Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 666-684, 11, 2024

Araştırma Makalesi

IJANSER
https://as-proceeding.com/index.php/ijanser
ISSN:2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 666-684, 11, 2024

Research Article

Development of Explanatory Artificial Intelligence Techniques for Automatic Detection of Factors Affecting Different Population Levels of Almond Leaf Bee *Cimbex quadrimaculata* Müller, 1766 (Hymenoptera: Cimbicidae)

Şemsettin Cebeci[1], İnanç ÖZGEN[1*] and Yunus Güral[2]

[1]*Department of Bioengineering, Engineering Faculty, Fırat University, 23119, Elazığ, Türkiye*
[2]*Department of Statistics, Science Faculty, Fırat University, 23119, Elazığ, Türkiye*

[*]*Email of corresponding author: inancozgen@gmail.com*

(Received: 11 December 2024, Accepted: 29 December 2024)

(5th International Conference on Scientific and Academic Research ICSAR 2024, December 23-24, 2024)

ATIF/REFERENCE: Cebeci, Ş., Özgen, İ. & Güral, Y. (2024). Development of Explanatory Artificial Intelligence Techniques for Automatic Detection of Factors Affecting Different Population Levels of Almond Leaf Bee *Cimbex Quadrimaculata* Müller, 1766 (Hymenoptera: Cimbicidae). *International Conference on Scientific and Academic Research*, 8(11), 666- 684.

Abstract – In this study; data sets of factors affecting different population levels of almond leaf bee *Cimbex quadrimaculata* (Hymenoptera: Cimbicidae) were taken from the almond fields of Nimri village of Keban district of Elazığ province and Eğil district of Diyarbakır province in 2020, 2021 and 2022. In the study; the independent variables affecting the population levels of the pest were altitude, temperature, humidity, wind intensity, tree age, tree height, tree crown projection, and the dependent variables were the larval level. The success classification values of Random Forest and Decision Tree algorithms were used. Larval numbers were graded as slightly contaminated, moderately contaminated, contaminated and highly contaminated with scale values of 0-3, 3-10, 10-30 and over 30 larvae per tree. These data were then analysed using decision tree algorithms from machine learning classification models. The analyses were interpreted according to Random Forest and Decision Tree algorithms in JASP 0.19.1 version programme. According to the results of the study; it was determined that the most important factor affecting the larval level in cultivated almonds in Eğil district was temperature, followed by humidity. The least influencing factors were tree height and tree crown projection. In the wild almond trees of Eğil District, humidity was the most important factor affecting the larval level, followed by temperature. In the cultivated almond trees of Keban District of Elazığ Province, it was observed that the factor affecting the larval level most was temperature, followed by humidity. In Wild Almond Trees of Keban District of Elazığ Province, it was observed that the factor affecting the larval level most was temperature, followed by humidity. In all parameters, the factors that least affect the larval population level were removed from the application in the data analysis programme and the Random Forest algorithm was applied again. In this application, the independent variables were altitude, temperature, humidity, wind intensity and tree age. As a result of the re-application of the algorithm, it was observed that the accuracy value and other values supporting this value increased. In the comparison of wild and cultivated almond classification values between locations, it was determined that the factors affecting the larvae level most in the application of success classification value determination of decision trees of wild almond trees in Eğil district of Diyarbakır province and Keban district of Elazığ province were determined as temperature,

humidity, altitude, tree age and wind intensity, respectively. In the application of determination of success classification value of decision trees of cultivated almond trees in Eğil district of Diyarbakır province and Keban district of Elazığ province, it was determined that the factors affecting the larval level most were temperature, humidity, tree age, altitude and tree crown projection, respectively. In the application of decision trees success classification value determination for cultivated almond trees in Eğil district of Diyarbakır province and Keban district of Elazığ province, it was determined that the factors affecting the larval level most were temperature, humidity, tree age, altitude and tree crown projection, respectively. In the success classification of wild and cultivated almonds between locations (Keban/Eğil), it was determined that the factors affecting the larvae level most in the decision tree success classification value determination application were temperature, humidity, tree age, altitude and wind intensity, respectively. As a result of all algorithm applications, temperature and humidity were found to be the factors affecting the larval level the most. The factors affecting the larval level the least were found to be tree height and tree crown projection.

Keywords – Cimbex quadrimaculata, Population, Artificial Intelligence Techniques

I. INTRODUCTION

Almond is an important cultivated plant for our country. Approximately 25% of our country's almond production is made in the Eastern and Southeastern Anatolia Region [1]. There are important diseases and pests that affect yield and quality in almonds. One of the most important of these pests is *Cimbex quadrimaculata* Müller, 1766 (Hymenoptera: Cimbicidae) (Figure 1).

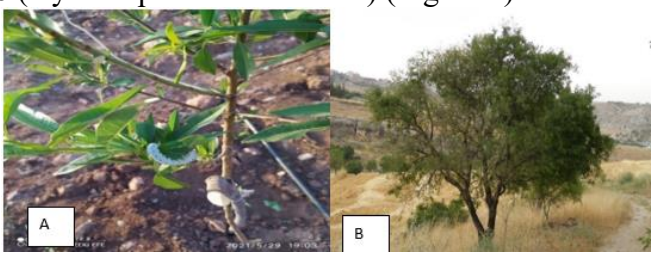

Figure 1. Almond leaf bee larva (A) and its damage (B)

To develop strategies for combating this pest, the TÜBİTAK 3001 project was conducted between 2020 and 2022 [2]. Within the scope of this project, data on the pest's biology, population fluctuations, natural enemies, and control measures were collected [3,4,5]. It was determined that the pest formed populations at different levels in Diyarbakır and Elazığ Provinces, where its populations were monitored [5,6]. In this study's dataset, data from native and cultivated almond trees in Elazığ and Diyarbakır Provinces were used, and population tracking was conducted. Critical factors affecting populations were analyzed using explainable artificial intelligence techniques [7,8]. Detailing these methods and determining the pest's impact on critical damage thresholds is essential for establishing critical control thresholds for this pest in cultivated and wild almond trees. In this thesis, factors influencing pest population development in both almond varieties were analyzed, and the ecological factor components affecting population fluctuations were examined using the results from the TÜBİTAK 3001 project dataset [1].

In other words, determining critical control thresholds through the use of explainable artificial intelligence techniques to identify critical population levels is of significant importance. Within the scope of the project, hybrid optimization methods based on mathematics and biology were adapted as classification analysis models for the targeted problem for the first time.

