol. 2016.

[5] F. Karray and C.W. De Silva, "Soft computing and intelligent systems Design: theory, tools and applications," Addison Wesley, New York, 2004.

[6] P. Doganis, A. Alexandridis, P. Patrinos, and H. Sarimveis, "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing," Journal of Food Engineering, vol. 75 (2), pp.196-204, 2006.

[7] L.R. Medsker and L. C. Jain, "Recurrent Neural Networks: design and applications," 1st ed. Boca Raton, FL, USA: CRC Press, Inc., pp. 41, 1999.

[8] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," Neural Comput., vol. 9 (8), pp. 1735-1780, 1997.

[9] Y. Peng, C. Hsu and P. Huang, "Developing crop price forecasting service using open data from Taiwan markets," 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI), Tainan, pp. 172-175, 2015.

[10] M. L. Maskey, T. B. Pathak and S. K. Dara, "Weather based strawberry yield forecasts at field scale using statistical and machine learning models," Atmosphere 2019, vol. 10, pp. 378, 2019.

[11] Z. Jiang et al., "Predicting county level corn yields using deep long short term memory models," ArXiv abs/1805.12044, 2018.

[12] E. Lin, "Testing for the usefulness of forecasts," J. Forecast., vol. 30, pp.469–489, 2011.

[13] pandas: powerful Python data analysis toolkit. https://pandas.pydata.org/pandas-docs/stable/index.html. Date: Nov 09, 2019 Version: 0.25.3

[14] J. H. Friedman, "Greedy function approximation: a gradient boosting machine", Feb. 1999.

[15] J. Chen et. al "Social media popularity prediction based on visual-textual features with XGBoost". The 27th ACM International Conference, 2019.

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," In Proceedings of the 22nd ACM DATA MIN KNOWL DISC, San Francisco, CA, pp. 785–794, 2016.

[17] Y. wang, M. long, J. wang, Z. gao, and S.Y. philip. "PredRNN: recurrent neuralnetworks for predictive learning using spatiotemporal lstms," In: Adv Neural Inf Process Syst. pp. 879–888 , 2017.

[18] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," Computers in Biology and Medicine,vol. 96, pp. 189-202, 2018.

[19] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv preprint arXiv:1803.01271, 2019.

[20] L. Zheng et. al, "A fault prediction of equipment based on CNN-LSTM network," IEEE ICEI, Nanjing, China, pp. 537-541. Aug. 2019.

[21] G. Weiss, Y. Goldberg and E. Yahav. "On the practical computational power of finite precision RNNs for language recognition," the 56th Proc Conf Assoc Comput Linguist Meet, vol. 2: Short Papers, 2018.

[22] Y. Zhang et al., "A text sentiment classification modeling method based on coordinated CNN-LSTM-Attention model," Chin. J. Electron., vol. 28 (1), pp. 120 – 126, Jan. 2019.

[23] X. Shi et. Al., "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," NIPS, pp. 802–810, 2015.

[24] A. Vaswani et al., "Attention is all you need," In 31st Conference on Adv Neural Inf Process Syst, pp. 6000–6010, 2017.

[25] G. Zheng, S. Mukherjee, X. L. Dong and F. Li, "OpenTag: open attribute value extraction from product profiles," In proceedings of the 24th ACM SIGKDD. pp. 1049-1058, 2018.

[26] S. D. Alwis, Y. Zhang, M. Na and G. Li, "Duo attention with deep learning on tomato yield prediction and factor interpretation," PRICAI: Trends in Artificial Intelligence. 16th Pacific Rim International Conference on Artificial Intelligence. pp. 704 – 715, 2019.

[27] R. Cai et al., "An CNN-LSTM attention approach to understanding user query intent from online health communities," IEEE ICDMW, New Orleans, LA, pp. 430-437, 2017.

[28] C. Zhang et al. "A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data". The Thirty-Third AAAI Conference on Artificial Intelligence, CoRR , Vol. abs/1811.08055 (2018). arxiv: 1811.08055

[29] Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. Econometric Reviews. 29(5–6):594–621. https://doi.org/10.1080/07474938.2010.481556

[30] S. Jharkharia and M. Shukla, "Agri-fresh produce supply chain management: a state-of-the-art literature review," INT. J. OPER. PROD. MAN. vol. 33(2), pp.114-158

[31] R. J. Hyndman and A. B. Koehler. "Another look at measures of forecast accuracy," Int. J. Forecast., vol. 22, pp. 679-688, 2006.

[32] A. M. Ticlavilca, D. M. Feuz, and M. McKee, "Forecasting agricultural commodity prices using multivariate Bayesian machine learning regression," Proc. NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO, April, 2010.

[33] D. L. Alexander, A. Tropsha, and D. A. Winkler, "Beware of R(2): Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models," J. Chem. Inf. Model., vol. 55(7), pp. 1316-1322.

[34] T. B. Pathak, S. K. Dara and A. Biscaro, "Evaluating correlations and development of meteorology based yield forecasting model for strawberry," ADV METEOROL, vol. 2016.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE