Şenol, H., Çolak, H., Çolak, E. (2024). Analysis of the Total Number of Licensed Athletes Using Artificial Neural Networks for the Future Years in Türkiye. *The Black Sea Journal of Sciences*, 14(4), 2153-2171.

The Black Sea Journal of Sciences, 14(4), 2153-2171, 2024. DOI: 10.31466/kfbd.1531019

KFBD

Karadeniz Fen Bilimleri Dergisi

The Black Sea Journal of Sciences

ISSN (Online): 2564-7377 https://dergipark.org.tr/tr/pub/kfbd

Araştırma Makalesi / Research Article

Analysis of the Total Number of Licensed Athletes Using Artificial Neural Networks for the Future Years in Türkiye

Halil ŞENOL[1], Halil ÇOLAK[2], Emre ÇOLAK[3*]

Abstract

The term sport refers to a collection of competitive or leisure physical activities conducted under certain rules and regulations to enhance individuals' physical capabilities, skills, and endurance. Official figures indicate that there are roughly 6.25 million licensed athletes in Türkiye as of 2022. The projection of this figure in the forthcoming years is crucial for facilitating more efficient sports policy planning. The entire number of athletes in Türkiye till 2040 was projected using artificial neural networks (ANN). The utilization of artificial neural networks to predict the number of athletes facilitates the production of projections for subsequent years. These estimations furnish essential data for the expansion of sports and the growth potential of the sports business. The study utilized the Levenberg-Marquardt and Bayesian Regularization techniques of ANN. By the year 2040, it is projected that Türkiye would have a minimum of 7.33 million athletes. Future research should quantify and analyze the number of athletes across various disciplines utilizing ANN algorithms.

Keywords: Machine learning, Estimate methods, Yoon algorithm, Sports, Number of athletes.

Türkiye'de Gelecek Yıllar için Toplam Lisanslı Sporcu Sayısının Yapay Sinir Ağları Kullanılarak Analizi

Öz

Spor terimi, bireylerin fiziksel yeteneklerini, becerilerini ve dayanıklılıklarını geliştirmek için belirli kurallar ve düzenlemeler çerçevesinde gerçekleştirilen rekabetçi veya eğlence amaçlı fiziksel aktiviteler bütünü olarak tanımlanmaktadır. Resmi kayıtlara göre Türkiye'de yaklaşık 6,25 milyon lisanslı sporcu bulunmaktadır (2022 yılı için). Önümüzdeki yıllarda bu sayının tahmin edilmesi, spor politikalarının daha etkili planlanmasına olanak sağlaması açısından önemlidir. Bu bağlamda, 2040 yılına kadar Türkiye'deki toplam sporcu sayısı yapay sinir ağları (YSA) ile tahmin edilmiştir. Sporcu sayısını tahmin etmek için YSA'ların uygulanması, gelecek yıllar için tahminlerin üretilmesini sağlamaktadır. Bu tahminler, sporun yaygınlaşması ve spor ekonomisinin büyüme potansiyeli için kritik bilgiler sağlamaktadır. Çalışmada YSA'nın Levenberg-Marquardt ve Bayes Regülarizasyon algoritmaları kullanılmıştır. 2040 yılına kadar Türkiye'de en az 7,33 milyon sporcu olacağı tahmin edilmektedir. Gelecek çalışmalarda farklı branşlardaki sporcu sayılarının YSA algoritmaları kullanılarak hesaplanıp tartışılması önerilmektedir.

Anahtar Kelimeler: Makine öğrenmesi, Tahmin yöntemleri, Yoon algoritması, Sporlar, Sporcu sayısı.

[1]Giresun University, Faculty of Engineering, Department of Energy Systems Engineering, Giresun, Türkiye, halil.senol@giresun.edu.tr
[2]Giresun University, Faculty of Sports Sciences, Department of Physical Education and Sports, Giresun, Türkiye, halil.colak@giresun.edu.tr
[3]Giresun University, Institute of Science, Department of Energy Systems Engineering, Giresun, Türkiye, emre.colak@giresun.edu.tr

[*]Sorumlu Yazar/Corresponding Author
Geliş/Received: 09.08.2024 **Kabul/Accepted:** 16.10.2024 **Yayın/Published:** 15.12.2024

1. Introduction

The term sport broadly refers to many activities necessitating diverse talents, abilities, and, occasionally, an element of chance. Individuals, pairs, or teams may conduct these exercises (Ryan & Doody, 2024). Sport serves as a crucial avenue for realizing objectives, such as acquiring prestige, popularity, financial rewards, and representing one's nation (Başkan, Özgül, Kolukısa, Çolak, & Başkan, 2020). Globally, engagement in sports and physical activities is crucial for the comprehensive physical, emotional, social, and mental health of the populace (Hulteen et al., 2017). An increasing amount of data indicates that comprehending sports might enhance social adaptation. Moreover, data indicates that engagement in sports correlates with enhanced moral development, which subsequently correlates with a decrease in criminal behavior and violence (Seefeldt & Ewing, 1997).

Sports analytics is a contemporary discipline that utilizes various analytical methods to investigate the extensive realm of sports (Exel & Dabnichki, 2024). The integration of many technologies and advances in sports has propelled the progress of both the disciplines and the players. This has coincided with the development of new inventions and the advancement of other technologies (Qi, Sajadi, Baghaei, Rezaei, & Li, 2024). An examination of the literature indicates that sports technologies are gaining prominence across several contexts. This includes the use of standardized sports equipment, advanced training tools, evaluating sports performance, building large sports arenas, developing performance analysis software, expanding the sports supplement industry, conducting anti-doping research, and implementing mathematical modeling and computer simulation (Gomis-Gomis, Pena-Pérez, & Pérez-Turpin, 2023; Kelly, Derrington, & Star, 2022). Every day, artificial intelligence, a subset of technology, advances and integrates into the world of sports (Çolak & Şenol, 2023). Thus, the application of artificial intelligence in sports and athlete development might enhance efficiency in processes such as athlete planning and preparation.

A key idea for comprehending the post-digital era is datafication. The word dataization refers to the ultimate phase of deep mediation (Mertala & Palsa, 2024). The emergence of data and big data eras has resulted in increased interest in machine learning (Zhou, Pan, Wang, & Vasilakos, 2017). The use of machine learning signifies substantial progress in artificial intelligence, impacting several domains within the sciences and engineering. The main goal of the system is to find and understand the complicated relationships among the various characteristics that constitute a complex system. This obviates the need for direct human interaction (Rezaei & Javadi, 2024). A primary advantage of machine learning is its ability to predict future occurrences and situations across several fields (Şenol, Çolak, Elibol, Hassaan, & El Nemr, 2024). Artificial neural networks (ANNs) are a subset of machine learning that emulate the functioning of the human brain, therefore embodying a type of artificial intelligence (Çolak & Çolak, 2024; Şenol, Çolak, & Oda, 2024). ANNs operate on principles analogous to those of the biological nervous system, which analyzes information and data to acquire and produce new knowledge (J. Zhu, Hu, Khezri, & Ghazali, 2024).