To this end, Random Forest and Decision Tree Algorithms, which are machine learning classifiers, were applied to determine the significance ranking of factors affecting the pest's population levels. The JASP 0.19.1.0 program was utilized to implement these classifiers.

As a result of this study, interpretations essential for "Forecasting and Early Warning" systems in pest management were made. Data on climatic factors affecting population levels—temperature, humidity, wind intensity, elevation, tree height, age, and canopy projection—were evaluated. Using the significance rankings of these factors, contributions were made to the predictive modeling of almond IPM (Integrated Pest Management) practices.

.

II. MATERIAL AND METHOD

Materials

The materials of this study consist of *Cimbex quadrimaculata,* climatic data obtained from various locations, and the JASP 0.19.1.0 software package.

JASP is an advanced, comprehensive, and effective program used for statistical analyses and data visualizations. When examining the results of the comprehensive and carefully planned application conducted using the JASP program for machine learning classification, the process of defining the dataset and identifying its various characteristics was meticulously carried out step by step. Following this, two different and effective models, Random Forest and Decision Tree, widely used in the field of machine learning, were developed. These models were systematically employed to accurately label examples within the dataset, successfully executing complex and multidimensional classification processes.e results obtained were analyzed and interpreted in detail using various appropriate metrics to more effectively evaluate model performance. Thus, the findings provided in-depth and comprehensive insights into the classification capabilities of the models.

Methods

The datasets evaluating the different population levels of the almond leaf bee, Cimbex quadrimaculata (Hymenoptera: Cimbicidae), were obtained from the TÜBİTAK-supported 3001 project numbered 1180124. The data analyzed for pest larval populations cover the period from the initial appearance of the larvae in nature to the stage when the larvae pupate and fall to the ground, during the years 2020, 2021, and 2022. These data were collected from wild and cultivated almond trees located in Nimri village, Keban district, Elazığ Province, and Eğil district, Diyarbakır Province (Figure 2).

Figure 2. Population monitoring of Cimbex quadrimaculata and its harmful larvae

The factors affecting the population fluctuations of the pest were analyzed using explainable artificial intelligence techniques. These data are presented in Table 1.

Table 1. Data sets used in the study

Locations	Elazığ (Keban/Nimri)	Diyarbakır (Eğil)
Wild / Cultivated	X	X
Temperature	X	X
Altitude	X	X
Humidity	X	X
Wind intensity	X	X
Tree age	X	X
Tree height	X	X
Habitat	X	X
Tree crown projection	X	X
Larva level	X	X
Yield	X	X

Detailing these methods and determining their effects on the critical damage thresholds of the pest are important for establishing critical control thresholds against this pest in cultivated and wild almond trees. In evaluating the factors affecting the pest's population, scale values determined based on field damage observations of the pest were used (Table 2).

Table 2. Scale Rating of Larvae Number

Larval numbers	Scale Number	Scale Definition (Loss Rating)
0-3	1	Slightly contaminated
3-10	2	Moderately contaminated
10-30	3	Contaminated
30-	4	Highly contaminated

Based on these scale values, the data were graded according to damage infestation rates. These data were then analyzed using decision tree algorithms, one of the machine learning classification models. The analyses were conducted using the JASP 0.19.1.0 software version, employing the Random Forest and Decision Tree algorithms. The success classification values of the decision tree algorithms were evaluated using seven different criteria. These criteria are presented in Table 3 below.

Table 3. Success classification values of decision tree algorithms

Accuracy ACC
Precision (Positive Predictive Value PPV)
Recall (True Positive Rate TPR)
False Positive Rate FPR
F1 Score
Matthews Correlation Coefficient MCC
Area Under Curve (AUC)

These criteria were analyzed and evaluated based on the accuracy percentage classification values of different datasets for cultivated and wild almond trees in Eğil district, Diyarbakır Province, and cultivated and wild almond trees in Keban district, Elazığ Province, within the applied algorithms.

III. RESULTS AND DISCUSSION

The findings were evaluated based on location, almond variety, and year, and decision tree algorithms were derived:

1-Evaluation of Factors Affecting Different Population Levels of the Almond Leaf Bee (Cimbex quadrimaculata Hymenoptera: Cimbicidae) in Cultivated Almond Trees in Eğil District, Diyarbakır Province, According to Decision Tree Algorithms

The independent variables affecting the pest population included elevation, temperature, humidity, wind intensity, tree age, tree height, and tree canopy projection. The dependent variable was the larval level. The success classification values of the Random Forest and Decision Tree algorithms are presented in Table 4.

Table 4. Success classification values of decision tree algorithms

	Random Forest	Decision Tree
Accuracy ACC	0.780	0.764
Precision (Positive Predictive Value PPV)	0.683	0,708
Sensitivity (True Positive Rate TPR)	0.671	0.646
Specificity (False Positive Rate FPR)	0.203	0.195
F1 Score	0.671	0.654
Matthews Correlation Coefficient MCC	0.314	0.428
ROC Area Under Curve (AUC)	0.846	0.713

Upon examining Table 4, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.780, indicating successful classification. The precision value is 0.683, the F1 score is 0.671, the MCC value is 0.314, and the ROC area value is 0.846.

In the Decision Tree algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.764, indicating successful classification. The precision value is 0.708, the F1 score is 0.654, the MCC value is 0.428, and the ROC area value is 0.713.

Between the Random Forest and Decision Tree algorithms, the Random Forest classification algorithm is seen to be the better decision tree in terms of accuracy determination (Table 4).

Since the Random Forest algorithm has higher average decrease in accuracy values during the decision-making process, the importance of the independent variable in affecting the dependent variable increases (Figure 3).

