Black Sea Journal of Agriculture
doi: 10.47115/bsagriculture.1536162

Open Access Journal
e-ISSN: 2618 – 6578

Research Article
Volume 7 - Issue 6: xxx-xxx / November 2024

DETERMINING THE EFFECTIVE METEOROLOGICAL PARAMETERS ON POTATO YIELD IN NİĞDE PROVINCE AND YIELD PREDICTION WITH MACHINE LEARNING AND DEEP LEARNING MODELS

Esma UZAN[1], Cevher ÖZDEN[2]*

[1]*Çukurova University, Faculty of Agriculture, Department of Agricultural Economy, 01330, Adana, Türkiye*
[2]*Çukurova University, Vocational School of Karaisalı, Department of Computer Technologies, 01330, Adana, Türkiye*

Abstract: In this study, we explore the impact of meteorological parameters on potato yield in Niğde province, a key agricultural region in Türkiye for potato production. Understanding the relationship between weather conditions and crop yield is crucial for optimizing agricultural practices and ensuring food security, especially in regions susceptible to climate variability. The study includes all meteorological parameters observed and recorded in Niğde Directorate of Meteorology, covering the period between 1990 and 2023. Through the analysis of historical weather data and potato yield records, we aim to identify the most influential meteorological factors affecting potato production. Then, artificial intelligence techniques such as Random Forest, Gradient Boost, Convolutional Neural Networks and Recurrent Neural Networks are utilized to predict the potato yield in order to provide valuable insights into how different weather patterns influence crop performance. The findings suggest that potato yield is correlated with meteorological parameters to some degree and AI techniques can predict the potato yield but lacks the precision. The reason is that potato production in Niğde is mostly done with irrigation. However, this further increases the risk that could result from the changes in groundwater levels and pollution in irrigation ponds for agricultural practices in Niğde.

Keywords: Potato yield, Correlation, Meteorology, Climate change, Artificial intelligence, Prediction

***Corresponding author:** Çukurova University, Vocational School of Karaisalı, Department of Computer Technologies, 01330, Adana, Türkiye
E mail: efeozden@gmail.com (C. ÖZDEN)
Esma UZAN https://orcid.org/0000-0003-1532-6268
Cevher ÖZDEN https://orcid.org/0000-0002-8445-4629

Received: August 20, 2024
Accepted: September 30, 2024
Published: November 15, 2024

Cite as: Uzan E, Özden C. 2024. Determining the effective meteorological parameters on potato yield in Niğde province and yield prediction with machine learning and deep learning models. BSJ Agri, 7(6): xxx-xxx.

1. Introduction

The potato is one of the most important and versatile crops in the world, it is rich in essential nutrients such as carbohydrates, vitamins, and minerals. Potatoes act as a crucial source of energy and has a significant place in global food security. Their adaptability to different growing conditions and ability to yield high production even in challenging environments make them a vital crop for sustaining populations, particularly in regions facing food scarcity. Therefore, potatoes are considered as one of the most important staple foods throughout the world and Türkiye. Niğde province is the largest potato production center in Türkiye and has the highest share in the country's potato production, and therefore, has a strategic importance in potato production (Table 1).

Potatoes yield is reported as 2107 kg/da for world in 2022, while it is 3739 kg/da for Türkiye and 3855 kg/da for Niğde (Faostat, 2022; TUİK, 2024). As Niğde is located in the interior part of Turkiye and has semi-arid climate characteristics, agricultural production is highly dependent on meteorological conditions. Mediterranean climate enables product yields to be more stable due to the predictable rainy season and long growing season, while yields are often less stable due to the unpredictability of rainfall and the higher likelihood of droughts in the semi-arid climate regions. Moreover, the limited and erratic rainfall can lead to crop failures, especially if the rain arrives late or not at all.

Parameters such as temperature, rainfall, humidity and sunshine duration play a critical role in crop production in terms of both quality and quantity. Especially the yield of subsoil crops such as potatoes is expected to be directly related to these meteorological parameters.