Literature contains works of machine learning techniques in sports and athletics. Bunker and Thabtah (2019) selected artificial neural networks as a machine learning technique for predicting basketball game outcomes. The results indicated that the artificial neural network model employed was appropriate for historical data but inadequate for future matches. Wilkens (2021) employed machine learning techniques like logistic regression, artificial neural networks, random forests, gradient boosting machines, and support vector machines to estimate tennis match outcomes. Nguyen et al. (2022) utilized machine learning on National Basketball Association players. The techniques employed were linear regression, gradient boosting machine, random forest, and ANN. Me and Unold (2011)) used machine learning, integrating fuzzy modeling with the immunity algorithm, to simulate tennis training. Zhu and Sun (2020) employed artificial neural networks, linear regression, and support vector machine techniques to assess athlete performance. Zhang and Li (2022) conducted a study on perovskite material analysis and data monitoring in training apparatus. They favored artificial neural networks, support vector machines, and genetic algorithms for this purpose. Ishwarya and Nithya (2021) employed methodologies including Random Forest, Support Vector Machine, and Multilayer Perceptron to evaluate athlete performance in cricket. Jauhiainen et al. (2021) performed estimation research on injury prediction for young athletes utilizing logistic regression and random forest methodologies. Kasera and Johari (2021) employed random forest and logistic regression to predict the victor of the Cricket World Cup and compared the effectiveness of the models. The existing literature estimates have incorporated investigations from several global regions. An examination of the existing literature reveals that analyses concerning the overall number of athletes in Türkiye are notably few, indicating a necessity for more research.

This study intends to project the number of athletes in Türkiye for the forthcoming years (up to 2040) utilizing ANN. The authors found a dearth of research in the literature regarding the estimation of the athletic population in Türkiye. Dalkılıç et al. (2017) utilized ANN to estimate the population of impaired athletes in Türkiye. Dalkılıç et al. (2017) employed data from 2007 to 2016 to estimate the population of athletes in underwater sports, water polo, and swimming in Türkiye utilizing ANN. Dalkılıç et al. (2017) employed ANN to project the number of wrestlers in Türkiye for the years 2017 and 2018, utilizing the actual athlete data from 2007 to 2016. Atasoy et al. (2017) utilized ANN to estimate the number of licensed athletes in 2017 and 2018, based on the data of combat sports athletes from 2007 to 2016 in Türkiye the completed research to date estimates the number of athletes in Türkiye. Only one research study in the literature has projected the number of athletes in Türkiye for the forthcoming years. Çolak and Şenol (2023) projected the number of athletes in Türkiye by 2030 utilizing the BR technique of ANN. To the authors' best knowledge, there exists just one research that projected the number of athletes in Türkiye for the next years, particularly in the domain of machine learning (Çolak & Şenol, 2023). Türkiye is formulating strategic plans and investments to attain success in sports at both national and international levels. Precise assessment of the athlete population will enhance the efficacy of these tactics and guarantee optimal resource allocation. Moreover, the application of modern technological methodologies, such as ANN, would provide a novel approach to monitoring athlete growth and performance, delivering superior accuracy and flexibility compared to conventional estimating techniques.

This study intends to project the number of athletes in Türkiye for the forthcoming years utilizing ANN. Six models will be developed utilizing data on athlete numbers, demographic information, economic variables, and historical sports club data, with an assessment of their correctness and dependability. The entire count of licensed athletes from 2019 to 2040 was examined using an ANN.

2. Theoretical Method

2.1. Artificial neural networks

ANNs have been crucial in the advancement of machine learning in recent years (Şenol, Çolak, Elibol, et al., 2024). ANNs are simplified versions of biological neural networks that are meant to copy brain functions like recognizing patterns, figuring out what things mean, sorting data, and predicting what will happen in the future (Şenol, Çolak, & Oda, 2024). ANNs emulate biological brain networks to represent both linear and nonlinear interactions (Şenol, Dereli, & Özbilgin, 2021). The fundamental components of the ANN are neurons interconnected in parallel that constitute each layer of the ANN (Şenol, 2021) An ANN has three primary layers: input, output, and hidden (Çolak & Çolak, 2024). The intermediary layer between the input and output layers comprises entities known as neurons (Şenol, Çolak, Elibol, et al., 2024). Weighted values interconnect neurons, which process and convey information. In ANN, neurons across layers emulate the function of synapses in the brain, facilitating the flow of information between neurons (Şenol, 2021). ANN methodologies adjust the models by varying the number of hidden layers according to the data architecture, thereby achieving a significant degree of adaptability to the data (Bas, Egrioglu, & Cansu, 2024). ANNs utilize diverse activation functions (f(x)). The study formulated activation functions as either tangent sigmoid (Eq. (1)) or linear (Eq. (2)) for the output layer and hidden neuron layer. The tangent sigmoid approximates nonlinear behavior and offers superior overall outcomes (Çolak & Çolak, 2024; Şenol, Çolak, & Oda, 2024). Figure 1 illustrates the process of the artificial neural network (ANN).

$$f(x) = x \tag{1}$$

$$f(x) = \frac{2}{1 + \exp^{-2e}} - 1 \tag{2}$$

The expression x in Eq. (1) represents the parameter data in the input layer.

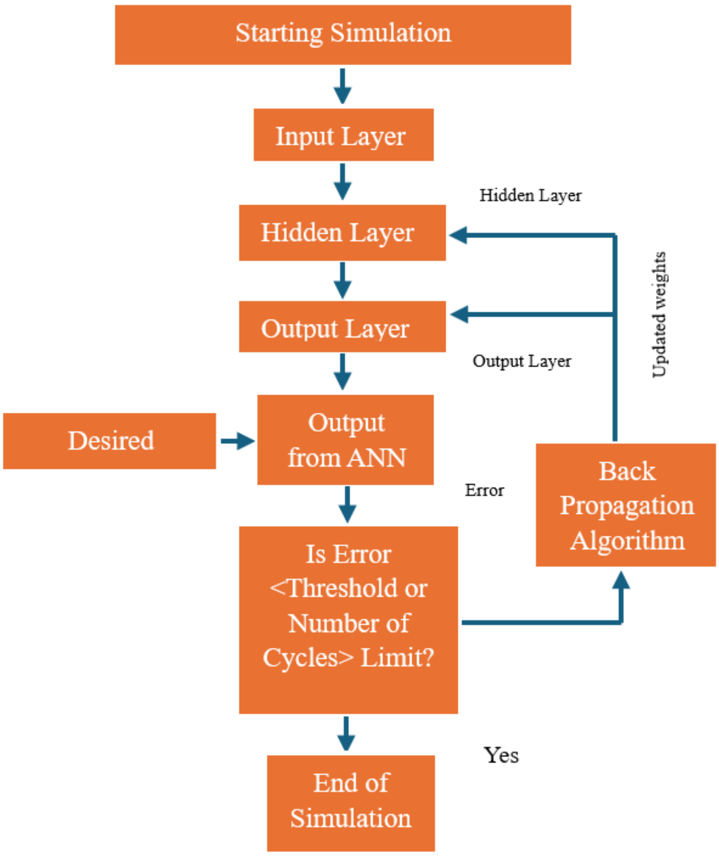

Fig. 1. Artificial Neural Network workflow diagram.

