

COMPARISON OF A STATISTICAL METHOD AND AN ARTIFICIAL INTELLIGENCE APPROACH IN TAXONOMICAL NEMATOLOGY FROM TURKIYE: USING A PAIR OF DETERMINED MORPHOMETRIC PARAMETERS

A. N. Tan¹ and A. Tan^{2*}

¹Program of Landscape and Ornamental Plants, Sakarya University of Applied Sciences, Vocational School of Sapanca, Sakarya, Türkiye

²Department of Geophysical Engineering, Natural Sciences Institute, Sakarya University, Sakarya, Türkiye

*Corresponding author's e-mail: aylin.tan@ogr.sakarya.edu.tr

ABSTRACT

In this study, mono and dual ovaries of plant parasitic nematodes in quince (*Cydonia oblonga* Mill.) (Rosales: Rosaceae) cultivated areas in Sakarya province of Türkiye, were investigated. A total of 230 female nematodes were used, which were obtained from the soil in July 2016 and 2017. The nematode which was examined exhibited the best relationship between the important parameters of the morphometric measurements. The mono and dual ovaries were discriminated by using the linear discriminate function (LDF) method and artificial neural networks (ANNs) approach. The pair of parameters were tried by using LDF method. Then it was observed that the pair of the tail length/tail diameter at anus or cloaca (c) and percentage of the distance of vulva from anterior (V%) parameters had the best correlation with each other considering the highest accuracy percentage obtained as 80% according to the LDF method. The c' versus (V%) of the nematode had a higher classification accuracy percentage for data set than others as 99% for LDF method and 91% for ANNs approach for the July 2016 set. Thus, it can be concluded that LDF method is as successful as ANNs approach.

Keywords: Artificial Neural Networks; Linear Discriminate Function; Nematode; Ovary; Quince

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Published first online April 15, 2023

Published final August 04, 2023

INTRODUCTION

Since plant parasitic nematodes are one of the organism groups with the highest variety of species among all the plant pests, they have priority in our studies. Plant parasitic nematodes are common in crop production areas and can be highly destructive when populations of some species rise above the economic threshold level (Gaugler and Bilgrami, 2004). Today, even with modern agricultural technologies in developed countries, the product loss can be 5-10% due to plant parasitic nematodes (Mitiku, 2018). Because of their microscopic dimensions and the undefined findings of an infection, plant parasitic nematodes live at sheltered areas. For that reason, the agricultural laborers and plant protection experts are needed frequently (Desaeger *et al.*, 2004).

In fact, root-damaging nematodes disrupt the plant's absorption of water and nutrients from the soil and increase the severity of the damage along with soil-borne plant disease factors (Dababat *et al.*, 2015). Hence, symptoms such as yellowing and wilting of the leaves, growth retardation, structural deformations in the root system, impediment of water and nutrient intake from the soil and eventual yield loss are observed at plants

damaged by nematodes (Daramola *et al.*, 2015). Depending on the type of nematode diagnosed, effective pest control method should be applied.

During the historical development of nematology, the systematic of nematodes has been in a dynamic structure that has been constantly changing in time (Ahmed *et al.*, 2015; Siddiqi, 2000). When technical and taxonomic experts use phenotypic characters and their options carefully, precisely and correctly in their morphological studies to identify taxa, identification of the organism can be as good as any biochemical or molecular method.

The morphologic parameters of the plant parasitic nematodes identified in a long process by a taxonomist need not be checked for accuracy percentage, while he or she is diagnosing the nematode specimen. However, the accuracy percentage results obtained by artificial neural networks (ANNs) approach and LDF method can be compared with real values. ANNs approach is one of the machine learning methods that have been widely used in recent years, and it is utilized in solving complex problems in various tasks from parameter and function estimation to classification (Kurtulmus *et al.*, 2020).

Besides, many nematode species were investigated by using ANNs approach (Sundararaju *et al.*, 2002; Akintayo *et al.*, 2018; Aragon *et al.*, 2019; Uhlemann *et al.*, 2020; Tan *et al.*, 2022). As a result of the scientific advances, the importance of the Artificial Intelligence was discovered and its results and the results of studies using LDF method have started to be compared (Keles, 2019; Tan *et al.*, 2022).

At the classical taxonomy, some morphological parameters are used for the determination of the ovary types of the female nematodes. In this study, all female nematodes have been determined according to the mono ovary and dual ovary in a population together. For that reason, it may cause some probable errors in scientific classification studies. To determine the real number of female nematodes in the study area, they should be diagnosed truly. Many different methods are available in the literature about discrimination of the population in the classical taxonomy. Therefore, some pairs of parameters that determine the type of the ovary needed to be tested due to the separation of the most important morphometric parameters by the taxonomical expert. The purpose of this study was the discrimination of mono and dual ovary by applying LDF method and ANNs approach. For the discrimination of the two groups according to the chosen pair of parameters, LDF method and ANNs approach have been used at the studies on earth sciences, too (Tan, 2021; Tan *et al.*, 2021a; Tan *et al.*, 2021b).

For that reason, we decided to compare the accuracy percentage results of some parameter pairs of LDF method and ANNs approach first. Calculated values of accuracy percentage were checked to define the best pair of parameters for presentation of the ovaries. Joint with some interdisciplinary departments, this may improve the quality of the population, helping to better identify errors in nematode taxonomy research.

In this study, soil samples were collected from quince (*C. oblonga* Mill.) growing areas in July 2016 and July 2017 in Pamukova and Geyve districts in Sakarya (Turkey) and were examined. With 174,038 tons of production and 6,568 ha of production area, Turkey is the homeland and the leader of quince production in the world and provides about 20% of world production (<http://faostat3.fao.org/>). Sakarya province ranks the first in quince production with 102,476 tons and constitutes 59% of our country's quince production (<http://www.tuik.gov.tr/>). Geyve Quince, symbolized by Geyve on June 17, 2020, was registered by the Turkish Patent and Trademark Office with geographical indication (Akal *et al.*, 2020). Geyve Quince, which has such a big economic importance for Sakarya province, needs to be cultivated healthily by farmers. It also needs to be controlled by the agricultural engineers regularly. That is because quince is exposed to various pests and diseases including soil-borne factors such as plant parasitic nematodes. Otherwise, the quince production

would affect the agricultural economy of Sakarya province negatively.

MATERIALS AND METHODS

In this study, a total of 50 soil samples were taken from quince cultivation areas in Geyve and Pamukova in Sakarya in July 2016 and July 2017 in the region bounded by 39.48-40.00°N and 30.03-30.21°E. A total of identified 230 female plant parasitic nematodes were diagnosed as *Helicotylenchus tunisiensis* Siddiqi, 1963 (Tylenchida: Haplolaimidae), *Merlinius brevidens* (Allen, 1955) Siddiqi, 1970 (Tylenchida: Belonolaimidae), *Pratylenchoides alkani* Yüksel, 1977 (Tylenchida: Pratylenchidae), *Rotylenchulus borealis* Loof and Oostenbrink, 1962 (Tylenchida: Haplolaimidae) and *Scuttylenchus quettensis* Maqbool, Ghazala and Fatima, 1984 (Tylenchida: Belonolaimidae) for the plant parasitic nematode species with dual ovary and *Boleodorus (B.) thyllactus* Thorne, 1941 (Tylenchida: Tylenchidae), *Irantylenchus clavidorus* Kheiri, 1972 (Tylenchida: Tylenchidae), and *Ditylenchus destructor* Thorne, 1945 (Tylenchida: Anguinidae) for the plant parasitic nematode species with mono ovary from these soil samples (Siddiqi, 2000) (Fig. 1).