In this regard, determination of meteorological parameters that affect potatoes yield is important for sustainability of regional agricultural practices. It will also provide an insight into the mitigation alternatives to be applied against the risks of climate change. Considering the potential impacts of climate change on production patterns and yields in traditional agricultural regions, study findings will provide an essential foundation for policies and water management strategies.

Table 1. Potato production (ton) (Faostat, 2022; TUIK, 2024)

	Niğde	% of Türkiye	Türkiye	% of World	World
2013	512644	13,0	3955294	1,1	365918796
2014	618853	14,9	4166000	1,1	368940950
2015	674773	14,2	4760000	1,3	365059121
2016	892297	18,8	4750000	1,3	353617596
2017	835200	17,4	4800000	1,3	368590224
2018	732188	16,1	4550000	1,2	364017608
2019	716180	14,4	4979824	1,4	366801517
2020	689312	13,3	5200000	1,4	369381444
2021	575627	11,3	5100000	1,4	373787150
2022	679653	13,1	5200000	1,4	374777763

Another important contribution of the study will be the comparative application of different Artificial Intelligence models on potato yield prediction using the identified parameters.

2. Materials and Methods

In the study, potatoes yield in Niğde between 1990 and 2023 is chosen as dependent (target) variable). The data between 2004 and 2023 are obtained from the TUIK website, while the data between 1990 and 2003 are curated from Agricultural Structure and Production Statistics Book published by TUIK (Figure 1).

The input variables consist of meteorological parameters observed and recorded in Niğde Provincial Directorate of Meteorology between 1990 and 2023. The parameters contain a total of 31 different observation metrics and are obtained from MEVBIS, official web site of Turkish State Meteorological Service. 9 of the parameters are excluded from the study as they contain missing series and corrupted data that cannot be replaced. The remaining 22 meteorological parameters are given Table 2.

In summary, monthly observations of 22 meteorological parameters are curated, which sums to a total of 264 observations for each year between 1990 and 2023. Some of the time series had zero observation during certain months, such as snow height measurement during summer period. These series have been excluded from the study as they will produce NaN in the analysis result. Furthermore, potato is planted during March-April and harvested until early October; therefore, only the months between March and October (154 observations) are included in the study.

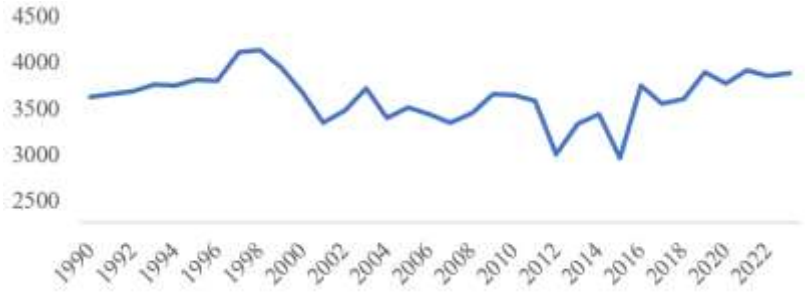

Figure 1. Potato yield (kg/da) in Niğde (1990-2023).

Table 2. Meteorological parameters (1990-2023)

No	Name	No	Name
1	Number of days with clear sky	12	Average soil temperature at 5 cm
2	Number of cloudy days	13	Average soil temperature at 20 cm
3	Number of days with frost	14	Average minimum temperature °C
4	Number of days with snow cover	15	Average daily maximum temperature
5	Number of days with snow precipitation	16	Average daily minimum temperature
6	Number of days with strong wind	17	Monthly average of daily total sunshine duration
7	Number of days with maximum temperature 30°C and above	18	Average temperature °C
8	Maximum temperature	19	Monthly total precipitation mm
9	Monthly maximum precipitation mm	20	Average wind speed
10	Number of days with minimum temperature 5°C and below	21	Average humidity %
11	Minimum temperature	22	Number of days with precipitation

Initially, correlation analysis has been performed to establish the relationship between inputs and target variables. The uncorrelated input variables are excluded from the dataset. Principle Component Analysis (PCA) is performed to see the possible components that could represent the inputs and increase the efficacy of correlation. Subsequently, Linear Regression, Random Forest and Gradient Boosting methods have been employed to predict the yield parameter using the inputs and components determined in the previous stage. CNN and RNN based deep learning models have been developed for this purpose, as well. The general structures of these models are given in the Figure 2.