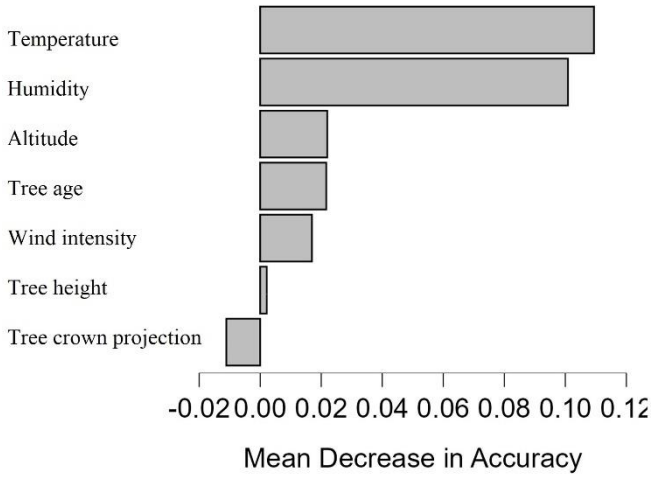

Figure 3. Mean decrease in accuuracy values

It is observed that the factor that most affects the larval level is temperature, followed by humidity (Figure 3). The factors that least affect the larval level are tree height and tree canopy projection. The factors least affecting the larval population level were removed from the data analysis program, and the Random Forest algorithm was applied again. In this application, the independent variables were altitude, temperature, humidity, wind intensity, and tree age. The dependent variable was taken as the larval level (Table 5).

Table 5. Successful classification values of random forest decision tree algorithms

	Random Forest 2
Accuracy ACC	0.886
Precision (Positive Predictive Value PPV)	0,847
Sensitivity (True Positive Rate TPR)	0.829
Specificity (False Positive Rate FPR)	0.090
F1 Score	0.833
Matthews Correlation Coefficient MCC	0.728
ROC Area Under Curve (AUC)	0.921

Upon examining Table 5, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.886, indicating successful classification. The precision value is 0.847, the F1 score is 0.833, the MCC value is 0.728, and the ROC area value is 0.921.

After removing the factors that least affect the larval level and reapplying the algorithm, it was observed that the accuracy value and other supporting values also increased (Figure 4).

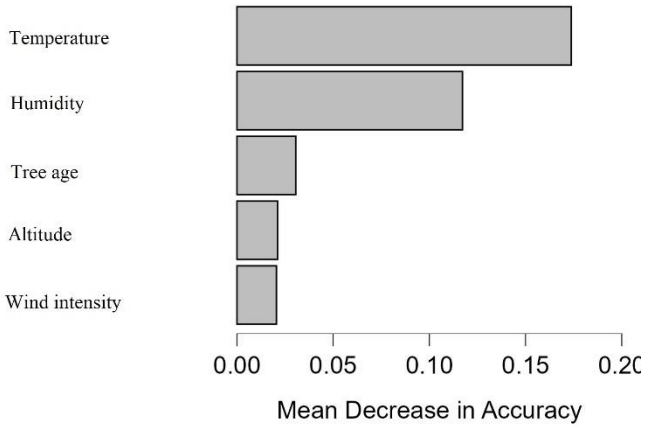

Figure 4. Mean decrease in accuuracy values

In the decision tree classification application for the culture almond trees in Diyarbakır Province, Eğil District, the factors that most affect the larval level were identified as temperature, humidity, tree age, altitude, and wind intensity, in that order (Figure 4).

2-Evaluation of Factors Affecting Different Population Levels of the Almond Leaf Bee (Cimbex quadrimaculata Hymenoptera: Cimbicidae) in Wild Almond Trees in Eğil District, Diyarbakır Province, According to Decision Tree Algorithms

The independent variables affecting the pest population included elevation, temperature, humidity, wind intensity, tree age, tree height, and tree canopy projection. The dependent variable was the larval level. The success classification values of the Random Forest and Decision Tree algorithms are presented in Table 6.

Table 6. Success classification values of decision tree algorithms

	Random Forest	Decision Tree
Accuracy ACC	0.848	0.805
Precision (Positive Predictive Value PPV)	0.689	0,713
Sensitivity (True Positive Rate TPR)	0.695	0.610
Specificity (False Positive Rate FPR)	0.117	0.147
F1 Score	0.677	0.552
Matthews Correlation Coefficient MCC	0.509	0.396
ROC Area Under Curve (AUC)	0.873	0.658

Upon examining Table 6, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.848, indicating successful classification. The precision value is 0.689, the F1 score is 0.677, the MCC value is 0.509, and the ROC area value is 0.873.

In the Decision Tree algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.805, indicating successful classification. The precision value is 0.713, the F1 score is 0.552, the MCC value is 0.396, and the ROC area value is 0.658.

Between the Random Forest and Decision Tree algorithms, the Random Forest classification algorithm is observed to be a better decision tree in terms of success in determining accuracy (Table 6). The Random Forest algorithm, due to its higher average decrease in accuracy values during the decision-making process, increases the importance of the independent variables in affecting the dependent variable (Figure 5).

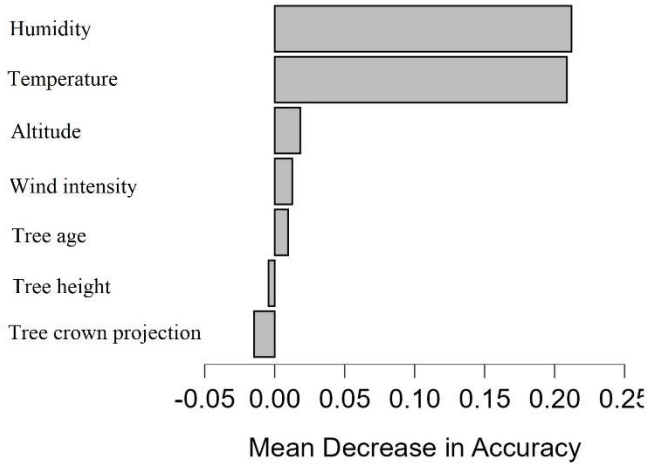

Figure 5. Mean decrease in accuuracy values

It is observed that the factor that most affects the larval level is humidity, followed by temperature (Figure 5). The factors that least affect the larval level are tree height and tree canopy projection. After removing the factors least affecting the larval population level from the data analysis program, the Random Forest algorithm was reapplied. In this application, the independent variables were altitude, temperature, humidity, wind intensity, and tree age. The dependent variable was taken as the larval level (Table 7).

Table 7. Successful classification values of random forest decision tree algorithms

	Random Forest 2
Accuracy ACC	0.884
Precision (Positive Predictive Value PPV)	0,773
Sensitivity (True Positive Rate TPR)	0.768
Specificity (False Positive Rate FPR)	0.086
F1 Score	0.750
Matthews Correlation Coefficient MCC	0.499
ROC Area Under Curve (AUC)	0.956

Upon examining Table 7, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.884, indicating successful classification. The precision value is 0.773, the F1 score is 0.750, the MCC value is 0.499, and the ROC area value is 0.956.

