Analysis of Corn Futures Models Across Differing Inputs

WO MING BOAZ CHEUNG, University of Toronto
GAVIN TIANYI PU, University of Toronto
YI FEI IVAN YE, University of Toronto

1 PROBLEM DEFINITION AND MOTIVATION

Modern commercialized agriculture is complex, with many moving parts, requiring large capital investments in infrastructure and land. Therefore, as these operations have evolved, there is a need for farmers to manage their risk portfolio, navigating through the volatility present in the market. Hedging against price fluctuations is caused by many different external factors, such as weather events, shifts in consumer demand, and geopolitical trends. Agriculture futures exist as a financial contract to set the selling of a product at a defined price and time, thus introducing stability into the system.

This paper will primarily focus on the futures markets within the United States. The crop most frequently traded on the Chicago Board of Trade is corn, due to its application for human consumption, animal feed, and biofuel use cases. Therefore, throughout this paper, corn is the baseline where we will attempt to predict its price based on a suite of external factors such as the prices of related agricultural products like oats, rice, and wheat. In addition, it will consider the weather patterns present in the primary corn-growing regions of Iowa and Illinois and the international corn prices in other primary corn-growing countries like Brazil and China, combined with the weather present within those two countries.

In this paper, we will explore the relationships that exist between the price and the different extraneous features mentioned. There will be a suite of models used to model this relationship. Starting from the baseline model of the feed-forward neural network model, the models that are also used to model this relationship include a convolutional neural network, a recurrent neural network, and a time series transformer model. By taking input from the data in the respective factors in the past and coming out with a prediction of the corn price at a specified time frame after a given time in the future.

From the results, we can figure out what features would have the most impact on the prices of corn futures in the U.S. In addition, from the test loss between the four models, we can also find out which model is the best at modeling these time series predictions without direct knowledge of the past price of the corn.

2 BACKGROUND & RELATED WORK

Machine learning methods have become increasingly applied to the field of agriculture. The number of agricultural research articles mentioning topics related to machine learning and deep learning have increased exponentially in the past decade. A broad range of data sources have been used for deep learning, including weather readings, camera images, and sounds. The utility of neural networks has made them applicable to many subfields such as crops, soil, irrigation, and livestock [4].

Our project in particular is inspired by a dissertation on estimating the price of corn futures. Weather and soil data from several states in the US were processed using CNN- and LSTM-related methods in order to predict percent changes in CBOT (Chicago Board of Trade) corn futures prices. While the neural network outputs were not accurate to the original data, the general price trend was still captured by the outputs, suggesting that incorporating weather data may supplement an existing model for price prediction so that its accuracy may increase [7].

Additional research has been performed on predicting attributes related to corn. Due to corn's economic importance as a staple food and animal feed, being able to predict corn prices and corn yields are of strategic interest. Wholesale prices of corn in China were predicted using an autoregressive neural network that outperformed other statistical models such as the random walk and the autoregressive (AR) model commonly used in time series analysis. However, the autoregressive neural network did not have lower mean squared error than an LSTM, though the difference in errors was not found to be significant [9]. Corn yield, a related and equally important measure, has also been able to be predicted using LSTMs and regular RNNs, though LSTMs generally perform better than RNNs for this task [5].

Authors' Contact Information: Wo Ming Boaz Cheung, University of Toronto; Gavin Tianyi Pu, University of Toronto; Yi Fei Ivan Ye, University of Toronto.

With the rise of globalization, country markets have also been able to interact and influence each other's prices. For instance, there is evidence supporting that agricultural markets for corn, wheat, and soybeans in the US and Italy are integrated. The relationship between these two markets can be explained by how a large quantity of agricultural goods moves from the US to Italy, and the CBOT may be recognized as a leader in determining the price of corn [6].

3 DATA SOURCE

Our data consists of four main components, the price of corn futures in the US [1], related crops futures in the US [2], international corn futures, and weather. Starting with the crops' future in the US our data is based from the Chicago board of trade. While for the international corn data, in China the exchange we used is based in Dalian since that's the region in China where most of the corn is grown. For Brazil we used the future prices from the B3 stock exchange located in São Paulo.These are the exchanges with the highest volume of trade through it in its respective countries, thus it will be representative of the prices that are within their economies.