In this study, some parameters such as overall body length (L), spear length (stylet), percentage of the distance from anterior to median bulb relative to length of esophagus (MB%), percentage of the length of male gonad relative to body length (T), distance from vulva to anus (VA), (T/VA), percentage of the length of anterior female gonad in relation to body length (G1), portion of body from anus or cloaca to posterior terminus (tail), body length / greatest body diameter (a), body length / distance from anterior to esophago-intestinal valve (b), body length / distance from anterior to base of esophageal glands (b'), body length / tail length (c), tail length / tail diameter at anus or cloaca (c'), conus of stomatostyle/total stomatostyle length (m), (distance of dorsal esophageal gland opening from stylet knobs x 100)/(length of stylet) (o) and percentage of the distance of vulva from anterior (V%) were used (De Man, 1880). And then the results were compared with each other using LDF method before the percentage accuracy was calculated using ANNs approach for the classification of mono and dual ovaries. L versus other pairs of parameters (as stylet, MB%, (T/VA), G1, the tail, a, b, b', c, c', m, o and V%, respectively) enabled the determination of LDF method using Statistical Package for the Social Sciences (SPSS) Analysis Program to discriminate mono and dual ovaries of this population (SPSS, 2005). Before ANNs approach was applied to the data set, the highest accuracy percentage among the pairs of parameters had to be chosen. Therefore, the highest values of these parameters were separated. After the best accuracy percentage was

determined, the ANNs approach was applied to this pair of parameters as c' versus $V\%$ for discrimination of mono

and dual ovary. Thus, results of both the LDF method and ANNs approach were compared, respectively.

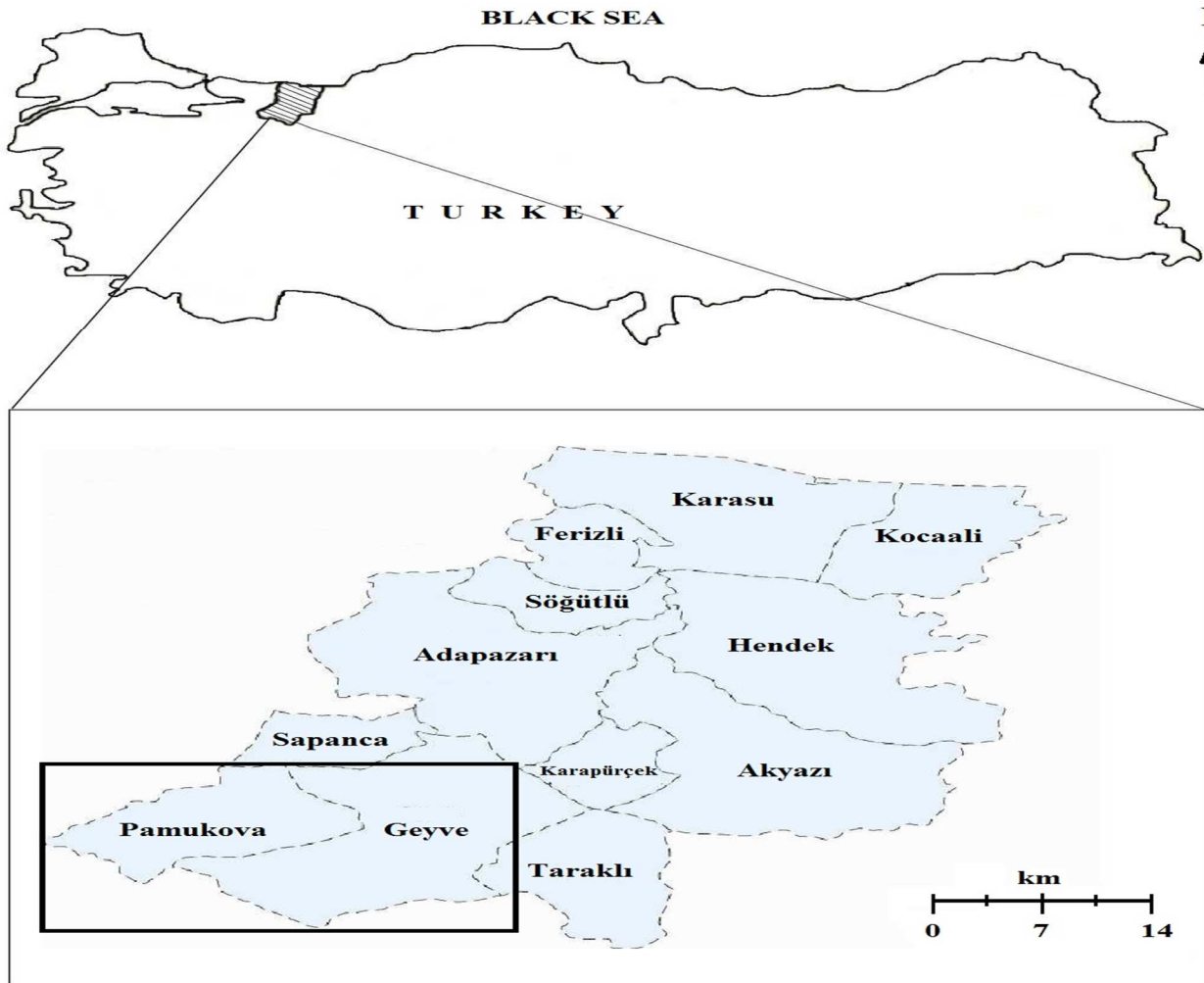


Fig. 1. The location map of the study area as shown inside of the black rectangle (Modified from Isik, 2007).

LDF method: LDF method was used to discriminate different data groups from each other (Fisher, 1936). Generally Linear Discriminate Functions were shown as again simplified in Eq. (1):

$$F_{LDF} = a + b_1X_1 + b_2X_2 + \dots + b_mX_m \quad (1)$$

Here, a is constant number, b_1, \dots, b_m are regression coefficients and X_m is the value of independent variable m .

X_1 : Normalized value of X_m discriminate parameters

The best pair of parameters was decided as c' versus $V\%$ for the data set because it had the highest accuracy percentage for the data number as 230. And then the graphic was drawn and mono and dual ovaries were distinguished using LDF method (Table 1).

For selection of the best pair of parameters, the stage which had the highest accuracy percentage was determined via LDF method using SPSS Analysis Program (SPSS, 2005). Before the best pair of parameters was selected, the values of accuracy percentage as four stages were separated. And then the highest value was chosen for a pair of c' versus $V\%$ parameters.