After removing the factors that least affect the larval level and reapplying the algorithm, it was observed that the accuracy value and the other supporting values also increased (Figure 6).

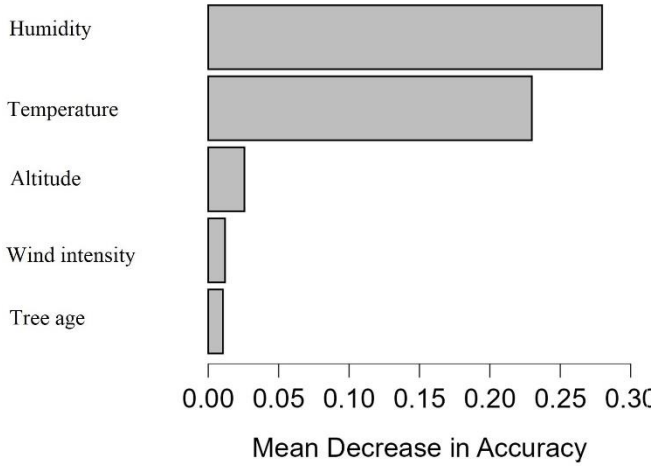

Figure 6. Mean decrease in accuuracy values

In the decision tree classification application for the wild almond trees in Diyarbakır Province, Eğil District, the factors that most affect the larval level were identified as humidity, temperature, altitude, wind intensity, and tree age, in that order (Figure 6).

3-Evaluation of Factors Affecting Different Population Levels of the Almond Leaf Bee (Cimbex quadrimaculata Hymenoptera: Cimbicidae) in Cultivated Almond Trees in Keban District, Elazığ Province, According to Decision Tree Algorithms

The independent variables affecting the pest population included elevation, temperature, humidity, wind intensity, tree age, tree height, and tree canopy projection. The dependent variable was the larval level. The success classification values of the Random Forest and Decision Tree algorithms are presented in Table 8.

Table 8. Success classification values of decision tree algorithms

	Random Forest	Decision Tree
Accuracy ACC	0.819	0.735
Precision (Positive Predictive Value PPV)	0.814	0,741
Sensitivity (True Positive Rate TPR)	0.819	0.735
Specificity (False Positive Rate FPR)	0.243	0.303
F1 Score	0.813	0.738
Matthews Correlation Coefficient MCC	0.551	0.385
ROC Area Under Curve (AUC)	0.742	0.697

Upon examining Table 8, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.819, indicating successful classification. The precision value is 0.814, the F1 score is 0.813, the MCC value is 0.551, and the ROC area value is 0.742.

In the Decision Tree algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.735, indicating successful classification. The precision value is 0.741, the F1 score is 0.738, the MCC value is 0.385, and the ROC area value is 0.697.

Between the Random Forest and Decision Tree algorithms, the Random Forest classification algorithm is observed to be a better decision tree in terms of success in determining accuracy (Table 8).

The Random Forest algorithm, due to its higher average decrease in accuracy values during the decision-making process, increases the importance of the independent variables in affecting the dependent variable (Figure 7).

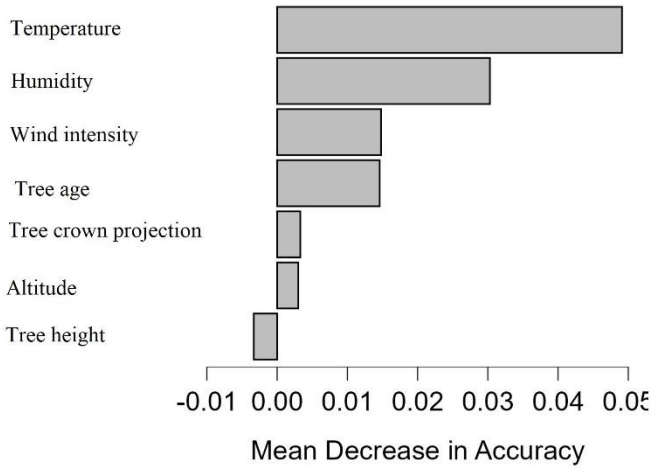

Figure 7. Mean decrease in accuuracy values

It is observed that the factor that most affects the larval level is temperature, followed by humidity (Figure 7). The factors that least affect the larval level are tree height and altitude. After removing the factors least affecting the larval population level from the data analysis program, the Random Forest algorithm was reapplied. In this application, the independent variables were temperature, humidity, wind intensity, tree age, and tree canopy projection. The dependent variable was taken as the larval level (Table 9).

Table 9. Successful classification values of random forest decision tree algorithms

	Random Forest 2
Accuracy ACC	0.819
Precision (Positive Predictive Value PPV)	0,812
Sensitivity (True Positive Rate TPR)	0.819
Specificity (False Positive Rate FPR)	0.259
F1 Score	0.812
Matthews Correlation Coefficient MCC	0.523
ROC Area Under Curve (AUC)	0.786

Upon examining Table 9, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.819, indicating successful classification. The precision value is 0.812, the F1 score is 0.812, the MCC value is 0.523, and the ROC area value is 0.786.

After removing the factors that least affect the larval level and reapplying the algorithm, it was observed that the accuracy value and the other supporting values also increased (Figure 8).

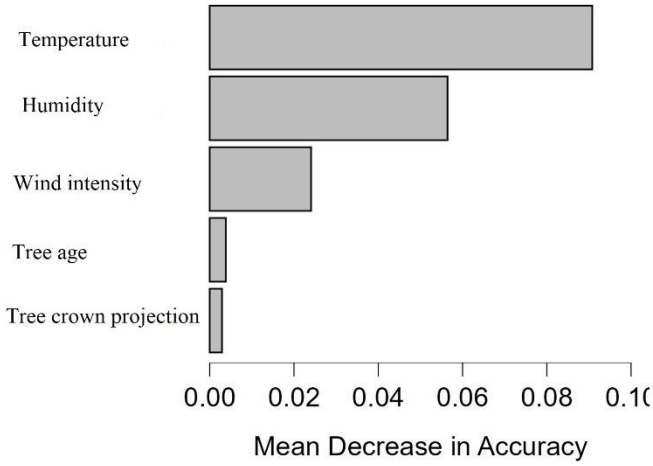

Figure 8. Mean decrease in accuuracy values

In the decision tree classification application for the cultivated almond trees in Elazığ Province, Keban District, the factors that most affect the larval level were identified as temperature, humidity, wind intensity, tree age, and tree canopy projection, in that order (Figure 8).