Nonlinear least squares problems, particularly in least squares curve fitting, employ the Levenberg-Marquardt (LM) method, which iteratively identifies the local minimum of a multivariate function. It is regarded as an intermediary technique between the gradient descent method and the Gauss-Newton algorithm. During network training, the LM propagation technique operates at the hidden layers before transmitting to the output layer. Initially, to rectify the randomly assigned output weight, it is essential to verify the accuracy of the data (Levenberg, 1944; Wu, 2008). BR is a mathematical technique that converts a nonlinear regression into a well-defined statistical issue akin to ridge regression (Burden & Winkler, 2009). ANN models were developed utilizing the LM and BR techniques. The clustering rates were established at 70%, 15%, and 15%. The ideal number of hidden neurons for each method was determined to be 8, 10, and 12, respectively. Trial and error established these values. While trial and error is a common method for determining the number of hidden layers and neurons within each layer, it is not always the most effective or efficient strategy (G. Zhang, Patuwo, & Hu, 1998). We designated demographic data such as population, gross domestic product (GDP), energy consumption (GWh), the number of sports clubs, economic indicators, and information on sports clubs as input factors. The goal of identifying these characteristics is to determine how many licensed athletes make up the output parameter. The writers regard the substantial number of licensed athletes as a reflection of a country's developmental status or potential for advancement. The independent variables included in the input parameters also signify a country's development status as either developed or developing.

Figure 2 depicts a typical architecture of an ANN for models. Population expansion, characterized by a rise in population size, leads to an expanded population base. This, in turn, leads to an increase in the number of individuals who can participate in sports, thereby expanding the population of athletes. An increase in GDP, signifying improved economic prosperity, may enable the deployment of additional resources to sports activities and the enhancement of sports infrastructure. This, in turn, increases the number of athletes. The increase in power usage can enable the development and use of more modern and technologically sophisticated sports facilities. The presence of well-illuminated and properly equipped venues promotes engagement in sports. The proliferation of sports clubs allows a larger segment of the public to participate in organized sports activities. This, in turn, directly contributes to the increase in the number of athletes.

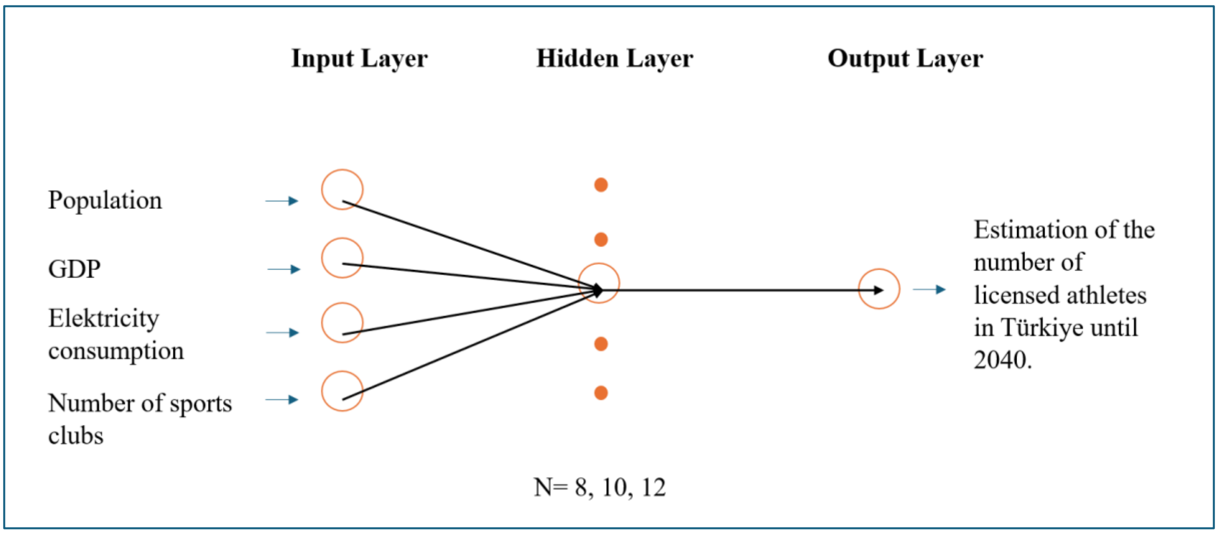

Fig. 2. Artificial Neural Network architecture (GDP: Gross Domestic Product).

2.2. Input data for artificial neural network

The data (inputs, outputs) utilized were sourced from many references (DrDataStats, 2024; TETI, 2024; TurkStats, 2024). Table 1 presents the data utilized for input and output.

Table 1. Dataset created for the years 2002-2018 (GDP: Gross Domestic Product).

Years	Population (10^6)	GDP (Dollar) (10^{10})	Elektricity Consumption (GWh) (10^3)	Number of sports clubs (10^3)	Number of athletes (10^3)
2002	65.3	23.0	132.6	6.04	278
2003	66.4	30.5	141.2	6.12	405
2004	67.6	39.0	150.0	6.36	608
2005	68.9	48.1	160.8	6.84	918
2006	70.2	52.6	174.6	7.53	1123
2007	70.6	64.9	190.2	8.59	1263
2008	71.5	74.2	198.4	9.41	1469
2009	72.6	61.7	193.4	9.98	1621
2010	73.7	73.2	210.4	10.50	1765
2011	74.7	77.4	229.4	11.07	1951
2012	75.6	78.6	241.9	11.70	2331
2013	76.7	82.3	246.1	12.41	2818
2014	77.7	79.9	257.2	31.21	3219
2015	78.7	72.0	264.2	13.71	3534
2016	79.8	86.3	273.7	13.44	3842
2017	80.8	85.1	290.4	14.01	4429
2018	82.0	79.7	291.6	15.83	4908

2.3. Relative importance analysis

The literature has not addressed the significant gap in the relative relevance analysis of ANN models (Suman, Singh, Mitra, & Kumar, 2024). Relative importance studies were performed to determine the degree to which the input parameters affect the output parameters in the models for the LM and BR algorithms. Yoon's approach was chosen for use in a relative significance equation (Eq. 3). Yoon's approach establishes the ratio of one input variable to another, accounting for the direction of contribution without considering absolute values (da Costa, de Lima, & Barbosa, 2021).

$$RI_{yoon}(\%)_{ik} = \frac{\sum_{j=1}^{h} W_{ij} V_{jk}}{\sum_{i=1}^{m} \left(\sum_{i=1}^{h=1} W_{ij} V_{jk} \right)} x100 \tag{3}$$

The symbol W_{ij} represents the connection weight between the input layer (label i) and the hidden layer (label j). Similarly, the notation W_{jk} (Çolak & Çolak, 2024; Şenol, Çolak, Elibol, et al., 2024) denotes the connection weight between the hidden unit, known as unit J, and the output unit, known as unit K.