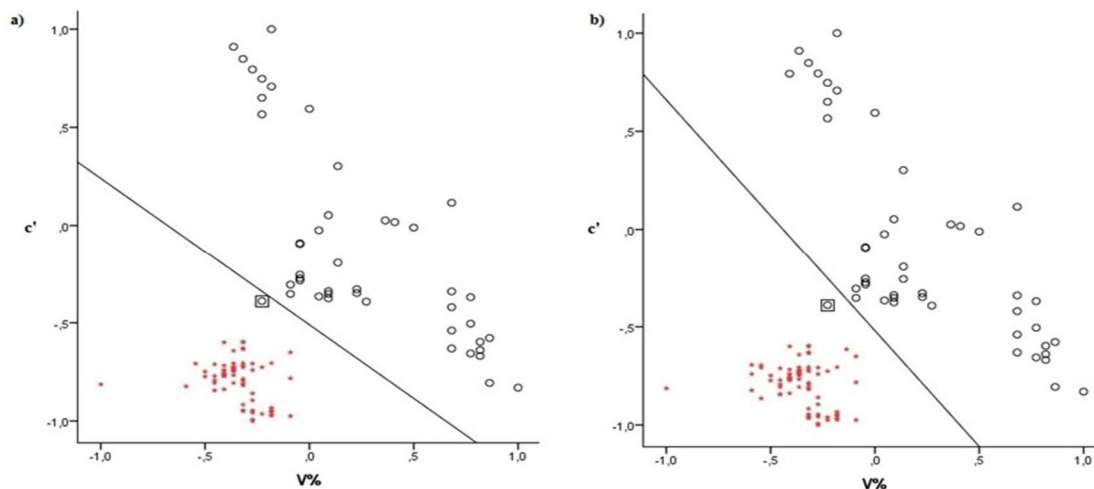
Before LDF method was used, the normalization process was applied to all data sets. The functions were drawn, and the accuracy percentages were calculated by using SPSS Analysis Program (SPSS, 2005). In this study, LDF method was applied to the data set first (Table 2 and Fig. 2).

Table 1. Selection of the best pair of parameters for the July 2016-2017, data set together using LDF method (Grey cells showed the highest value of the accuracy percentage using LDF method).

Pair of parameters	Accuracy (%)	1 st Stage selected	Accuracy (%)	2 nd Stage selected	Accuracy (%)	3 rd Stage selected	Accuracy (%)	4 th Stage selected	Accuracy (%)
	LDF method	pair of parameters	LDF method	pair of parameters	LDF method	pair of parameters	LDF method	pair of parameters	LDF method
L-Stylet	78	(T/VA)-G1	80	G1-Tail	80	Tail-V%	80	c'-V%	80
L-%MB	54	(T/VA)-Tail	80	G1-V%	80	Tail-c'	80		
L-(T/VA)	80	(T/VA)-V%	80	G1-c'	80				
L-G1	80	(T/VA)-c'	80						
L-Tail	80								
L-a	67								
L-V%	78								
L-b	73								
L-b'	70								
L-c	72								
L-c'	80								
L-m	64								
L-o	60								

Table 2. The results of the discriminant analysis using LDF method for pairs of Criteria 1: c' versus V% parameters for the July 2016 data set, Criteria 2: c' versus V% parameters for the July 2017 data set and Criteria 3: c' versus V% parameters for the July 2016-2017 data set. The original grouped cases were correctly classified for two criteria as 99%, 99% and 80%, respectively.

Criterion	Type	Type	Predicted group membership		Total
			Dual Ovary (DO)	Mono Ovary (MO)	
1	Original number	DO	64	0	64
		MO	1	44	45
	%	DO	100.0	0	100
		MO	2.2	97.8	100
2	Original number	DO	74	0	74
		MO	1	46	47
	%	DO	100.0	0	100
		MO	2.1	97.9	100
3	Original number	DO	141	42	183
		MO	3	44	47
	%	DO	77.0	23.0	100
		MO	6.4	93.6	100



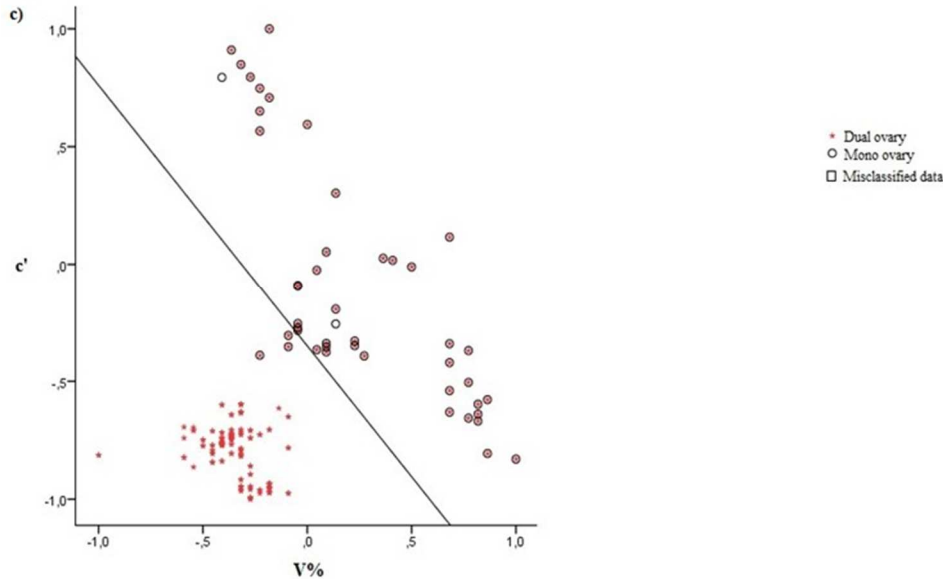


Fig. 2. Plots showed distribution for data set using LDF method a) c' versus $V\%$ parameters for the July 2016 data set b) c' versus $V\%$ parameters for the July 2017 data set and c) c' versus $V\%$ parameters for parameters for the July 2016-2017 data set, respectively. The accuracy percentages were obtained as 99%, 99% and 80% for pairs of c' versus $V\%$ parameters, respectively.

ANNs approach: The ANNs approach was used to compare the results of accuracy percentage of the other method. This method was applied to the data set for the first time. In this study the BPNNs learning algorithm was used. That's because it had some advantages such as reducing errors backwards, namely from output to input

(Cetin *et al.*, 2006). Furthermore, it had a simple neural network topology (Cayakan, 2012). In this study, the BPNNs learning algorithm was used. Generally, members of the network architecture were shown as in Fig. 3 (Rumelhart *et al.*, 1986; Gulbag, 2006).