Our weather is collected through the national weather service from the US government. Where we focus on the data is around the main agricultural regions in the US, namely regions in both Iowa and Illinois as they produce the most amount of corn a year [3].

4 BASELINE MODEL

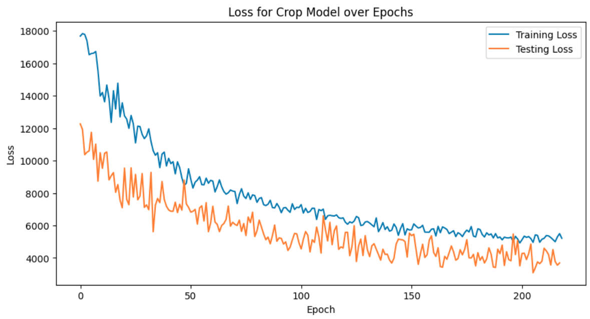
The baseline model was created using a feedforward neural network. Which is a basic design of neural networks that enable other more complex and deeper architectures that will be later explored in the paper to compare back to. Our setup of the feedforward neural network involved three hidden layers with hidden sizes of respectively 70, 50, and 30. Where between each layer there is an activation layer using leaky ReLU to introduce non-linearity to the neural network. Along with a dropout layer initialized to .1 to prevent overfitting. Our loss function is defined as a Mean Squared Loss function between the target and the output of the model.

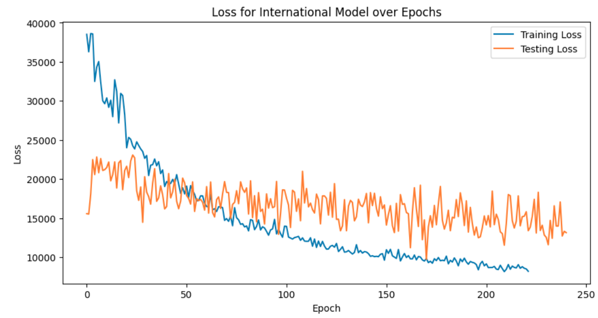
The feedforward model was training for 250 epochs with a learning rate of .0002 for the crop and international crop model and learning rate of .0001. The reasoning for this really low learning rate is due to the sensitivity of the weights.

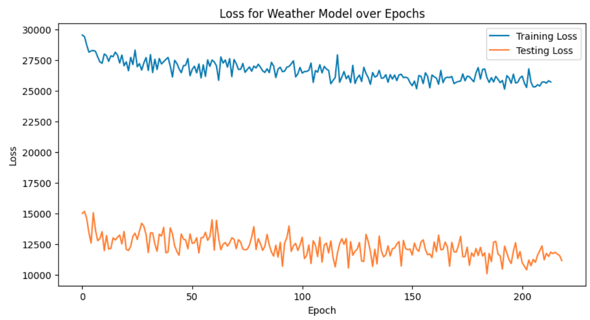
The batch size used to train the model was 64 with 250 epochs, and as shown in the graph below are the results for the loss values after training the model. As seen, the loss that is present when given the input of related crops in the US market results in a lower loss value compared to the other two categories of inputs. Showing a strong relationship within the different crops that are being traded in the Chicago board of trade. While international corn prices and weather have a higher loss value thus showing that there is a weaker relationship in their input's ability to predict the corn prices.

Feed Forward Model	**Train Loss**	**Test Loss**
Crops	5221.1027	4371.0909
International	7975.5408	14012.1665
Weather	25379.7700	11373.3968

As seen from the graphs below, the loss for crop model represents the inputs of related crops of oat, rice, and wheat. Shown the curve of lowering and the model learning the representation of these crops and how they are related to corn prices. While for international corn prices and weather the testing loss never effectively learns from its initial state and stays roughly the same.