The initial CNN model consisted of two Conv 1D layers each followed by a max pooling layer and ReLU activation, and two Fully Connected layers (Figure 2). The second RNN model is based on Long Short Term Memory (LSTM) architecture. It consisted of two LSTM layers and one last fully connected layer (Figure 3). Learning rate is set to 0.001, Adam optimization and Mean Squared Error (MSE) functions are preferred for training. Each model is trained for 100 epochs.

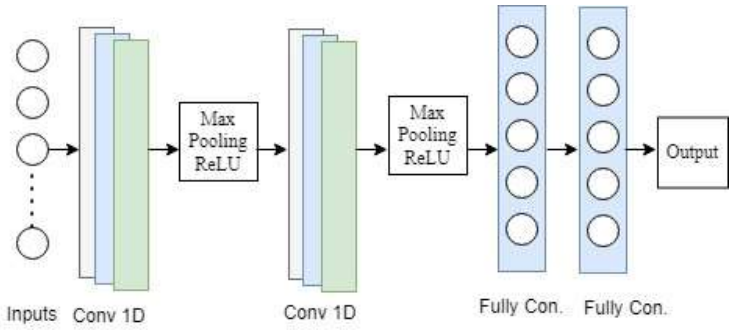

Figure 2. CNN model.

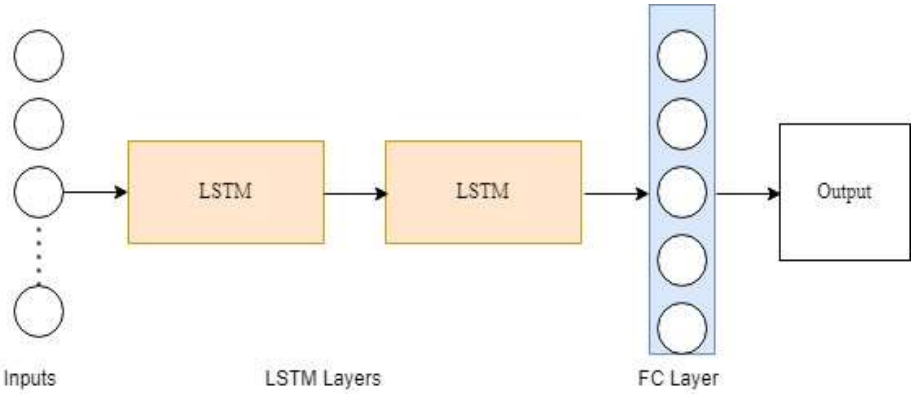

Figure 3. LSTM based RNN model.

3. Results and Discussion

Firstly, correlations between meteorological parameters and potato yield have been determined through performing correlation analysis with Scikit-learn framework. Correlation coefficient (CC) ranges between -1 (perfect negative correlation) and +1 (perfect positive correlation).

- 0.0 - 0.1: very weak and non-relation
- 0.1 - 0.3: weak relation
- 0.3-0.5: medium relation
- 0.5- 0.7: strong relation
- 0.7-0.9: very strong relation
- 0.9-1.0: perfect relation

Accordingly, absolute value of correlation coefficient is below 0.1 for 27 out of 154 observations. The non-correlated parameters are:

- Maximum Monthly precipitation mm Manual (August)
- Number of cloudy days (April)
- Number of days with a minimum temperature of 5 C and below (April)
- Maximum temperature (June)
- Average humidity % (August)
- Number of days with clear sky (April)
- Monthly total precipitation mm (April)
- Monthly total precipitation mm (July)
- Minimum temperature (August)
- Average daily minimum temperature (May)
- Average minimum temperature (May)
- Monthly average of total daily sunshine time (September)
- Monthly average wind speed (March)
- Monthly maximum precipitation amount mm (July)
- Average minimum temperature (April)
- Average daily minimum temperature (April)
- Monthly maximum precipitation amount mm (October)
- Number of cloudy days (May)