2.4. Model performance analysis

The performance of the six models developed using the LM and BR algorithms of ANN was evaluated using the correlation coefficient (R^2) and the mean square error (MSE) criterion. The R^2 value ranges from 0 to 1.00, with values around 1.00 signifying a high-performing model (Çolak & Çolak, 2024). A reduced MSE value signifies a minimal degree of estimating mistakes (Tufaner & Avşar, 2016). The equations in Eqs. 4 and 5 encompass the model performance requirements.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (y_i - \hat{Y}_i)^2} \tag{4}$$

$$MSE = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n} \tag{5}$$

In Equations 4 and 5, n is the number of data $\hat{y}_i$ is the forecasted value, and y_i is the actual value.

3. Results and Discussion

3.1. Forecasting input using curve fitting method

For forecasting the number of licensed athletes in Türkiye for the forthcoming years, the curve-fitting approach was utilized on the input variables, yielding predictions for the period between 2019 and 2040. Table 2 presents the projections of the input variables through 2040.

Table 2. Forecasting of input parameters by curve fitting till 2040 (GDP: Gross Domestic Product).

Years	Population (10^6)	GDP (Dollar) (10^{10})	Elektricity Consumption (GWh) (10^3)	Number of Sports Clubs (10^3)
2019	82.9	96.6	306.9	18.6
2020	83.9	100.0	317.2	19.4
2021	84.9	104.0	327.5	20.1
2022	85.9	107.0	337.8	20.9
2023	86.9	111.0	348.1	21.7
2024	88.0	114.0	358.3	22.5
2025	89.0	118.0	368.6	23.3
2026	90.0	121.0	378.9	24.1
2027	91.0	125.0	389.2	24.9
2028	92.0	128.0	399.5	25.7
2029	93.1	132.0	409.7	26.5
2030	94.1	135.0	420.0	27.2
2031	95.1	138.0	430.3	28.0
2032	96.1	142.0	440.6	28.8
2033	97.1	146.0	450.9	29.6
2034	98.2	149.0	461.1	30.4
2035	99.2	152.0	471.4	31.2
2036	100.2	156.0	481.7	32.0
2037	101.2	160.0	492.0	32.8
2038	102.2	163.0	502.3	33.6
2039	103.3	167.0	512.5	34.4
2040	104.3	170.0	522.8	35.1

An analysis of the data in Table 2 enables the assessment of the impacts of the increases in Türkiye's population, economic growth, power consumption, and sports clubs from 2019 to 2030 on the number of athletes. The overall development and growth tendencies throughout this period may have also facilitated the rise in the number of athletes. Türkiye's population rose from 82.87 in 2019 to 104.3 in 2040. This signifies an estimated rise of 25.9%. The accessibility of sports and participation opportunities for a broader demographic strongly correlate with the impact of population expansion on the quantity of athletes. The growing population has led to a rise in sports club membership and an increase in the number of athletes. Türkiye's GDP has risen from $966 billion in 2019 to $1.70 trillion in 2040. This signifies an estimated rise of 76%. Economic growth facilitates the enhancement of sports infrastructure, the expansion of sports facilities, and an increase in the budget designated for sports activities. Increased resources enable a greater number of individuals to engage in sports and pursue athleticism. Furthermore, when economic prosperity rises, individuals may have greater time and financial resources to dedicate to sports (Gündoğdu & Devecioğlu, 2008). Electricity usage rose from 306.9 GWh in 2019 to 522.8 GWh in 2040. This signifies a 70.3% gain. The rise in power usage signifies advancements and expansion in the industrial and service sectors. The establishment and upkeep of contemporary, well-structured sporting facilities facilitate a rise in electricity usage. Enhanced facilities promote more participation in sports and elevate the athlete population (Çolak & Şenol, 2023). The quantity of sports clubs rose from 18,561 in 2019 to 35,143 in 2030. This indicates an augmentation of 89.3%. The dissolution of a sports club facilitates more participation in sports among a greater number of individuals. Each newly established sports club possesses a specific capability for athlete recruitment and provides professional training and development options for athletes. The rise in sports club membership can directly facilitate the growth of athletes (Taştan & Yiğit, 2023).

3.2. Forecasting of the number of licensed athletes with Levenberg-Marquardt algorithm

The LM technique was employed to predict the quantity of licensed athletes in Türkiye from 2019 to 2040 (MATLAB R2019a). In the ANN architecture, four independent variables and one dependent variable were selected. Trial-and-error methodologies determined the neuron counts to be 8, 10, and 12, respectively. The three models for LM were developed based on neuron counts of 8, 10, and 12. Setting the neuron count to eight (4-8-1) resulted in an overall regression performance of 0.9976 (Fig 3). The analysis of the MSE graph revealed that the minimum MSE value over the 12 iterations occurred in the 12th iteration, based on the parameters provided. In 2019, there was a 0.28% decline in the number of athletes compared to 2018 and a projected rise of about 105.9% in 2040. In the 4-10-1 model, the total regression coefficient was determined to be 0.99792. The minimum MSE value for the seventh iteration occurred in the fifth iteration. The model predicted a 7.4% decrease in the number of athletes in 2019 compared to 2018, which was followed by a roughly 5% increase in 2040. Upon selecting 12 neurons, the model's overall regression was determined to be 0.99728. The 6th iteration yielded the lowest MSE. The 4-12-1 model projected a 10.68% rise in the number of athletes in 2019 vs. 2018 and an almost 53% increase by 2040.