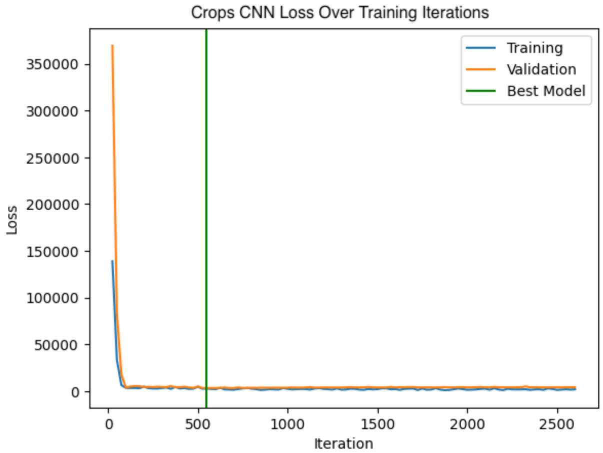
5 CNN

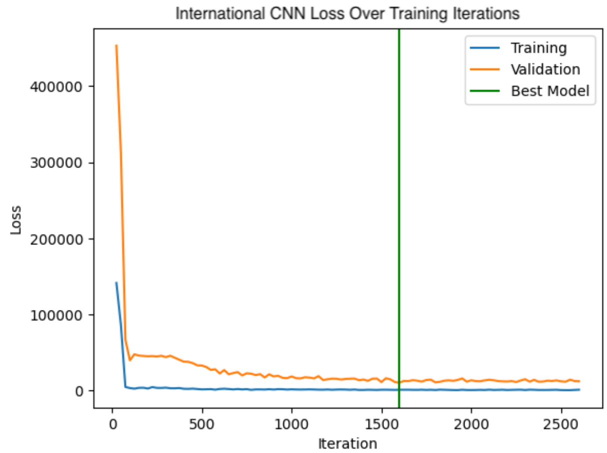
Unlike feedforward neural networks, CNNs contain layers that perform convolution and pooling. CNNs have been widely successful in recognizing patterns in imaging data. While their application to sequential data has been less discussed, it is not entirely unheard of, and CNNs have been previously applied in the same problem domain [7].

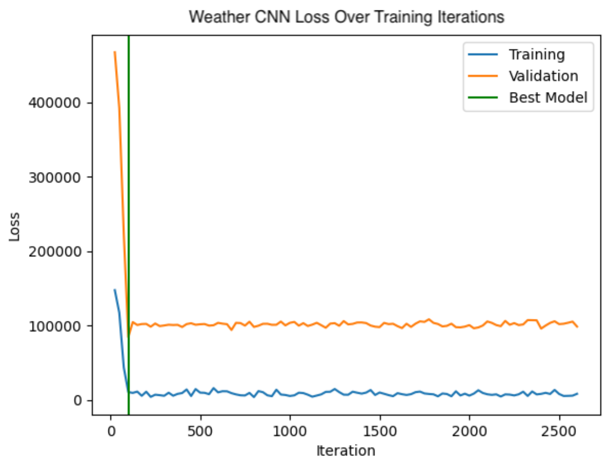
The dates in our data are not continuous because stock prices are not usually recorded on weekends, and several values in the international weather datasets were missing. In all of our models, we chose to only use dates for which all data were present. On average, 18 sequential dates in our data span about 30 days. We decided to use the CNN to predict corn futures prices of a given date using the past 30 days, which could therefore be approximated by using 18 sequential dates.

Our inputs consisted of multiple one-dimensional vectors of 18 dates worth of data. For instance, when predicting corn futures prices using prices of related crops, we used four separate vectors, one for each crop (red wheat, oats, rough rice, and wheat), as input. We then applied separate convolution layers on each vector. Each convolution layer used a kernel size of 4 and produced 2 channels in the output from the single channel inputs. Average pooling was performed on each of the results. We then applied another round of convolution using the same methodology but different parameters: the inputs contained 2 channels instead of 1, the kernel size was set to 2, and we requested 4 channels in the output. After two rounds of convolution and average pooling, we reshaped each modified input back into a one-dimensional vector (each of which contained 12 entries) and concatenated them together into a single vector. This vector was then used as an input to a small feedforward neural network with one hidden layer (which was always half the size of the input layer) to get a single output corresponding to the corn futures price on the latest date contained in the original input.

We trained 10 different CNNs for each category of inputs (30 models in total). Each CNN was trained using the same hyper-parameters: the batch size was 64, the learning rate was 0.001, and 100 epochs were used. To perform regularization, we used a weight decay of 0.0001 and early stopping. For each input category, we selected the model with the lowest validation loss at any iteration during training. The learning curves of each model are shown below, as well as values for training and test losses.