4-Evaluation of Factors Affecting Different Population Levels of the Almond Leaf Bee (Cimbex quadrimaculata Hymenoptera: Cimbicidae) in Wild Almond Trees in Keban District, Elazığ Province, According to Decision Tree Algorithms

The independent variables affecting the pest population included elevation, temperature, humidity, wind intensity, tree age, tree height, and tree canopy projection. The dependent variable was the larval level. The success classification values of the Random Forest and Decision Tree algorithms are presented in Table 10.

Table 10. Success classification values of decision tree algorithms

	Random Forest	Decision Tree
Accuracy ACC	0.780	0.756
Precision (Positive Predictive Value PPV)	0.668	0,642
Sensitivity (True Positive Rate TPR)	0.671	0.634
Specificity (False Positive Rate FPR)	0.180	0.196
F1 Score	0.667	0.613
Matthews Correlation Coefficient MCC	0.441	0.335
ROC Area Under Curve (AUC)	0.868	0.744

Upon examining Table 10, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.780, indicating successful classification. The precision value is 0.668, the F1 score is 0.667, the MCC value is 0.441, and the ROC area value is 0.868.

In the Decision Tree algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.756, indicating successful classification. The precision value is 0.642, the F1 score is 0.613, the MCC value is 0.335, and the ROC area value is 0.744.

Between the Random Forest and Decision Tree algorithms, the Random Forest classification algorithm is observed to be a better decision tree in terms of success in determining accuracy (Table 10).

The Random Forest algorithm, due to its higher average decrease in accuracy values during the decision-making process, increases the importance of the independent variables in affecting the dependent variable (Figure 9).

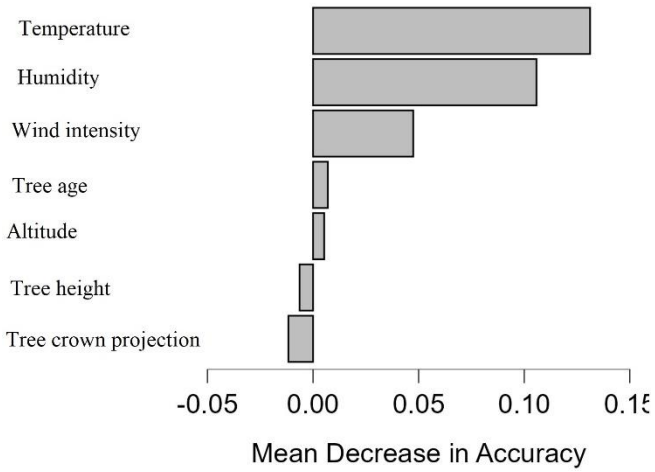

Figure 9. Mean decrease in accuuracy values

It is observed that the factor that most affects the larval level is temperature, followed by humidity (Figure 9). The factors that least affect the larval level are tree height and tree canopy projection. After removing the factors least affecting the larval population level from the data analysis program, the Random Forest algorithm was reapplied. In this application, the independent variables were altitude, temperature, humidity, wind intensity, and tree age. The dependent variable was taken as the larval level (Table 11).

Table 11. Successful classification values of random forest decision tree algorithms

	Random Forest 2
Accuracy ACC	0.821
Precision (Positive Predictive Value PPV)	0,730
Sensitivity (True Positive Rate TPR)	0.732
Specificity (False Positive Rate FPR)	0.141
F1 Score	0.716
Matthews Correlation Coefficient MCC	0.531
ROC Area Under Curve (AUC)	0.921

Upon examining Table 11, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.821, indicating successful classification. The precision value is 0.730, the F1 score is 0.716, the MCC value is 0.531, and the ROC area value is 0.921.

After removing the factors that least affect the larval level and reapplying the algorithm, it was observed that the accuracy value and the other supporting values also increased (Figure 10).

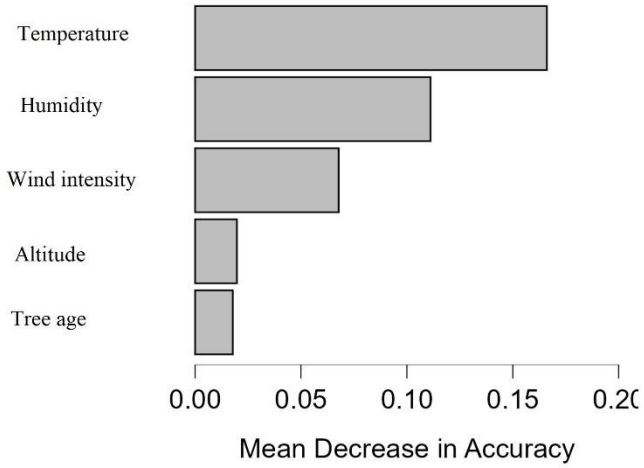

Figure 10. Mean decrease in accuuracy values

In the decision tree classification application for the wild almond trees in Keban District, Elazığ Province, the factors that most affect the larval level were identified as temperature, humidity, wind intensity, altitude, and tree age, in that order (Figure 10).

5-Evaluation of Factors Affecting Different Population Levels of the Almond Leaf Bee (Cimbex quadrimaculata Hymenoptera: Cimbicidae) in Wild Almond Trees in Eğil District, Diyarbakır Province and Keban District, Elazığ Province, According to Decision Tree Algorithms

The independent variables affecting the pest population included elevation, temperature, humidity, wind intensity, tree age, tree height, and tree canopy projection. The dependent variable was the larval level. The success classification values of the Random Forest and Decision Tree algorithms are presented in Table 12.

Table 12. Success classification values of decision tree algorithms

	Random Forest	Decision Tree
Accuracy ACC	0.838	0.826
Precision (Positive Predictive Value PPV)	0.663	0,669
Sensitivity (True Positive Rate TPR)	0.677	0.652
Specificity (False Positive Rate FPR)	0.124	0.132
F1 Score	0.668	0.657
Matthews Correlation Coefficient MCC	0.467	0.348
ROC Area Under Curve (AUC)	0.894	0.676

Upon examining Table 12, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.838, indicating successful classification. The precision value is 0.663, the F1 score is 0.668, the MCC value is 0.467, and the ROC area value is 0.894.

In the Decision Tree algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.826, indicating successful classification. The precision value is 0.669, the F1 score is 0.657, the MCC value is 0.348, and the ROC area value is 0.676.