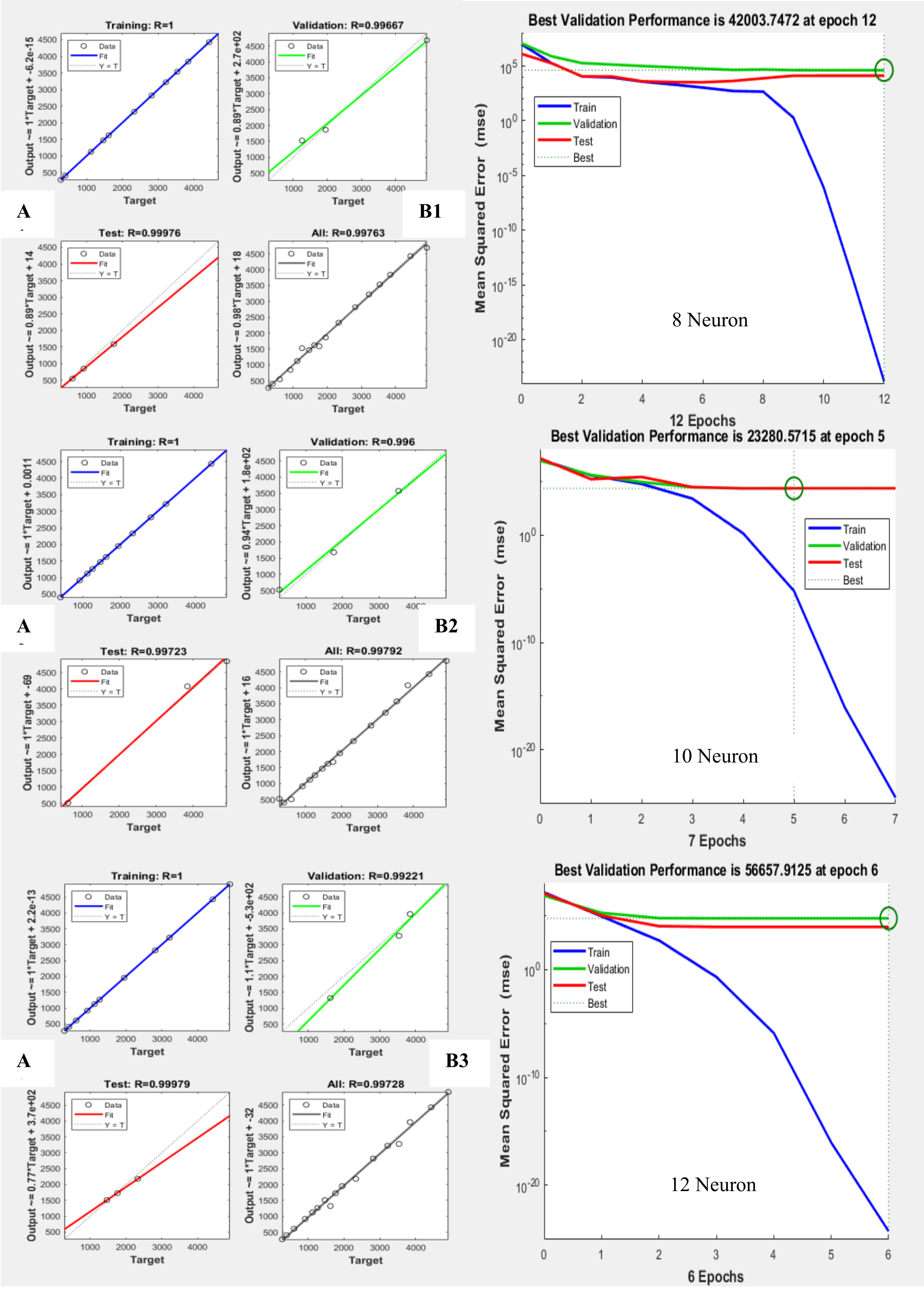

Fig. 3. General regression (A) and MSE graphs (B) of the Levenberg-Marquardt algorithm (A1, B1 for 8 neurons, A2, B2, for 10 neurons, A3, B3 for 12 neurons).

3.3. Forecasting of the number of licensed athletes with Bayesian Regularization algorithm

For forecasting the quantity of licensed athletes in Türkiye from 2019 to 2040, the BR algorithm, an ANN-based methodology, was chosen as the optimal approach. A trial-and-error methodology determined the neuron counts to be 8, 10, and 12, respectively. Three models were developed for BR utilizing the designated neuron quantities. The choice of 8 neurons yielded an overall regression performance of 0.99856. The MSE graph analysis showed that the 34th iteration achieved the minimum MSE value of 711. The values were subsequently employed to estimate the outcomes. The number of athletes rose by 5.71% in 2019 relative to 2018 and is projected to climb by 35.09% by 2040. Upon picking hidden neuron 10, the regression coefficient was estimated to be 0.99850. The 42nd iteration of the 1000 repetitions recorded the minimum MSE value (Fig 4). The model predicted a 13% increase in athletes in 2019 and a 67.82% increase by 2040 compared to 2018. With the selection of 12 neurons, the model's general regression was calculated to be 0.99854. The minimum mean square error (MSE) of the model was observed in the 431st iteration of 955 total iterations. The quantity of athletes in the 4-12-1 model rose by 5.45% in 2019 relative to 2018 and by 67% in 2040.