Feed Forward Model	**Train Loss**	**Test Loss**
Crops	5221.1027	4371.0909
International	7975.5408	14012.1665
Weather	25379.7700	11373.3968

CNN	**Train Loss**	**Test Loss**
Crops	2723.3715	3591.6704
International	926.4779	7990.9023
Weather	10582.3583	31210.2832

The CNN architecture appears to perform best on related US crops, where the training and validation losses of the best model were both below 3200. When used on a test set, we found the test loss to be slightly less than 3600, still indicating good performance and generalization to new data. The CNN did not perform well on using US weather to predict corn prices. The training loss was in the 10000s, while the validation loss was in the 80000s. Neither loss became lower even with increased training, suggesting that the model could not learn any further from the data. The test loss for this model was in the 30000s, indicating further that the

CNN model did not perform well using US weather data as input. Finally, for international corn and weather inputs, the model performed extremely well on the training set with a loss of less than 1000, although the validation loss was in the 10000s for the best model. We calculate the test loss to be less than 8000, indicating that this model still performs well at predicting US corn futures prices, though not as well as the CNN model training on solely US crops.

6 TIME SERIES TRANSFORMER

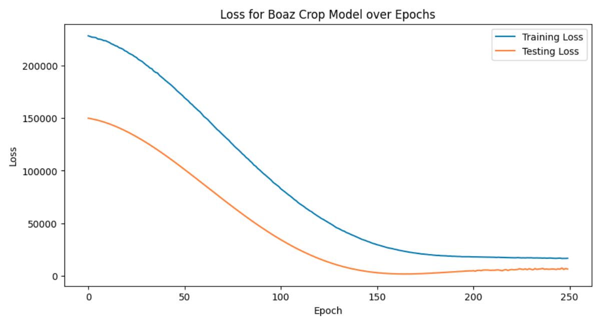
Transformers is a neural network architecture that is first intended to be used for natural language processing tasks. Based on the concept of self-attention mechanisms to obtain all the history at the hidden layers. The time series transformer used is an extension of that idea in the form of predicting sequence of numerical values from a sequential input instead of a sequence of words like in NLP. Using the strength of the transformer to look for differences and patterns within the past data in the sequence and categorize those patterns.

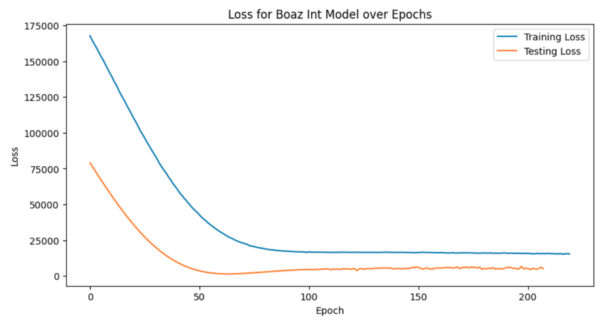
The model starts with an encoder layer which contains two parts of the base feed-forward network that is similar to the base model that was explored earlier in the paper. Along with the highlight of the model in the self-attention mechanism where it is able to calculate the attention score based on the input sequence. Enabling the model to determine which part of the past sequence is important in computing the current values. The hyperparameter nhead controls in here how many parallel attention computation are being performed to capture different perspectives of the data. With the current implementation the nhead parameter is set to be equal to the amount of parameters. Enabling the model to pay attention to each of the features that is present within the input data.

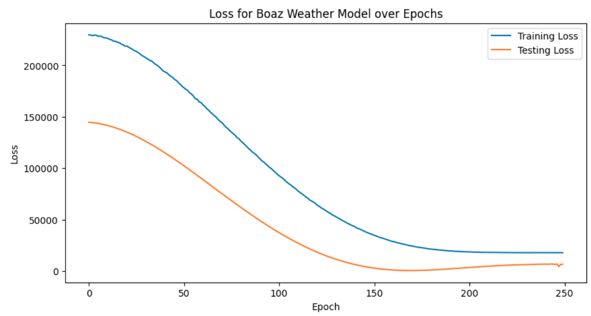
Then the encoding layers are stacked onto each other depending on the hyperparameter of num_layers to allow the data to flow sequentially through these layers to allow the model to represent the complex dependencies within the sequential data that is being imputed. For this model to balance between efficiency and model complexity we have chosen the value of depth 3 as well which is the same depth as the baseline model.