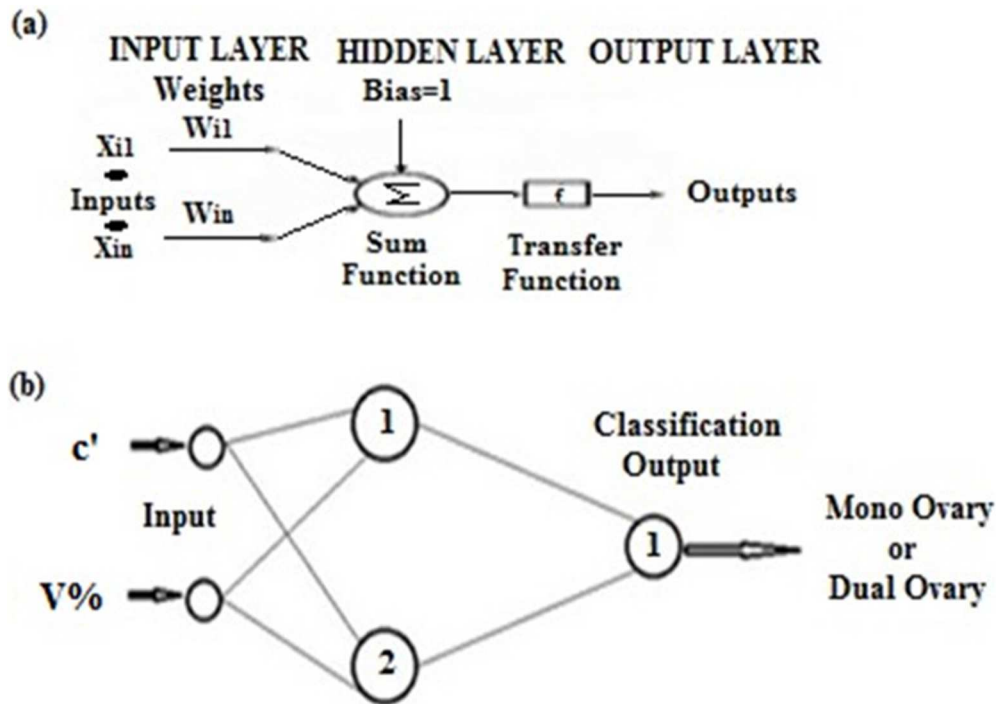


Fig. 3. (a) Members of the network architecture, a neural network structure for types of the ovary (b) c' versus $V\%$ (Modified from Gulbag, 2006).

Pairs of parameters were used for the first time in this study as one of them was the input parameter for testing and the other was the output parameter as the type. These pairs of parameters were determined as c' versus $V\%$ (Fig. 3).

After the learning algorithm was chosen, the data set started to be prepared as “the training data” and “the testing data” for ANNs approach. Different researchers have arranged their data using values of different percentages to separate training data and test data. In other words, there is not a special rule to separate the data (Gulbag, 2006; Yildirim, 2013; Tan, 2021; Tan et al., 2021a; Tan et al., 2021b; Tan et al., 2022). In this study, the data set was arranged by using one station that belonged to the data set randomly. It was decided to use

70% of all data as training data and 30% of all data as testing data.

The July 2016 data set had 109, the July 2017 data set had 121 and the July 2016-2017 data set had 230 numbers for types of the ovary. These data sets were separated into two parts as training data (Number of 76 data for the July 2016 data set; number of 85 data for the July 2017 data set; and number of 161 data for the July 2016- 2017 data set, respectively) and as testing data (Number of 33 data for the July 2016 data set; number of 36 data for the July 2017 data set; and number of 69 data for the July 2016-2017 data set, respectively). That is to say, the number of training data set was of 70% using ANNs approach in this study (Table 3).

Table 3. Number of events in training set, testing set, misclassified testing set and misclassified quarry blast for all data sets by using ANNs approach (For pairs of Criteria 1: c' versus $V\%$ for the July 2016 data set; Criteria 2: c' versus $V\%$ for the July 2017 data set and c) c' versus $V\%$ for the 2016-2017 data set, respectively).

Criterion	The number of all data sets	The number of training sets	The number of testing sets	The number of misclassified testing sets	Accuracy (%) (ANNs approach)
1	109	76	33	3	91
2	121	85	36	5	86
3	230	161	69	6	91

All results were obtained using ANNs approach on MATLAB (MATLAB, 2011). The results of accuracy percentage were obtained using this method and applied to all of data k-fold cross validation technique, too (James et al., 2017). ANNs approach was requested to be validated again. So suitable results that had high values of accuracy percentage values between 50% and 91% were obtained, i.e., the results of ANNs approach were very successful. To obtain the network architecture of the artificial neural network, the selection of the number of neurons (Nn) was an important criterion in the ANNs approach (Kermani et al., 2005; Gulbag, 2006). That's

because it was one of the substantial factors for the discrimination of different data sets (Cetin et al., 2006).

Furthermore, Nn was decided by trial & error method (Yildirim, 2013; Kaftan et al., 2017). And then Nn which had the highest accuracy percentage was taken for the determined ANNs model (Gulbag, 2006). In the literature, researchers have used different intervals using different increments for Nn (Gulbag, 2006; Kuyuk et al., 2009; Yildirim, 2013; Kaftan et al., 2017; Tan, 2021; Tan et al., 2021a; Tan et al., 2021b; Tan et al., 2022). In this study, it was increased by 5 between 1 and 25 and then results were compared with each other for the pair of parameters separately (Table 4).

Table 4. The number of neurons (Nn) according to the Accuracy percentage results according to ANNs approach for pairs of of Criteria 1: c' versus $V\%$ for the 2016 July data set; Criteria 2: c' versus $V\%$ for the 2017 July data set and c) c' versus $V\%$ for the 2016-2017 July data set, respectively.

Criterion	Accuracy (%) for Nn:5	Accuracy (%) for Nn:10	Accuracy (%) for Nn:15	Accuracy (%) for Nn:20	Accuracy (%) for Nn:25
1	82	91	76	73	88
2	86	81	64	50	69
3	81	88	91	70	90

The training was continued until the determination coefficient (R^2) approximated to 1. When the suitable value was obtained, the network model was stopped and started to be tested (Table 5).

For the July 2016 data set, Nn was selected as 10 for the pair of c' versus $V\%$ parameters.

Furthermore, for the July 2017 data set, Nn was selected as 5 for the pair of c' versus $V\%$ and 15 for the

pair of c' versus $V\%$ parameters for the July 2016-2017 data set, respectively. That's because N_n was not high for a pair of a parameter. Namely, the topology of the

network was not complex and was close to 1 as R^2 (Table 6).

Table 5. The variation of R^2 according to N_n that were obtained using ANNs approach (For pairs of Criteria 1: c' versus $V\%$ for the 2016 July data set; Criteria 2: c' versus $V\%$ for the 2017 July data set and c) c' versus $V\%$ for the 2016-2017 July data set, respectively).

Criterion	R^2 (Nn:5)	R^2 (Nn:10)	R^2 (Nn:15)	R^2 (Nn:20)	R^2 (Nn:25)
1	1	1	1	1	1
2	0.96	0.97	1	1	1
3	0.6	0.53	0.29	0.23	0.34

Table 6. The selected N_n according to the Accuracy percentage results for pairs of c' versus $V\%$ parameters (Criteria 1: for the 2016 July data set; Criteria 2: for the 2017 July data set and Criteria 3: for the 2016-2017 July data set, respectively).