- Monthly total precipitation mm (September)
- Number of rainy days (July)
- Average minimum temperature (June)
- Average daily minimum temperature (June)
- Average daily minimum temperature (October)
- Number of snowy days (May)
- Number of days with strong wind (October)
- Monthly maximum precipitation amount mm (April)
- Monthly average wind speed (October)

The correlation coefficient is slightly over 0.6 for only one parameter (Monthly average of total daily sunshine time -October), which is followed by Average Relative Humidity (June) and Number of Days with Frost (March). Average Soil Temperature (-20cm) parameters (August-September-June) are negatively correlated with potatoes yield, and the absolute coefficient changes between 0.3-0.4 for most of the other parameters (Figure 4). The results indicate that meteorological conditions have only somewhat important but not decisive effects on potato yield in Niğde.

R-square is a statistical metric that gives the proportion of the variance in the dependent target variable explained by the independent variables employed in the model. It changes between 0 and 1. If R-square is 0, it means that the model explains none of the variance and provides no useful information for prediction. On the other hand, if R-square is 1, it means that the models can explain all of the variance and provides perfect information for prediction (Shanmugavalli and Ignatia, 2025). The interpretation of this metric is highly depended on the field of research. It is expected to be above 0.9 for research fields where precise prediction is required like physics and mathematics, while the values between 0.5-0.7 may be found acceptable in social sciences and the values between 0.3-0.5 could be reasonable for studies that contain chaotic data such as human behavior.

In this study, the highest r-square (0.678) is obtained with CNN Deep Learning model, followed by Gradient Boosting (0.542). Both values are acceptable to some degree considering the chaotic nature of weather data. Despite this fact, the models cannot provide precise prediction for potato yield, which indicates that potato production in Niğde is not so much dependent on weather. In fact, potato production is mostly done with irrigation (Sen and Gungor, 2019). In the past 20 years, a number of Ponds have been built in the region (Kocalar, 2022) and many projects have been instigated to increase effective use of water. Potato production has always been profitable in Niğde; however, a recent study reports that absolute and relative profitability degrees have shown a tendency to decrease with the increase in production costs (Karsan and Gul, 2017). Dams are crucial for water resource management. They provide a stable water supply and when precipitation patterns are becoming increasingly unpredictable, dams can store water during periods of excess rainfall and release it during droughts, which helps to maintain a reliable water supply (Selek et al., 2016). Kocalar (2022) reported that water scarcity is increasingly spreading in the region and the groundwater level in Niğde province has been falling to much lower levels each year. Moreover, there are increasing signs of pollution in the irrigation ponds caused by industry and construction, which threatens agricultural practices in the Niğde.