Between the Random Forest and Decision Tree algorithms, the Random Forest classification algorithm is observed to be a better decision tree in terms of success in determining accuracy (Table 12).

The Random Forest algorithm, due to its higher average decrease in accuracy values during the decision-making process, increases the importance of the independent variables in affecting the dependent variable (Figure 11).

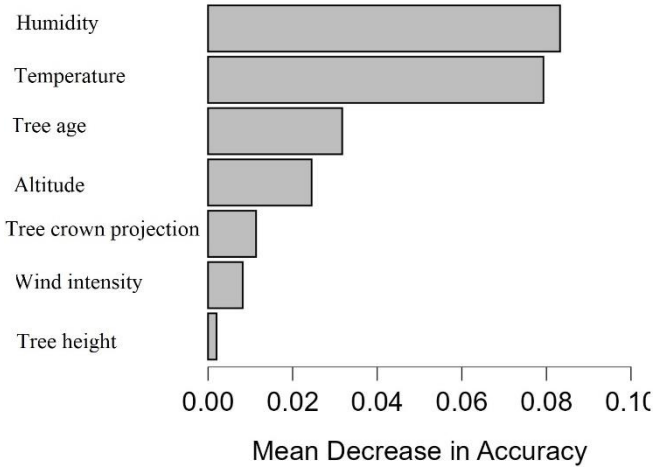

Figure 11. Mean decrease in accuuracy values

It is observed that the factor that most affects the larval level is humidity, followed by temperature (Figure 11). The factors that least affect the larval level are tree height and tree canopy projection. After removing the factors that least affect the larval population level from the data analysis program, the Random Forest algorithm was reapplied. In this application, the independent variables were altitude, temperature, humidity, wind intensity, and tree age. The dependent variable was taken as the larval level (Table 13).

Table 13. Successful classification values of random forest decision tree algorithms

	Random Forest 2
Accuracy ACC	0.863
Precision (Positive Predictive Value PPV)	0,728
Sensitivity (True Positive Rate TPR)	0.726
Specificity (False Positive Rate FPR)	0.098
F1 Score	0.725
Matthews Correlation Coefficient MCC	0.434
ROC Area Under Curve (AUC)	0.925

Upon examining Table 13, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.863, indicating successful classification. The precision value is 0.728, the F1 score is 0.725, the MCC value is 0.434, and the ROC area value is 0.925.

After removing the factors that least affect the larval level and reapplying the algorithm, it was observed that the accuracy value and the other supporting values also increased (Figure 12).

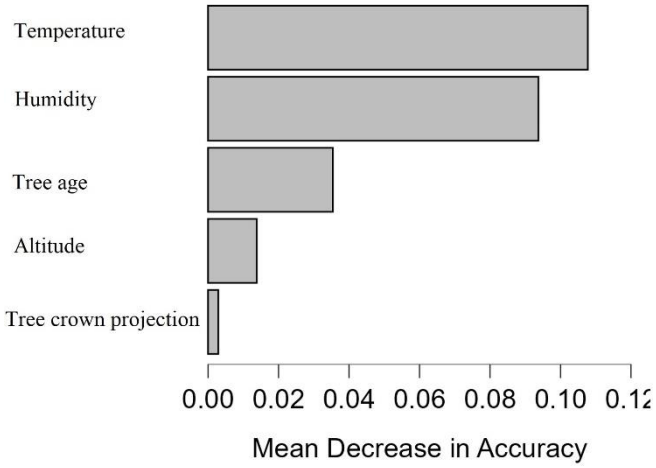

Figure 12. Mean decrease in accuuracy values

In the decision tree classification application for the wild almond trees in Eğil District, Diyarbakır Province, and Keban District, Elazığ Province, the factors that most affect the larval level were identified as temperature, humidity, altitude, tree age, and wind intensity, in that order (Figure 12)

6-Evaluation of Factors Affecting Different Population Levels of the Almond Leaf Bee (Cimbex quadrimaculata Hymenoptera: Cimbicidae) in Cultivated Almond Trees in Eğil District, Diyarbakır Province and Keban District, Elazığ Province, According to Decision Tree Algorithms

The independent variables affecting the pest population included elevation, temperature, humidity, wind intensity, tree age, tree height, and tree canopy projection. The dependent variable was the larval level. The success classification values of the Random Forest and Decision Tree algorithms are presented in Table 14.

Table 14. Success classification values of decision tree algorithms

	Random Forest	Decision Tree
Accuracy ACC	0.810	0.794
Precision (Positive Predictive Value PPV)	0.709	0,699
Sensitivity (True Positive Rate TPR)	0.715	0.691
Specificity (False Positive Rate FPR)	0.195	0.190
F1 Score	0.706	0.692
Matthews Correlation Coefficient MCC	0.495	0.452
ROC Area Under Curve (AUC)	0.813	0.750

Upon examining Table 14, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.810, indicating successful classification. The precision value is 0.709, the F1 score is 0.706, the MCC value is 0.495, and the ROC area value is 0.813.

In the Decision Tree algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.794, indicating successful classification. The precision value is 0.699, the F1 score is 0.692, the MCC value is 0.452, and the ROC area value is 0.750.

Between the Random Forest and Decision Tree algorithms, the Random Forest classification algorithm is observed to be a better decision tree in terms of success in determining accuracy (Table 14).

The Random Forest algorithm, due to its higher average decrease in accuracy values during the decision-making process, increases the importance of the independent variables in affecting the dependent variable (Figure 13).

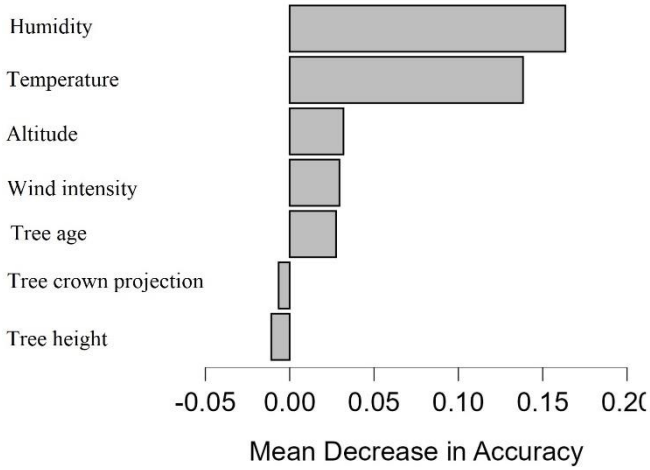

Figure 13. Mean decrease in accuuracy values

It is observed that the factor that most affects the larval level is humidity, followed by temperature (Figure 13). The factors that least affect the larval level are tree height and wind intensity. After removing the factors that least affect the larval population level from the data analysis program, the Random Forest algorithm was reapplied. In this application, the independent variables were altitude, temperature, humidity, tree canopy projection, and tree age. The dependent variable was taken as the larval level (Table 15).