Criterion	The selected N_n	Accuracy (%) (ANNs approach)
1	10	91
2	5	86
3	15	91

Furthermore, the Levenberg-Marquardt training algorithm and Hyperbolic Tangent-Sigmoid activation function were used in this study (Kermani *et al.*, 2005; Kuyuk *et al.*, 2009). This algorithm had an important application in MATLAB software (Levenberg, 1944; Marquardt, 1963; Charrier *et al.*, 2007; MATLAB, 2011; James *et al.*, 2017). Selected activation function, denoted

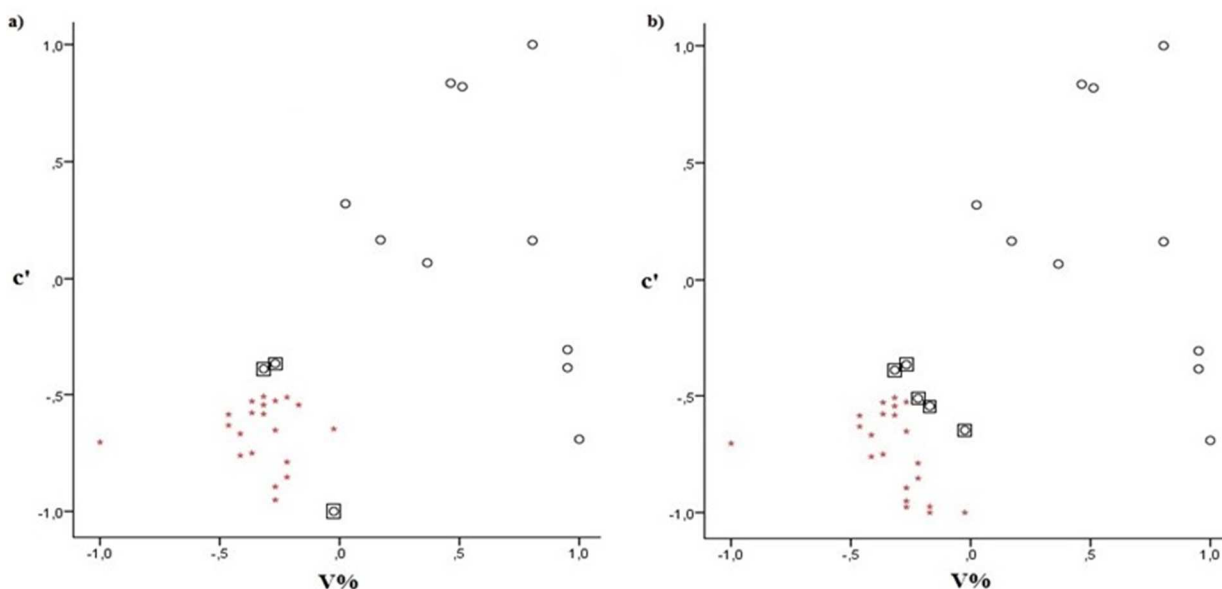
by $\varphi(x)$, defined the output of a neuron in terms of the induced local field. The hyperbolic tangent sigmoid function can be used, defined by using Eq. (2):

$$\varphi(x) = \frac{2}{1+e^{(-2x)}} - 1 \quad (2)$$

Here, $\varphi(x)$: Hyperbolic Tangent Sigmoid activation function (Gradshteyn and Ryzhik, 2007).

Moreover, the normalization process was applied to every data and a significant percentage of the data was selected randomly as the training data. Hence, remaining part was chosen as the testing data (Kermani *et al.*, 2005). After obtained outputs were compared with tested outputs, the accuracy percentage was calculated (Fig. 4).

Moreover, number of testing data, N_n , misclassified data, R^2 , performance, epoch and accuracy percentage values for data set were investigated by using ANNs approach for comparison for pairs of criteria (Table 7).



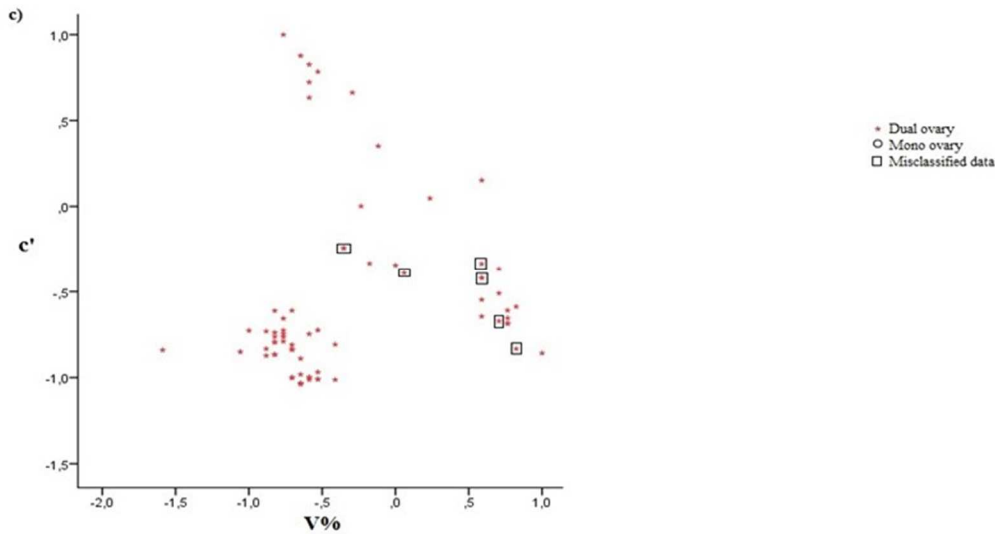


Fig. 4. Plots showed distribution for data set using ANNs approach a) c' versus $V\%$ (Selected $Nn:10$) for the 2016 July data set b) c' versus $V\%$ (Selected $Nn:5$) for the 2017 July data set and c) c' versus $V\%$ (Selected $Nn:15$) for the 2016-2017 July data set, respectively. The accuracy percentages were obtained as 91%, 86% and 91% for pairs of c' versus $V\%$ parameters, respectively.

Table 7. Comparison of number of testings, Nn , misclassified data, R^2 , performance, epoch and accuracy percentage data values for data set using ANNs approach (For pairs of Criteria 1: c' versus $V\%$ for the 2016 July data set; Criteria 2: c' versus $V\%$ for the 2017 July data set and c) c' versus $V\%$ for the 2016-2017 July data set, respectively).

Criterion	Number of testing data	Nn	Number of misclassified data	R^2	Performance	Epoch	Accuracy (%) ANNs approach
1	33	5	6	1	$3.27*(10)^{-008}$	13	82
		10	3	1	$6.4*(10)^{-012}$	14	91
		15	8	1	$8.5*(10)^{-014}$	12	76
		20	9	1	$3*(10)^{-013}$	15	73
		25	4	1	$1.02*(10)^{-012}$	15	88
						(Average) 82	
2	36	5	5	0.96	0.058	7	86
		10	7	0.97	$1.6*(10)^{-014}$	16	81
		15	13	1	$5.8*(10)^{-006}$	12	64
		20	18	1	$8.5*(10)^{-007}$	16	50
		25	11	1	$6*(10)^{-005}$	6	69
						(Average) 70	
3	69	5	13	0.60	0.10	10	81
		10	8	0.53	0.14	1	88
		15	6	0.29	0.12	2	91
		20	21	0.23	0.17	0	70
		25	7	0.34	0.22	2	90
						(Average) 84	

And then the comparison of ANNs approach results with the results of the LDF method is shown in the Table 8.

Table 8. Comparison of the accuracy percentage values for data set according to LDF method and ANNs approach. (For pairs of Criteria 1: c' versus V% for the 2016 July data set; Criteria 2: c' versus V% for the 2017 July data set and c) c' versus V% for the 2016-2017 July data set, respectively.