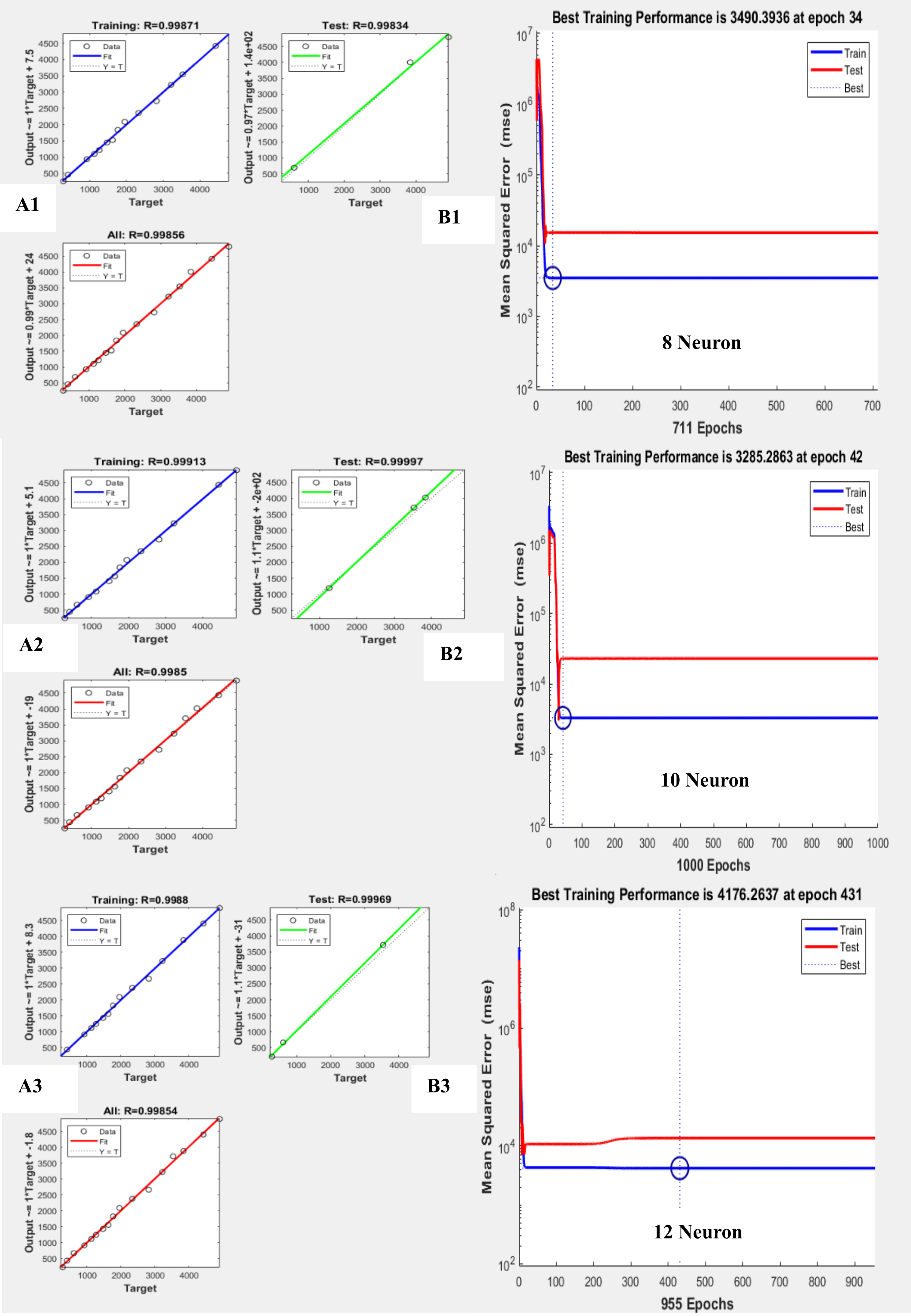

Fig 4. General regression (A) and MSE graphs (B) of the Bayesian Regularization algorithm (A1, B1 for 8 neurons, A2, B2, for 10 neurons, A3, B3 for 12 neurons).

3.4. Relative importance results

A sensitivity analysis was performed on the LM and BR models to determine the impact of input parameters on output parameters. This investigation utilized the Yoon algorithm. This investigation utilized the Yoon algorithm, an efficient technique for assessing the impact of various inputs on the output, for both the LM and BR models. Analysis of the 8-hidden-neuron variant of the LM model revealed that the input exerting the most significant influence on the output parameter was GDP. The variable with the smallest influence was the quantity of sporting clubs. The examination of the LM model with 10 hidden neurons revealed that power usage significantly influenced the output parameter, but the number of sports clubs had the least impact. Previous research revealed a divergent result in the 12-hidden-neuron variant of the LM model. The model revealed that the number of sporting clubs had the most substantial influence, while power usage had the least significant effect.

The BR model identified the population as the most influential input on the output parameter across all eight hidden neuron configurations, with the number of sports clubs exerting the least influence. The evaluations of the BR model with 10 and 12 hidden neurons yielded results comparable to those of the model with 8 hidden neurons. In all iterations, the population was recognized as the input exerting the most substantial influence, while the quantity of sports clubs was shown to have the minimal effect.

The studies illustrate the comparative influence of the many factors on the outcomes obtained across different model settings. The LM and BR models' efficiency tests give us important information for choosing the best model and optimizing the parameters. This helps us make smart decisions that will improve the model's accuracy (Table 3).

The LM series organizes its neurons in a linear configuration, each displaying a multi-layered architecture. In the BR system, the units exerting the most significant influence on neuron variables are linked (population), signifying a changeable configuration.

Table 3. Sensitivity analysis results applied to input parameters (GDP: Gross Domestic Product, LM: Levenberg-Marquardt, BR: Bayesian Regularization).

Algorithm-Parameters	Population (%)	GDP (%)	Türkiye’s Electricity Consumption (%)	Number of sports clubs (%)
LM 8 neuron	21.3	32.7	25.3	20.7
LM 10 neuron	27.2	27.0	29.4	16.3
LM 12 neuron	22.2	27.9	19.3	30.5
BR 8 neuron	48.7	15.2	24.5	11.5
BR 10 neuron	53.0	22.6	18.3	6.0
BR 12 neuron	45.3	24.6	18.3	11.9

3.5. Analysis of model performance

Table 4 presents the MSE and R^2 values derived from the implementation of two different models, LM and BR, with varied neuron counts (8, 10, 12). As the number of neurons in the LM model increases, the MSE value rises while the R^2 value declines. The MSE score of the LM 12 model attained 11,607, signifying the largest error figure recorded. This signifies that the model is less proficient at predicting the data set compared to the other models. In contrast, although the augmentation of neurons in the BR model led to a decrease in MSE values, the R^2 values remained elevated. The BR-8 model achieved the minimum MSE value of 5576. This model is regarded as the most accurate estimation of the data set. Regarding the R^2 values, both models exhibit significant effectiveness in estimating the dataset.

Table 4. Table of performance criteria of Artificial Neural Network models (LM: Levenberg-Marquardt, BR: Bayesian Regularization)

Models- Criterion	MSE	R^2
LM 8 neuron	9763	0.9949
LM 10 neuron	8208	0.9956
LM 12 neuron	11,607	0.9939
BR 8 neuron	**5576**	**0.9970**
BR 10 neuron	6742	0.9964
BR 12 neuron	5824	0.9969

3.6. Comparative analysis of licensed athletes in Türkiye

The model estimations are depicted in various hues in Fig. 5. Both strategies demonstrate an increasing trajectory and produce comparable outcomes. The 8-neuron model for LM initially provides a highly accurate estimation of the actual data. Nonetheless, it forecasts a substantial rise during the years after 2030. This may indicate that the model is potentially overfitting, as the estimations significantly exceed the prevailing trend. With 10 neurons, its performance appears to align most closely with the actual data. It accurately aligns with the actual data both initially and in subsequent years. In the 12-neuron model, the estimate is marginally greater than in the 10-neuron LM model. While it may somewhat overstate the time after 2035, it remains within a plausible range overall. In BR models, the 8-neuron configuration yields lower estimates than the 8-neuron LM model. While it appears effective in aligning with real data initially, it subsequently falls short of actual data post-2030. The 10-Neuron model appears to yield the most accurate estimations relative to the actual data and demonstrates the most consistent upward trend. This model effectively encapsulated the dataset's overall outcomes and may furnish a plausible future projection. The projections of the 12-neuron model indicated more severe breakdowns compared to the other BR models. The models that provide the most accurate estimations for actual data and most effectively represent the overall trend are the LM 10 neuron and BR 10 neuron models. The authors suggest that employing model features and training with additional data may mitigate overfitting and underfitting durations.