Since this is a time-series prediction there needs to be a mechanism that considers the sequence but that isn't built into transforms unlike RNNs which are built in. That's what positional encoding comes in. Based on the paper "Attention is All you need" [8], we implemented learned positional encodings which are parameters that are updated during the training of the model to fit it to this specific sequence. Enabling the model to represent the temporal dependencies that exist within that sequence. In addition domain-specific encoding for this corn option price data would be able to exist to have the best tool for the situation.

Considering the results, the test loss from the time series transformer performs slightly worse in modeling the corn prices with the input of related crops while performing better in modeling the corn prices given the input of international corn and weather patterns present in corn growing regions of the US. Thus showing the capacity for the time series transformer to represent more complex patterns that are present within the data. The training loss however, for the time series transformer seems to be higher across the board. There could be a couple of reasons as to why this is the case. It could be due to the lack of enough training data as this is a very deep and complex model and there are only around two thousand data points each with three to five parameters. Alternatively the reason could be within the data, since there could be anomalies within the training data set such as one off special events that spikes the prices. Where the time series transformer is intentionally not fitting to in order to better promote better generalization.







Feed Forward Model	Train Loss	Test Loss
Crops	5221.1027	4371.0909
International	7975.5408	14012.1665
Weather	25379.7700	11373.3968

Time Series Transformer	Train Loss	Test Loss
Crops	16527.2976	6215.6713
International	15398.3531	5097.5148
Weather	17948.1323	6644.0408

It can be seen from all three of the graphs that the model is learning from a really high initial loss. Where in the beginning there is a really high rate at which the loss is decreasing and then slowing down towards the middle and converging towards the absolute minima. The models are trained till the training rate converges to a stable value, since with the different generalization mechanisms within the model it is important for it to learn all the features within the data.

Compared to the other models, the time series transformer started with a higher loss value suggesting that the initial weights were near zero and didn't update as rapidly as the other models. Taking around 20 epochs to reach similar loss levels. For the three different categorizations of data, weather takes the longest to flatten out at around 200 epochs while international corn flatten out at around 50 epochs and related crops flattening out at 125 epochs. This shows that there could be more features and patterns that take longer for the model to learn.

7 RNN

Since the data can be represented as a sequence where the next date follows the current date, a recurrent neural network was chosen as one of the models used to predict corn futures as it can handle temporal dependence and store memory. These features of a RNN are useful because the model can capture dependency within the data where past price trends can affect future ones.

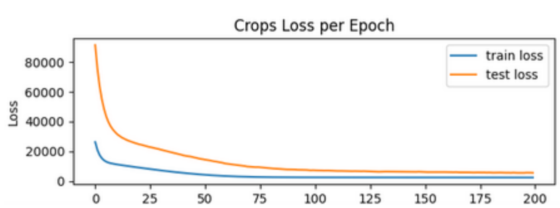
Despite LSTMs generally performing better than RNNs for this task [5], RNN was chosen because of a lack of data when it came to related crops. The data collected only spanned 10 years, counting only days where the market is active resulting in a limited dataset leading to comparable performance between LSTM and RNN. The data collected was normalized using MinMaxScaler() to provide uniformity between the feature values helping it balance the weights properly. It was then split 85% for train and 15% for test, which corresponds to approximately a little over a year for the testing data set. The resulting prediction was then unnormalized with inverse_transform() to make it more readable, the model still uses the normalized values when performing backward pass.

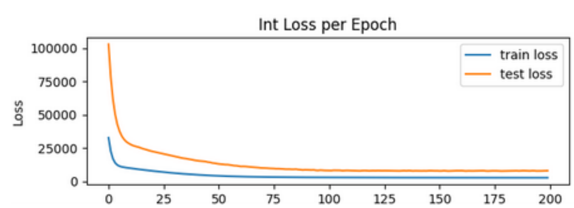
The model involves 4 layers where the input first passes through a RNN layer followed by a tanh activation function to help scale the output. Afterwards, a dropout layer is used to prevent overfitting, and lastly a fully connected layer to map the hidden state of size 100 to the output of the price. The loss function used was mean squared error for easy interpretation, since finding the average difference predicted only requires a square root. The batch size used was 64 with 200 number of epochs and a learning rate being 0.0001. Despite the rather simplistic hyperparameters, a slight increase of the learning rate to 0.0002 causes minor fluctuations in loss for all 3 models.