Criterion	Method and approach	Accuracy (%)
1	LDF	99
	ANNs	91
2	LDF	99
	ANNs	86
3	LDF	80
	ANNs	91

RESULTS

The identification of the type of ovaries was achieved using LDF method and ANNs approach. In this study, firstly the best pair of parameters was determined using LDF method according to the highest accuracy percentage value. For selection of the best pair of parameters, the stage which had the highest accuracy percentage was considered by using LDF method. Before the best pair of parameters were selected, the values of accuracy percentage were separated as four stages. Then the highest value was chosen as 80% for a pair of c' versus V% parameters. Later, three datasets were prepared as the July 2016, the July 2017 and the July 2016-2017 (Table 1). Sixty-four (59%) out of total 109 female nematodes were described as dual ovaries and 45 (41%) of them were described as mono ovaries for the 2016 July data set. Seventy-four (61%) out of total 121 female nematodes were described as dual ovaries and 47 (39%) of them were described as mono ovaries for the July 2017 data set. One hundred-eighty-three (80%) out of total 230 female nematodes were described as dual ovaries and 47 (20%) of them were described as mono ovaries for the July 2016-2017 data set, too. Numbers of the types of female ovaries were compared with each other in the Sakarya district. The results of the classification method between the types of the ovaries using LDF method for pairs of criteria: 1- c' versus V% for the July 2016 data set; 2- c' versus V% for the July 2017 data set and 3- c' versus V% for the July 2016-2017 data set were shown in Table 2, respectively.

The c' versus (V%) of the nematode had a higher classification accuracy percentage for dataset than others as 99% for LDF method and 91% for ANNs approach for the July 2016 data set. The LDF method was as successful as the ANNs approach. A new perspective, thus, was introduced to the principles of the taxonomical nematology. This study was achieved for Sakarya by

using these pairs of parameters with these new method and approach together at the nematological studies in both Turkey and the world for the first time.

In the first criterion, 64 dual ovaries were classified correctly, and 1 dual ovary was misclassified as mono ovary. 44 mono ovaries were classified correctly, and 1 mono ovary was misclassified as dual ovary. So, the accuracy percentage was obtained as 99% for the July 2016 data set by using LDF method. In the second criterion, 74 dual ovaries were classified correctly, and 1 dual ovary was misclassified as mono ovary. 46 mono ovaries were classified correctly, and 1 mono ovary was misclassified as dual ovary. Thus, the accuracy percentage was obtained as 99% for the July 2017 data set by using LDF method. In the third criterion, 141 dual ovaries were classified correctly, and 3 dual ovaries were misclassified as mono ovary. 44 mono ovaries were classified correctly, and 3 mono ovaries were misclassified as dual ovary. Therefore, the accuracy percentage was obtained as 80% for the July 2016-2017 data set by using LDF method (Table 2).

After mono and dual ovary were distinguished by using LDF method, ANNs approach was applied for the same pair of parameters. Firstly, it had to be decided for Nn, then we created test and training data set for the two criteria in Table 3, respectively. The values of the accuracy percentage for ANNs approach were also given in Table 4. The accuracy percentage values varied from 50% to 91%. The values of the number of neurons which were raised by 5 between 5 and 25 were given in Table 4. Next, Nn versus the determination coefficient (R^2) per dataset for c' versus V% (Table 5). R^2 values varied from 0.23 to 1 in that table. This situation indicated that BPNNs learning algorithm was successful for those parameters on that structure of the network topology. The comparison of R^2 values that were obtained using ANNs approach for pairs of parameters in this study area and the comparison of R^2 versus the number of neurons were alone not enough to decide. Table 5 shows that this relationship was only a stopping criterion to stop the training stage of the architecture network.

Nn were decided as 10, 5 and 15 at the network architectures for pair of criteria: 1- c' versus V% for the July 2016 data set; 2- c' versus V% for the July 2017 data set and 3- c' versus V% for the July 2016-2017 data set, respectively. Because the average accuracy percentage were the highest as 91%, 86% and 91% for the July 2016 data set, for the July 2017 data set and for the July 2016-2017 data set, respectively (Table 6).

When the table of the number of testing data, Nn, misclassified data, R^2 , performance, epoch and accuracy percentage values for three datasets were obtained using ANNs approach, they were compared with each other for pairs of criteria (Table 7). It can be concluded from this table that if Nn and number of testing data increased, number of misclassified data and

performance also increased but R^2 and epoch decreased. Next, average of the accuracy percentage values changed irregularly in that situation for N_n that increased by 5 between 1 and 25.

Additionally, c' versus $V\%$ values of the accuracy percentage for LDF method and ANNs approach were shown for the July 2016 data set, for the July 2017 data set and for the July 2016-2017 data set in Table 8. According to c' versus $V\%$, the accuracy percentage values were obtained using LDF method and ANNs approach as 99% and 91% for the July 2016 data set, respectively. And then according to c' versus $V\%$, the accuracy percentage values were obtained as 99% and 86% for the July 2017 data set. Also, c' versus $V\%$ the accuracy percentage values were obtained using LDF method and ANNs approach as 80% and 91% for the July 2016 and the July 2017 data sets, respectively. Values of pair of the c' versus $V\%$ for ANNs approach were plotted in Figure 4 for the July 2016 data set; the July 2017 data set and the July 2016-2017 data set, respectively.

DISCUSSION

LDF method was one of the most popular and successful techniques for the classification of different groups among the natural sciences in the world. Keles (2019) used ANNs approach and Discriminate Analysis classification methods, and their success rates were 89.1% and 92.7% on X axis, 92.7% and 92.7% on Y axis, 86.8% and 88.7% on Z axis, respectively for classification of hazelnut varieties. Tan *et al.* (2022) obtained success rates as among 91-97% for LDF method and 94-100% for ANNs approach for discrimination of the ovary types of some nematodes, too.

In this study, the values of N_n which were increased by 5 between 5 and 25 were shown in Table 4-5-6. N_n versus R^2 values per data set were shown in Table 5 for pair of criteria: 1- c' versus $V\%$ for the July 2016 data set; 2- c' versus $V\%$ for the July 2017 data set, and 3- c' versus $V\%$ for the July 2016-2017 data set, respectively. R^2 values varied from 0.23 to 1 in these tables, too. It means that BPNNs learning algorithm was successful for these pairs of parameters on that structure of the topology in the area considered in this study. Tan *et al.* (2022) found the R^2 values between 0.85 and 1 in their study, which exhibits that the BPNN learning algorithm was successful for this network topology, too. For that reason, learning occurred successfully, and everything went on well through the process, too.

When it was compared that the accuracy percentage values for three criteria (c' versus $V\%$ for the July 2016 data set, the July 2017 data set and the July 2016-2017 data sets, respectively), the pair of c' versus $V\%$ had higher classification accuracy percentage values for both of the July 2016 datasets (99% for LDF method

and 91% for ANNs approach) than the other data sets in Table 8.

Additionally, misclassified data was near the limit of the discrimination area between two different groups called as “the mono ovary” and “the dual ovary” (As shown in Figure 2 and Figure 4) for both LDF method and ANNs approach. This is a new finding in this study according to these classification techniques, too.