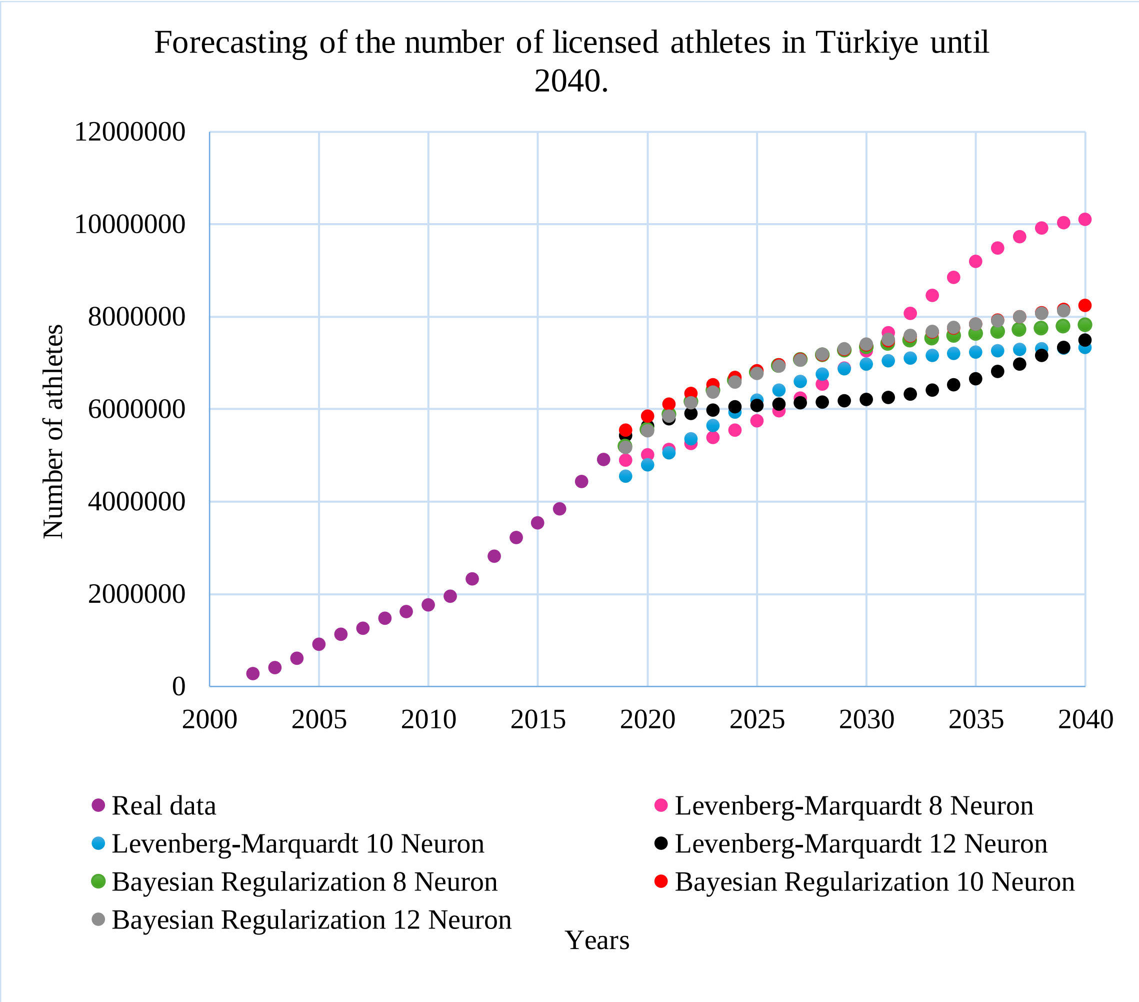

Fig. 5. Estimation of the number of licensed athletes in Türkiye until 2040 using ANN.

4. Conclusions

This study employed six ANN models with varying neuron counts to forecast the number of licensed athletes in Türkiye by 2040. The ANN models were constructed with LM and BR algorithms, with neuron architectures of 8, 10, and 12 evaluated for both techniques. The analysis reveals that the 10-neuron configuration, among the models created with the LM algorithm, produced forecasts most closely aligned with the actual data. This configuration demonstrated significant concordance with the dataset, particularly up to 2018, and displayed a consistent upward trend during this timeframe. The 12-neuron model, conversely, indicated issues with overfitting by producing overestimations, particularly for the period following 2030. Among the models developed using the BR algorithm, the 10-neuron configuration exhibited the most favorable behaviors and realistic predictions. Upon running models with both low and high neuron counts, the 10-neuron structure demonstrated superior improvement over the others. This means that estimates will be very accurate until 2040.

The study demonstrated that ANN might serve as an effective instrument for forecasting the number of licensed athletes in Türkiye for future years. To ensure calculation precision, model selection and neuron quantity optimization are crucial. Future research should focus on enhancing model performance using diverse validation strategies and evaluating them on more extensive datasets.

Furthermore, precisely quantifying the number of licensed athletes is essential for athlete training and the formulation of sports regulations in Türkiye. These forecasts enable government and pertinent sports organizations to better efficiently allocate resources, make capital investments, and support programs for youth athletes. The rising number of athletes illustrates Türkiye's potential for competition in sports at both national and international levels. This study offers critical insights to assist policymakers in deliberately shaping future sports development initiatives. It serves as a crucial reference for identifying measures to enhance the training and development chances for athletes.

Authors' Contributions

All authors contributed equally to the study.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The author declares that this study complies with Research and Publication Ethics.

References

Atasoy, M., Dalkılıç, M., & Uğraş, S. (2017). Estımatıon of Lıcensed Sportsman-Woman in Area of Martıal Sports by Artıfıcıal Neural Networks. *Kilis 7 December University Journal of Physical Education and Sports Sciences, 1*(1), 33-37.

Bas, E., Egrioglu, E., & Cansu, T. (2024). Robust training of median dendritic artificial neural networks for time series forecasting. *Expert Systems with Applications, 238*, 122080.

Başkan, A. H., Özgül, F., Kolukısa, Ş., Çolak, H., & Başkan, A. H. (2020). Giresun University Faculty of Sport Sciences Students Usefulperceptions of theconcept of "Sports Management". *Eurasian Journal of Researches in Social and Economics, 7*(10), 58-67.

Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied computing and informatics, 15*(1), 27-33.