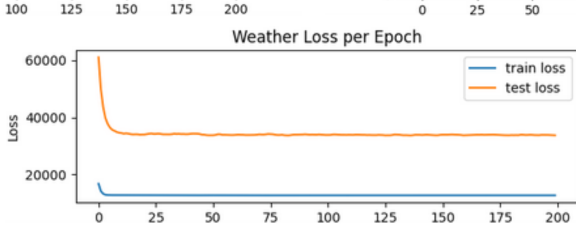
Feed Forward Model	Train Loss	Test Loss
Crops	5221.1027	4371.0909
International	7975.5408	14012.1665
Weather	25379.7700	11373.3968

RNN	Train Loss	Test Loss
Crops	2378.7358	3810.6604
International	2645.7793	7896.0229
Weather	12616.4111	33845.2148

When comparing the performance, from the table above, to the feed forward neural network baseline model, it can be observed that the RNN's loss is superior when it comes to the models utilizing crops or international factors. However, it pales in comparison when considering weather as the input. This is because weather data has little impact on the corn futures market [7], making the RNN less effective as it attempts to learn non-existing temporal correlations, complicating the training process.







The weather loss figure above supports the notion of weather data having no impact on predictions as the test loss does not approach the train loss compared to the crops loss figure and the international loss figure. The latter two also show a significant reduction is its loss with convergence between the two losses indicating that there exists temporal relationships that can be learned.

8 CONCLUSIONS & FURTHER WORK

Feed Forward Model	Train Loss	Test Loss
Crops	5221.1027	4371.0909
International	7975.5408	14012.1665
Weather	25379.7700	11373.3968

CNN	Train Loss	Test Loss
Crops	2723.3715	3591.6704
International	926.4779	7990.9023
Weather	10582.3583	31210.2832

Time Series Transformer	Train Loss	Test Loss
Crops	16527.2976	6215.6713
International	15398.3531	5097.5148
Weather	17948.1323	6644.0408

RNN	Train Loss	Test Loss
Crops	2378.7358	3810.6604
International	2645.7793	7896.0229
Weather	12616.4111	33845.2148

We verified the conclusion from Singh's paper about weather features having little correlation to the price of corn futures [7]. Even with the Time Series Transformer having a test loss of 6644.0408, meaning the predicted value was on average $81 off of the true value, it is not confident enough of a prediction to be used for any real decision making. All models performed better than the baseline of the Feed Forward Network, with the exception of the aforementioned weather features. However, while the test loss might be considered acceptable, the train loss on the other hand struggles for the Time Series Transformer. Ideally, our models would have been able to make proper predictions within a certain margin of error, but the data trend of corn futures and the economy as a whole is incredibly difficult to model due to external events such as the COVID-19 pandemic.

Several extensions of this paper can be addressed, such as exploring different combinations of input features to see if there are any potential interactions between features leading to better prediction accuracy. This can be further explored by adding more adjacent features to the ones already present, for instance natural disaster information to supplement the weather features. Additionally, leveraging ensemble methods or creating more complex/hybrid models could also have a large effect, since each model seems to have particular strengths and weaknesses when it comes to different types of input.

REFERENCES

[1] [n. d.]. Corn Futures. https://cn.investing.com/commodities/us-corn.

[2] [n. d.]. Historical Data. https://www.nasdaq.com/market-activity/quotes/historical.

[3] [n. d.]. National Weather Service. https://www.weather.gov/.

[4] K Dokic, L Blaskovic, and D Mandusic. 2020. From machine learning to deep learning in agriculture–the quantitative review of trends. In *IOP Conference Series: Earth and Environmental Science*, Vol. 614. IOP Publishing, 012138.

[5] Zehui Jiang, Chao Liu, Baskar Ganapathysubramanian, Dermot J Hayes, and Soumik Sarkar. 2020. Predicting county-scale maize yields with publicly available data. *Scientific Reports* 10, 1 (2020), 14957.

[6] Franco Rosa, Michela Vasciaveo, Robert D Weaver, et al. 2014. Agricultural and oil commodities: price transmission and market integration between US and Italy. *BAE* 3, 2 (2014), 93–117.

[7] Sriramjee Singh. 2022. *Estimating the impact of weather on CBOT corn futures prices using machine learning*. Ph. D. Dissertation. Iowa State University.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[9] Xiaojie Xu and Yun Zhang. 2023. Yellow corn wholesale price forecasts via the neural network. *EconomiA* 24, 1 (2023), 44–67.