Moreover, the BPNN algorithm was one of the most popular and successful techniques for the classification of different groups at multidisciplinary sciences in the world, too.) Some researchers investigated some datasets using BPNN Learning algorithm successfully between absolutely 80-100%, too (Yildirim *et al.*, 2011; Tan, 2021; Tan *et al.*, 2021a; Tan *et al.*, 2021b; Tan *et al.*, 2022). Akyuz (2019) used the BPNN learning algorithm and obtained the R^2 values of the ANN models as above 99% that is accepted to be successful.

Consequently, in this study the accuracy percentage values of the LDF method were as successful as results using ANNs approach. But ANNs approach was more successful than LDF method for classification of mono and dual ovaries of plant parasitic nematodes. Hence, it was concluded that the mono varies were distinguished very well in this study and it may improve the nematological cultivation studies.

Moreover, the results were obtained using LDF method and ANNs approach. The highest classification accuracy percentage values were achieved for the July 2016 dataset (99% for LDF method and 91% for ANNs approach). Hence, this training algorithm proved to be successful.

Additionally, Tan *et al.* (2022) achieved successful accuracy percentage values (as 91% and 97% for LDF method and 97% and 100% for ANNs approach) by using LDF method and ANNs approach together for the first time in the world in nematology study area, too.

Conclusion: In this study, a pair of parameters, which were some morphometric measurement data, had been obtained using the human logic. It was applied for the classification of the plant parasitic nematodes. The LDF method which was a discrimination method had been applied for determination of the correlation between that pair of parameters. And then it was proved that ANNs approach could approximate the analyzing capability of the human logic successfully. So, a new perspective has already been introduced to the principles of the taxonomical nematology. Physical properties gave high accuracy degrees in our study, which shows that a database created in this way can also be used in classical taxonomy research in the future. The results of our study demonstrate that the identification method tested is cheap, quick and dependable with high accuracy and for this reason it can be considered as an alternative to

innovative applications in identifying ovary types of plant parasitic nematodes. A classification method for the plant parasitic nematodes on the quince cultivated areas based on ANNs approach was proposed in this study. The parameters of some morphological measurements of female members of plant parasitic nematodes obtained in the research were used. The results have showed that ANNs approach and LDF method achieved higher classification accuracy percentage values for the July 2016 dataset (99% for LDF method and 91% for ANNs approach) than the other datasets for the pair of c' versus $V\%$. The contributions of the study can be expressed as follows: Back-propagation algorithm was introduced, and the training data and testing data of ANNs were used. Identification and classification techniques of ANNs approach and LDF method that can easily be adapted into practice were chosen to classify the parameters of some morphological measurements of female members of the plant parasitic nematodes. The preferred identification and classification properties may be used in the design of many different network architectures for the ANNs from a simple state to a complex regulation. It is believed that this method can be an alternative to eliminating the complexity of classification problems in the plant parasitic nematodes and will also contribute to the time management. Moreover, investigations should give priority to using more parameters of some morphological measurements of the plant parasitic nematodes species and their different data, too.

Authors' contribution: Ayşe Nur TAN, Ph. D. obtained the taxonomical results from the study fields, collecting soil samples, isolating, fixing, preparing, counting, calculating and measuring some physical parameters of the nematodes and then diagnosing in the laboratory and finally examining some scientific journals and writing some captions of the manuscript on the computer. Next, Aylin TAN, Ph. D. arranged the data sets normalizing all data set for the statistical analysis and the artificial neural networks and then examining some scientific journals and writing some captions of the manuscript on the computer. Both authors controlled the foreign language of this manuscript.

REFERENCES

- Ahmed, M., M. Sarp, T. Prior, G. Karssen and M. Back (2015). Nematode taxonomy: from morphology to metabarcoding. *SOIL Discussions, An Interactive Open-Access J.* 2: 1175-1220. DOI: <http://www.doi.org/10.5194/soild-2-1175-2015>
- Akal, M., B. Gokce and S. Celik (2020). Geyve district quince producers survey. *Sakarya University J. Business Institute.2* (2): 41-49 (in Turkish). DOI: <http://www.doi.org/10.47542/sauied.769688>
- Akintayo, A., G. L. Tylka, A. K. Singh, B. Ganapathysubramanian, A. Singh and S. Sarkar (2018). A deep learning framework to discern and count microscopic nematode eggs. *Nature Scientific Reports.* 8: 9145. DOI: <http://www.doi.org/10.1038/s41598-018-27272-w>
- Akyuz, I. (2019). Future projection and the sales of industrial wood in Turkey: Artificial neural networks. *Turkish J. Agriculture and Forestry.* 43: 368-377. DOI: <http://www.doi.org/10.3906/tar-1901-20>
- Aragon, D., R. Landa, L. Saire, G. Kemper and C.D. Carpio (2019). A neural-network based algorithm oriented to identifying the damage degree caused by the *Meloidogyne incognita* nematode in digital Images of Vegetable Roots. *CONIITI Int.* 123-127. DOI: <http://www.doi.org/10.1109/CONIITI48476.2019.8960622>
- Cayakan, C. (2012). Partial saturation estimation to be applied in sands for liquefaction improvement by the method of artificial neural networks. M.Sc. thesis (unpublished). Dept. of Civil Engineering, Istanbul Technical University, Istanbul (in Turkish). Available at: <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Cetin, M., A. Ugur and S. Bayzan (2006). Heuristic approach of the Backpropagation algorithm in high-feed artificial neural networks. IV. Wisdom and Academic Informatics Symposium Int. 190-197 (in Turkish). Available at: <https://www.researchgate.net/profile/Sahin-Bayzan>
- Charrier, C., G. Lebru and O. Lezoray (2007). Selection of features by a machine learning expert to design a color image quality metrics. *Video Processing and Quality Metrics for Consumer Electronics (VPQM) Int.* 113-119. Available at: https://lezoray.users.greyc.fr/Publis/charrier_VPQM2007.pdf
- Dababat, A. A., H. Miminjanov and R. W. Smiley (2015). Nematodes of small grain cereals: Current status and research. *FAO Publishers; Turkey.* Available at: https://www.researchgate.net/publication/283714250_NEMATODES_OF_SMALL_GRAIN_CEREALS_CURRENT_STATUS_AND_RESEARCH?_sg%5B0%5D=7cGS9fiQs33FepNLg1siT3rT47MdVyLBsOogRS84pgjuH7jwQa_Cal7RVgJ6W7TcGc3iLsjr8cOj87XMhJxmin1diHJ3vwsU5utmETCLIRDsQul5xucjA7s.J7aQ_tUdQL6Muls8JnsDZ5Gqn2k0Fol0G9gi3JGcnkIzIblIEfajJT5npi0RB_8m86kUWjjHZms-GO2_T4miIA