Table 15. Successful classification values of random forest decision tree algorithms

	Random Forest 2
Accuracy ACC	0.859
Precision (Positive Predictive Value PPV)	0,793
Sensitivity (True Positive Rate TPR)	0.788
Specificity (False Positive Rate FPR)	0.150
F1 Score	0.784
Matthews Correlation Coefficient MCC	0.558
ROC Area Under Curve (AUC)	0.864

Upon examining Table 15, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.859, indicating successful classification. The precision value is 0.793, the F1 score is 0.784, the MCC value is 0.558, and the ROC area value is 0.864.

After removing the factors that least affect the larval level and reapplying the algorithm, it was observed that the accuracy value and the other supporting values also increased (Figure 14).

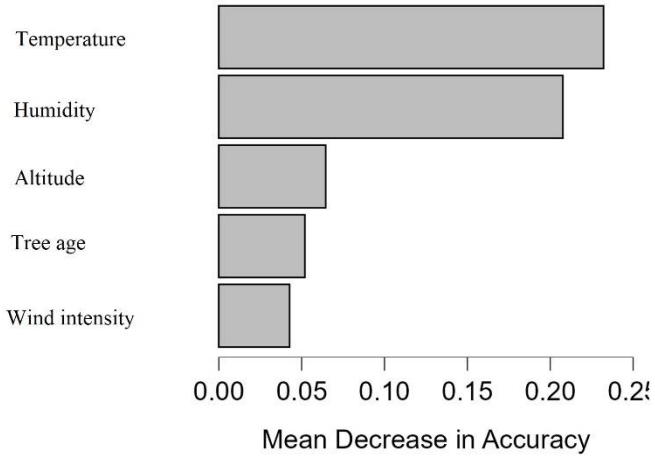

Figure 14. Mean decrease in accuuracy values

In the decision tree classification application for the cultivated almond trees in Eğil District, Diyarbakır Province, and Keban District, Elazığ Province, the factors that most affect the larval level were identified as temperature, humidity, tree age, altitude, and tree canopy projection, in that order (Figure 14)

7-Evaluation of Factors Affecting Different Population Levels of the Almond Leaf Bee (Cimbex quadrimaculata Hymenoptera: Cimbicidae) in Cultivated and Wild Almond Trees in Eğil District, Diyarbakır Province and Keban District, Elazığ Province, According to Decision Tree Algorithms

The independent variables affecting the pest population included elevation, temperature, humidity, wind intensity, tree age, tree height, and tree canopy projection. The dependent variable was the larval level. The success classification values of the Random Forest and Decision Tree algorithms are presented in Table 16.

Table 16. Success classification values of decision tree algorithms

	Random Forest	Decision Tree
Accuracy ACC	0.856	0.850
Precision (Positive Predictive Value PPV)	0.696	0,707
Sensitivity (True Positive Rate TPR)	0.711	0.699
Specificity (False Positive Rate FPR)	0.123	0.130
F1 Score	0.696	0.703
Matthews Correlation Coefficient MCC	0.396	0.351
ROC Area Under Curve (AUC)	0.906	0.681

Upon examining Table 16, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.856, indicating successful classification. The precision value is 0.696, the F1 score is 0.696, the MCC value is 0.396, and the ROC area value is 0.906.

In the Decision Tree algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.850, indicating successful classification. The precision value is 0.707, the F1 score is 0.703, the MCC value is 0.351, and the ROC area value is 0.681.

Between the Random Forest and Decision Tree algorithms, the Random Forest classification algorithm is observed to be a better decision tree in terms of success in determining accuracy (Table 16).

The Random Forest algorithm, due to its higher average decrease in accuracy values during the decision-making process, increases the importance of the independent variables in affecting the dependent variable (Figure 15).

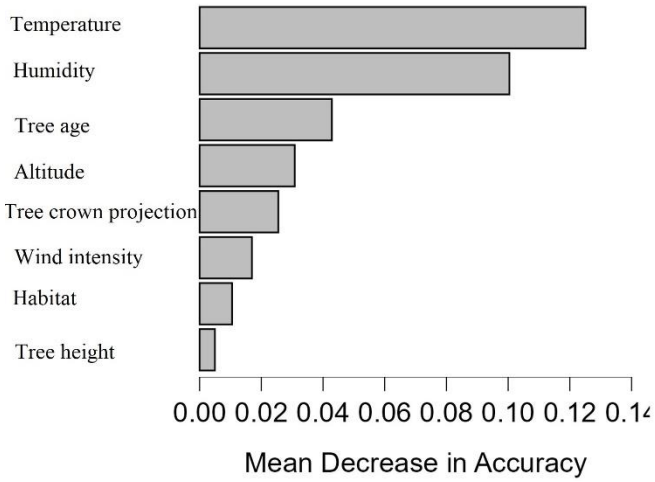

Figure 15. Mean decrease in accuuracy values

It is observed that the factor that most affects the larval level is temperature, followed by humidity (Figure 15). The factors that least affect the larval level are tree height and habitat. After removing the factors that least affect the larval population level from the data analysis program, the Random Forest algorithm was reapplied. In this application, the independent variables were altitude, temperature, humidity, wind intensity, tree canopy projection, and tree age. The dependent variable was taken as the larval level (Table 17).

Table 17. Successful classification values of random forest decision tree algorithms

	Random Forest 2
Accuracy ACC	0.869
Precision (Positive Predictive Value PPV)	0,734
Sensitivity (True Positive Rate TPR)	0.739
Specificity (False Positive Rate FPR)	0.113
F1 Score	0.729
Matthews Correlation Coefficient MCC	0.580
ROC Area Under Curve (AUC)	0.917

Upon examining Table 17, it is observed that in the Random Forest algorithm, the sensitivity value is high, while the specificity value is low. The accuracy value is found to be 0.869, indicating successful classification. The precision value is 0.734, the F1 score is 0.729, the MCC value is 0.580, and the ROC area value is 0.917.