Burden, F., & Winkler, D. (2009). Bayesian regularization of neural networks. *Artificial neural networks: methods and applications*, 23-42.

Çolak, H., & Çolak, E. (2024). Estimation of Prevalence Distribution of Pre-obesity by Gender in Türkiye Using Artificial Neural Networks and Time Series Analysis. *The Black Sea Journal of Sciences 14*(3), 1340-1359.

Çolak, H., & Şenol, H. (2023). *Estimating the Number of Licensed Athletes in Turkey with Artificial Neural Networks until 2030*: Academic Evaluations in the Field of Sports Sciences - 7, Duvar Publications.

da Costa, N. L., de Lima, M. D., & Barbosa, R. (2021). Evaluation of feature selection methods based on artificial neural network weights. *Expert Systems with Applications, 168*, 114312.

Dalkılıç, M., Atasoy, M., Yİğİt, Ş., & Mamak, H. (2017). Estimation of Number of Disabled Licensed Sports by Artificial Neural Networks. *The Journal of Academic Social Science*.

Dalkılıç, M., Kargün, M., Kızar, O., & Genç, H. (2017). Estimation of Licensed Number of Number of Competitors in the Wrestling of Artificial Neural Networks. *Kilis 7 December University Journal of Physical Education and Sports Sciences, 1*(1), 15-19.

Dalkılıç, M., Mamak, H., Atasoy, M., & Mihriay, M. (2017). Forecast of Licensed Number of Living Sports in Water Trophies, Water Ball and Swimming Areas by Artificial Neural Networks.

DrDataStats. (2024). www.drdatastats.com.

Exel, J., & Dabnichki, P. (2024). Precision Sports Science: What Is Next for Data Analytics for Athlete Performance and Well-Being Optimization? *Applied Sciences, 14*(8), 3361.

Gomis-Gomis, M. J., Pena-Pérez, X., & Pérez-Turpin, J. A. (2023). Sustainability and sports science: A new way for a better future. *Sustainability and Sports Science Journal, 1*(1), 1-2.

Hulteen, R. M., Smith, J. J., Morgan, P. J., Barnett, L. M., Hallal, P. C., Colyvas, K., & Lubans, D. R. (2017). Global participation in sport and leisure-time physical activities: A systematic review and meta-analysis. *Preventive medicine, 95*, 14-25.

Ishwarya, K., & Nithya, A. A. (2021). *Relative analysis and performance of machine learning approaches in sports.* Paper presented at the 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA).

Jauhiainen, S., Kauppi, J.-P., Leppänen, M., Pasanen, K., Parkkari, J., Vasankari, T., . . . Äyrämö, S. (2021). New machine learning approach for detection of injury risk factors in young team sport athletes. *International journal of sports medicine, 42*(02), 175-182.

Kasera, M., & Johari, R. (2021). *Prediction using machine learning in sports: a case study.* Paper presented at the Data Analytics and Management: Proceedings of ICDAM.

Kelly, S. J., Derrington, S., & Star, S. (2022). Governance challenges in esports: a best practice framework for addressing integrity and wellbeing issues. *International Journal of Sport Policy and Politics, 14*(1), 151-168.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics, 2*(2), 164-168.

Me, E., & Unold, O. (2011). Machine learning approach to model sport training. *Computers in human behavior, 27*(5), 1499-1506.

Mertala, P., & Palsa, L. (2024). Running free: recreational runners' reasons for non-use of digital sports technology. *Sport in Society, 27*(3), 329-345.

Nguyen, N. H., Nguyen, D. T. A., Ma, B., & Hu, J. (2022). The application of machine learning and deep learning in sport: predicting NBA players' performance and popularity. *Journal of Information and Telecommunication, 6*(2), 217-235.

Qi, Y., Sajadi, S. M., Baghaei, S., Rezaei, R., & Li, W. (2024). Digital technologies in sports: Opportunities, challenges, and strategies for safeguarding athlete wellbeing and competitive integrity in the digital era. *Technology in Society*, 102496.

Ryan, L., & Doody, O. (2024). The treatment, outcomes and management of hand, wrist, finger, and thumb injuries in the professional/amateur contact sport athletes: A scoping review. *International Journal of Orthopaedic and Trauma Nursing*, 101108.

Seefeldt, V. D., & Ewing, M. E. (1997). Youth Sports in America: An Overview. *President's council on physical fitness and sports research digest*.

Suman, S., Singh, S., Mitra, S., & Kumar, M. (2024). Deep neural network model for predicting thermal-hydraulic performance of a solar air heater with artificial roughness: Sensitivity, generalization capacity, and computational efficiency. *Process Safety and Environmental Protection*.

Şenol, H. (2021). Methane yield prediction of ultrasonic pretreated sewage sludge by means of an artificial neural network. *Energy, 215*, 119173.

Şenol, H., Çolak, E., Elibol, E. A., Hassaan, M. A., & El Nemr, A. (2024). Optimisation of biochar dose in anaerobic co-digestion of green algae and cattle manure using artificial neural networks and response surface methodology. *Chemical Engineering Journal*, 152750.

Şenol, H., Çolak, E., & Oda, V. (2024). Forecasting of biogas potential using artificial neural networks and time series models for Türkiye to 2035. *Energy*, 131949.

Şenol, H., Dereli, M. A., & Özbilgin, F. (2021). Investigation of the distribution of bovine manure-based biomethane potential using an artificial neural network in Turkey to 2030. *Renewable and Sustainable Energy Reviews, 149*, 111338.

TETI. (2024). Türkiye Electricity Transmission Inc. https://www.teias.gov.tr/.

Tufaner, F., & Avşar, Y. (2016). Effects of co-substrate on biogas production from cattle manure: a review. *International journal of environmental science and technology, 13*, 2303-2312.

TurkStats. (2024). www.tuik.gov.tr.

Wilkens, S. (2021). Sports prediction and betting models in the machine learning age: The case of tennis. *Journal of Sports Analytics, 7*(2), 99-117.

Wu, J.-M. (2008). Multilayer potts perceptrons with Levenberg–Marquardt learning. *IEEE transactions on neural networks, 19*(12), 2032-2043.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting, 14*(1), 35-62.

Zhang, L., & Li, N. (2022). Material analysis and big data monitoring of sports training equipment based on machine learning algorithm. *Neural Computing and Applications, 34*(4), 2749-2763.

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing, 237*, 350-361.

Zhu, J., Hu, C., Khezri, E., & Ghazali, M. M. M. (2024). Edge intelligence-assisted animation design with large models: a survey. *Journal of Cloud Computing, 13*(1), 48.

Zhu, P., & Sun, F. (2020). *Sports athletes' performance prediction model based on machine learning algorithm.* Paper presented at the International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019: Applications and Techniques in Cyber Intelligence 7.