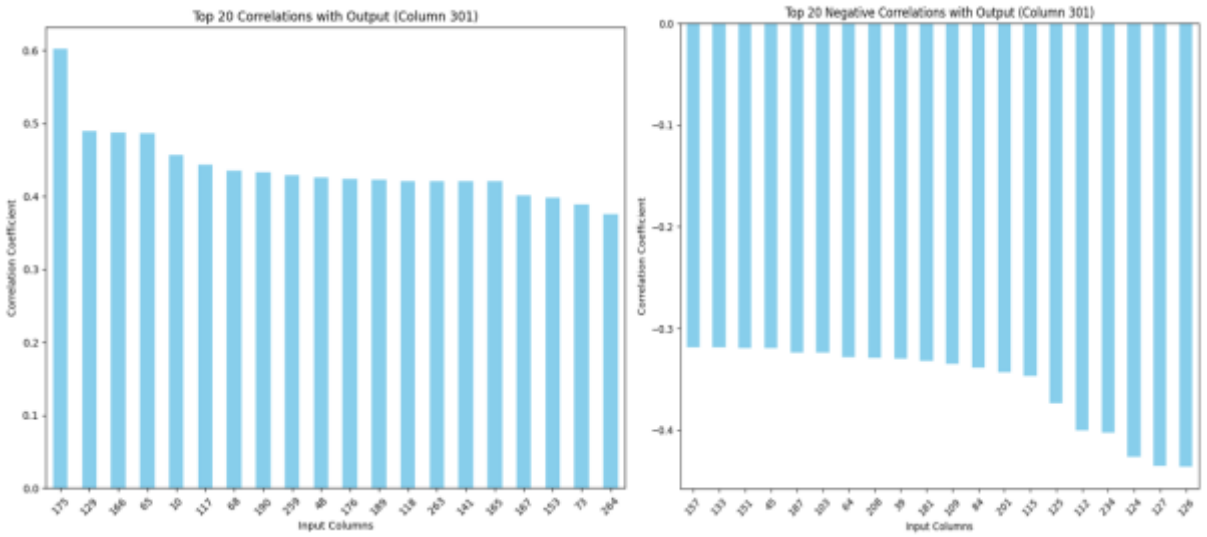

Figure 4. Top 20 positive and negative correlations.

Table 3. Yield estimation analysis results

Metrics	Linear Regression	Random Forest	Gradient Boosting	CNN Deep Learning	RNN Deep Learning
MAE	125.83	127.67	82.98	79.39	109.67
MSE	35570.43	30401.47	39613.09	24821.38	29647.39
R^2	0.202	0.318	0.542	0.678	0.481

4. Conclusion

The study includes all meteorological parameters that have been observed in Niğde province and covers a long period of time (33 years). In this regard, the study is quite comprehensive. We have found that potatoes yield is positively correlated to such meteorological parameters as Sunshine duration and relative humidity and negatively correlated to soil temperature at 20 cm depth. However, the absolute correlation coefficient mostly changes between 0.3 and 0.4 for most of the parameters. This restricts the performance of models in predicting potatoes yield. The reason is that potatoes production is heavily dependent on irrigation in the region. However, the recent findings of other studies producers in the regions could not find enough resources of irrigation in the near future as it is today because the groundwater resources recede to deeper levels each year, which could result in sinkholes that are seen in Konya a neighboring province to Niğde. Furthermore, irrigation ponds built in the regions are threatened by the pollution caused by industry and construction. Green deal agreement is another possible threat to potatoes producers. Water footprint is an important component of the agreement, and certain restrictions could be posed if the irrigation resources are not improved and ameliorated in the Niğde. Climate change will make the endeavors even harder in the future because the projections show severe drought in the region.

Author Contributions

The percentage of the author(s) contributions is presented below. All authors reviewed and approved the final version of the manuscript.

	E.U.	C.Ö.
C	80	20
D	60	40
S	60	40
DCP	80	20
DAI	60	40
L	90	10
W	70	30
CR	50	50
SR	20	80

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

FAOSTAT. 2022. Crop production database. URL= https://www.fao.org/faostat/en/#data/QCL (accessed date: August 18, 2024).

Karsan A, Gul M. 2017. Patates üretim maliyetleri ve karlılığındaki değişim: Niğde ili örneği. Turkish J Agri Food Sci Technol, 5(5): 530-535. https://doi.org/10.24925/turjaf.v5i5.530-535.1029

Kocalar AC. 2022. Niğde su varlığındaki değişimin sürdürülebilirliği-Akkaya Baraj Göleti. E-J Vocat Colleges, 12(2): 76-90.

MEVBIS. 2024. Meteorological Data Service. URL: https://mevbis.mgm.gov.tr/ (accessed date: August 18, 2024).

Selek B, Demirel Yazici D, Aksu H, Özdemir AD. 2016. Seyhan Dam, Turkey, and climate change adaptation strategies. In: Tortajada C. (eds) Increasing resilience to climate variability and change. Water resources development and management. Springer, Singapore, pp: 205-231. https://doi.org/10.1007/978-981-10-1914-2_10

Sen B, Güngör S. 2019. Effects of climate change over potato (Solanum tuberosum L.) production. 4th International Congress on Biosystems Engineering (ICOBEN2019), September 24-27, Hatay, Türkiye, pp: 263-271.

Shanmugavalli M, Ignatia KM. 2023. Comparative study among MAPE, RMSE and R square over the treatment techniques undergone for PCOS influenced women. Rec Patents Eng, 19: e041223224190. https://doi.org/10.2174/0118722121269786231120122435

TUIK. 2023. Bitkisel üretim istatistikleri. URL= https://biruni.tuik.gov.tr/ (accessed date: August 18, 2024).