- Daramola, F., J. Popoola, A. O. Eni and O. Sulaiman (2015). Characterization of Root-knot Nematodes (*Meloidogyne* spp.) associated with *Abelmoschus esculentus*, *Celosia argentea* and *Corchorus olitorius*. Asian J. Biology Sciences. 8: 42–50. DOI: <http://www.doi.org/10.3923/ajbs.2015.42.50>
- De Man, J. G. (1880). The native nematodes living freely in the pure earth and in fresh water. Preliminary report and descriptive-systematic report. Tijdschrift der Nederlandsche Dierkundige Vereeniging.5: 1-104 (in Dutch). Available at: <https://www.biodiversitylibrary.org/bibliography/8982>
- Desaeger, J., M. R. Rao and J. Bridge (2004). Nematodes and other soilborne pathogens in agroforestry. In: van Noordwijk M, Cadisch G and Ong. CK (eds.) Below-Ground Interactions in Tropical Agroecosystems: Concepts and Models with Multiple Plant Components. CABI Publishing, UK. DOI: <http://www.doi.org/10.1079/9780851996738.0263>
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Human Genetics. 7(2): 179-188. DOI: <http://www.doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Gaugler, R. and A. L. Bilgrami (2004). Nematode Behaviour. CABI Press; New York (USA). 432 p. Available: <https://www.cabidigitallibrary.org/doi/book/10.1079/9780851998183.0000#>
- Gradshteyn, I. S. and I. M. Ryzhik (2007). Table of integrals, series, and products. Academic Press; USA, 1161 p. Available: http://fisica.ciens.ucv.ve/~svincenz/TISPISGIM_R.pdf
- Gulbag, A. (2006). Quantitative determination of volatile organic compounds by artificial neural network and fuzzy logic-based algorithms. Ph. D. thesis (unpublished). Dept. of Computer Engineering, Sakarya University, Sakarya (in Turkish). Available at: <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Isik, S. (2007). Agricultural geography of Sakarya. M.Sc. thesis (unpublished). Dept. of Geography, Sakarya University, Sakarya (in Turkish). Available at: <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- James, G., D. Witten, T. Hastie and R. Tibshirani (2017). An introduction to statistical learning with application. Springer Publication; England, 440 p. Available at: https://hastie.su.domains/ISLR2/ISLRv2_website_e.pdf
- Kaftan, I., M. Salk and Y. Senol (2017). Processing of earthquake catalog data of Western Turkey with artificial neural networks and adaptive neuro-fuzzy inference system. Arabian Geophysical Geosciences. 10: 243. DOI: [10.1007/s12517-017-3021-1](https://doi.org/10.1007/s12517-017-3021-1)
- Keles, O. (2019). Classification of some varieties of nuts using artificial neural networks and discriminant analysis. M.Sc. thesis (unpublished). Dept. of Agricultural Engineering, Ondokuz Mayıs University, Samsun (in Turkish). Available at: <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Kermani, B. G., S. S. Schiffman and H. G. Nagle (2005). Performance of the Levenberg–Marquardt neural network training method in electronic nose applications. Sensors and Actuators B: Chemical. 110 (1): 13-22. DOI: <https://doi.org/10.1016/j.snb.2005.01.008>
- Kurtulmus, F., A. Polat and N. Izli (2020). Modeling of drying speed and humidity parameters in drying apricots by different drying methods using Artificial Neural Networks. COMU J. Agric. Fac. 8 (2): 261-269 (in Turkish). DOI: [10.33202/comuagri.733166](https://doi.org/10.33202/comuagri.733166)
- Kuyuk, H. S., E. Yildirim, G. Horasan and E. Dogan (2009). Investigation of earthquake and quarry blasting data by reaction surface, multivariate regression and learning vector quantization methods. Sakarya Earthquake Symposium Int.1-10. (In Turkish). Available at: https://www.researchgate.net/publication/295605801_Investigation_on_discrimination_of_quarry_blast_and_earthquake_data_by_response_surface_multi-regression_and_learning_vector_quantization_methods
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. Quartet Applied Mathematics. 2: 164-168. DOI: <https://doi.org/10.1090/qam/10666>
- Marquardt, D. W. (1963). An algorithm for Least-Squares estimation of nonlinear parameters. J. The Society for Industrial and Applied Mathematics. 11(2): 431-441. Available at: <https://www.jstor.org/stable/2098941>
- MATLAB, (2011). Release, The neural network toolbox The Math Works, Increments, Natick Massachusetts, United States. Available at: <https://www.mathworks.com/products/matlab.html>
- Mitiku, M. (2018). Plant-Parasitic Nematodes and their Management: A Review. Agri Res & Tech:

- Open Access J. 16 (2): ARTOAJ.MS.ID.555980. DOI:10.19080/ARTOAJ.2018.16.55580
- Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). Learning Internal Representations by Error Propagation. In: Rumelhart DE and McClelland JL (Eds.). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press., Massachusetts (USA). Available at: <https://ieeexplore.ieee.org/document/6302929>
- Siddiqi, M. R. (2000). Tylenchida parasites of plants and insects. CAB International; Wallingford (UK). Available at: <https://www.cabidigitallibrary.org/doi/book/10.1079/9780851992020.0000>
- SPSS, (2005). V.17.0, SPSS for Windows. SPSS Increments (Statistical Package for the Social Sciences). Available at: <https://www.ibm.com/products/spss-statistics>
- Sundararaju, R., R. L. Devi and M. Manikemalai (2002). Analysis of Best Treatment and Variety Based on Nematode Population on Banana using Artificial Neural Networks. Indian J. Nematology. 32(1): 78-101. Available at: <https://www.indianjournals.com/ijor.aspx?target=ijor:ijn&volume=32&issue=1&article=019>
- Tan, A. (2021). Differentiation of earthquakes and explosions in different regions of Turkey. Ph. D. thesis (unpublished). Dept. of Geophysical Engineering, Sakarya University, Sakarya (in Turkish). Available at: <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Tan, A., G. Horasan, D. Kalafat and A. Gulbag (2021a). Discrimination of earthquakes and quarries in Kula District Manisa, Turkey) and its vicinity by using linear discriminate function method and artificial neural networks. Bulletin of the Mineral Research and Exploration. 164: 75-92. DOI: <https://doi.org/10.19111/bulletinofmre.757701>
- Tan, A., G. Horasan, D. Kalafat and A. Gulbag (2021b). Discrimination of earthquakes and quarries in the Edirne district (Turkey) and its vicinity by using a linear discriminate function method and artificial neural networks. Acta Geophysica. 69: 17-27. DOI: 10.1007/s11600-020-00519-9
- Tan, A. N., A. Tan and H. Susurluk, (2022). First application of two distinguishment techniques: Using Linear Discriminate Function method and Artificial Neural Networks approach according to the ovary types for some plant parasitic nematodes. Harran J. Agri. and Food Sci., 46 (2): 1-14. DOI: 10.29050/harranziraat.1025087
- Uhlemann, J., O. Cawley and T. Kakouli-Duarte (2020). Nematode identification using Artificial Neural Networks. Conference on Deep Learning Theory and Applications Int. pp. 1-22. DOI: DOI: 10.5220/0009776600130022
- Yildirim, E., A. Gulbag, G. Horasan and E. Dogan (2011). Discrimination of quarry blasts and earthquakes in the vicinity of Istanbul using soft computing techniques. Computers and Geosciences. 37: 1209-1217. DOI: <https://doi.org/10.1016/j.cageo.2010.09.005>
- Yildirim, E. (2013). Classification of seismic waves by determining the ground properties from the damping character. Ph. D. thesis (unpublished). Dept. of Geophysical Engineering, Sakarya University, Sakarya (in Turkish). Available at: <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- URL: <http://faostat3.fao.org/> (accessed: January 25, 2022).
- URL: <http://www.tuik.gov.tr/> (accessed: January 20, 2022).