After removing the factors that least affect the larval level and reapplying the algorithm, it was observed that the accuracy value and the other supporting values also increased (Figure 16).

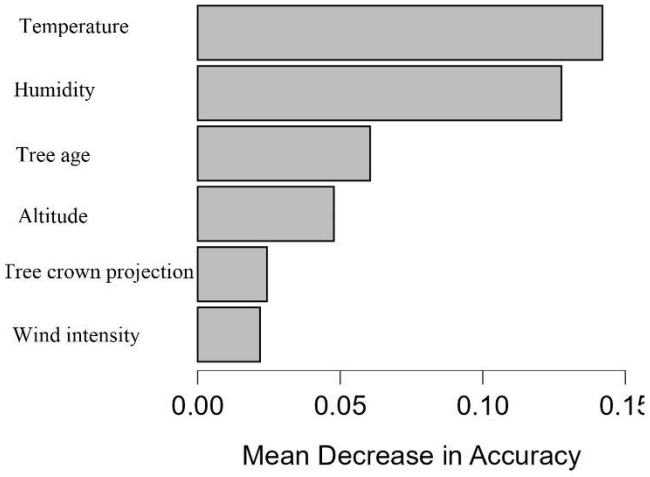

Figure 16. Mean decrease in accuuracy values

In the decision tree classification application for the cultivated and wild almond trees in the Eğil District of Diyarbakır Province and the Keban District of Elazığ Province, the factors that most affect the larval level were identified as temperature, humidity, tree age, altitude, and wind intensity, in that order (Figure 16).

IV. CONCLUSIONS

In conclusion, the factors affecting the larval level of the almond leaf beetle Cimbex quadrimaculata in wild and cultivated almond trees were examined through two locations and a three-year dataset, using the Random Forest and Decision Tree algorithms. The study areas included the cultural and wild almond trees in Diyarbakır's Eğil district, and in the Nimri village of Keban district in Elazığ, where separate applications were conducted.

In the application for cultivated almond trees in Diyarbakır's Eğil district, the classification success was determined to be 0.886. The factors that most influenced the larval level were temperature, humidity, tree

age, elevation, and wind intensity. The least influential factors were tree height and canopy projection.

For wild almond trees in Diyarbakır's Eğil district, the classification success was determined to be 0.884. The most influential factors were humidity, temperature, elevation, wind intensity, and tree age, while the least influential factors were tree height and canopy projection.

For cultivated almond trees in Elazığ's Keban district, the classification success was found to be 0.819. The most influential factors on larval level were temperature, humidity, wind intensity, tree age, and canopy projection, with tree height and elevation being the least influential.

For wild almond trees in Elazığ's Keban district, the classification success was 0.821. The factors most affecting the larval level were temperature, humidity, wind intensity, elevation, and tree age, while tree height and canopy projection had the least impact.

For the wild almond trees in both Diyarbakır's Eğil district and Elazığ's Keban district, the classification success was 0.863. The factors most affecting the larval level were temperature, humidity, elevation, tree age, and wind intensity, while tree height and canopy projection were the least impactful.

For the cultivated almond trees in both Diyarbakır and Elazığ, the classification success was 0.859. The factors that most influenced the larval level were temperature, humidity, tree age, elevation, and canopy projection, while tree height and canopy projection were the least influential.

Finally, for both wild and cultivated almond trees in Diyarbakır's Eğil district and Elazığ's Keban district, the classification success was 0.869. The most influential factors on larval levels were temperature, humidity, tree age, elevation, canopy projection, and wind intensity, while the least influential were tree height and habitat.

The data from this study is significant for evaluating artificial intelligence data in the control of almond pests.

ACKNOWLEDGEMENTS

This study was supported by TUBITAK 118O124 through project number 3001. All authors would like to thank the Scientific Council of TÜBİTAK.

REFERENCES

[1] Özgen, İ. TÜBİTAK 118O124 no'lu 3001 proje sonuç raporu. 126 s.

[2] Bolu, H.,. Distrtibution, Life History and Biology of Almond Sawfly *Cimbex quadrimaculata* (Müller, 1766), Hymenoptera: Cimbicidae). Scientific Papers. Series A. Agronomy, 2016:Vol. LIX: 219-222.

[3] Özgen, İ., Efe, M. Z., Hassan, E. and E. Karaman, Utilization of pomegranate flower phenology in determining the pest management timing of *Cimbex qaudrimaculata* Müller, 1766 (Hymenoptera: Cimbicidae) International Journal of Zoological and Entomological Letters; 2021:1(1): 62-64.

[4] Özgen, İ., Güral, Y. and T. Arslan,. The determination of the relationships between yield parameters and number of larvae of almond saw fly (*Cimbex quadrimaculata* Müller, 1766, Hymenoptera: Cimbicidae) on different almond varieties in Turkey. International Journal of Zoological and Entomological Letters 2022; 2(1): 01-037.

[5] Özgen, İ., Topdemir, A., Baydoğan, G., and E. Karaman, . Population in Almond Sawfly *Cimbex quadrimaculata* Müller, Hymenoptera: Cimbicidae in two Almond Varieties in Diyarbakır and Elazığ Provinces in Turkey. Journal of Research (Science), 2021. Vol. 32, No. 3-4.

[6] Özgen I, Dioli P, Koç I, Topdemir A. Some Heteroptera (Hemiptera) species that are potential natural enemies of *Cimbex quadrimaculata* (Müller, 1766) (Hymenoptera: Cimbicidae). J. Het. Turk,2021b: 3(2): 151–156 (in English).

[7] Özgen I, Efe M. Z, Hassan E, Karaman E. Utilization of pomegranate flower phenology in determining the pest management timing of *Cimbex quadrimaculata* Müller, 1766 (Hymenoptera: Cimbicidae). Int. J. Zool. Entomol. Lett. 2021b: 1(1):62–64 (in English)ang X., Nisar T., Hou Y., Gou X., Sun L., Guo Y. *Pomegranate peel pectin can be used as an effective emulsifier*. Food Hydrocolloids, 2018: 85, 30-38.

[8] Özgen I, Danışman T, Baydoğan G. Spider species feeding on *Cimbex quadrimaculata* (Müller, 1766) (Hymenoptera: Cimbicidae). Acta Entomol Zool. 2022: 3 (1):1–3 (in English).