APPLICATION OF NEURAL NETWORKS IN HALF MARATHON AND MARATHON

Dubljanin Milan, Savić Milan, Dejanović Milan
Faculty of Sciences and Mathematics, University of Priština - Kosovska Mitrovica, Serbia

ORIGINAL SCIENTEFIC ARTICLE
DOI: 10.5937/ATAVPA25243D

***Abstract**: Artificial intelligence is the fastest-growing field of computer science. Its use is growing and yet to be fully uncovered. Using deep learning and datasets of specific data, artificial neural networks can be trained to help users in everyday life. This paper aims to make neural networks to help get the best time in a half marathon and marathons. Many factors influence marathon race times, such as temperature, precipitation, sunshine or atmospheric pressure, sex, age and physical preparation and current condition of runners. A dataset from smartwatches will be used to determine the condition of runners, as well as a dataset Berlin marathon dataset. Combining the analysis of these two datasets, neural networks will predict if the ideal time was achieved.*
***Key words**: neural networks, artificial intelligence, dataset, smartwatches, marathon*

INTRODUCTION

Neural networks are computational models inspired by the structure and functionality of neurons in the human brain. Each network consists of layers of perceptrons, where each perceptron receives inputs, applies weights, and generates an output. A typical deep neural network consists of an input layer, one or more hidden layers, and an output layer, enabling it to learn complex patterns from data (Goodfellow, et al., 2016). Perceptrons function similarly to biological neurons, combining multiple inputs to generate a single output, making them the fundamental building blocks of neural network operations.

Artificial intelligence (AI) has become integral to modern life, revolutionising industries and everyday activities. AI applications, including voice assistants like Alexa and Google Assistant, automatic license plate recognition systems, and machine learning models like ChatGPT, demonstrate its ubiquity and transformative potential (Russell & Norvig, 2020). Within AI, deep learning is a specialised subfield that utilises neural networks to analyse large datasets, enabling systems to learn and make predictions (Goodfellow et al., 2016).

This study uses deep learning to predict marathon and half marathon finishing times by analysing weather, biometric and environmental data. To achieve this, the dataset combines meteorological variables with simulated biometric data of runners derived from established mathematical models. These models include the Karvonen formula (Karvonen et al., 1957), which uses heart rate reserve (HRR) and personalised percentages to estimate the time spent in various heart rate zones. The five zones analysed are Zone 1 (Very Light), Zone 2 (Light), Zone 3 (Moderate), Zone 4 (Hard), and Zone 5 (Maximum). This approach enables the creation of a comprehensive dataset suitable for training neural networks to predict marathon performance.

The dataset used in this research encompasses data from the Berlin Marathon (1974–2019), including meteorological variables (e.g., average, maximum, and minimum temperatures, precipitation, atmospheric pressure, hours of sunshine, and cloud coverage) and anonymised runner statistics (e.g., gender and age group). Since historical heart rate data was unavailable, simulated heart rate zone data was generated to augment the dataset, enabling a more detailed analysis of runner performance and physiological responses.

Riegel (1981) developed a formula to estimate race finishing times across different distances, taking into account a runner's capabilities. This formula has been widely used in sports and medical research to evaluate and compare runner performance over various distances. Dracopoulos's study (2017) demonstrated that neural networks could significantly improve the accuracy of marathon time predictions compared to traditional methods like Riegel's formula. Similarly, our work aims to extend this approach by integrating additional factors such as meteorological conditions and detailed heart rate zone analyses, further enhancing prediction accuracy.

Akay et al. (2011) developed an artificial neural network (ANN) model to predict maximal oxygen uptake (VO2max) in adults during a submaximal treadmill running test. By utilising input data such as gender, age, body weight, running speed, and heart rate, their model achieved significantly higher accuracy than traditional multiple linear regression models. Their study highlights the advantages of ANN in providing safer, faster, and more efficient methods over costly and potentially risky maximal tests. This underscores the transformative potential of neural networks in enhancing predictive capabilities across diverse domains of sports science.

Lerebourg et al. (2022) compared the performance of ANN and KNN algorithms for predicting race outcomes based on specific input parameters. Their findings revealed an exceptionally high correlation between actual and predicted results ($r > 0.90$), emphasising the substantial potential of AI in improving the accuracy of race performance predictions. By employing a simple neural network architecture and leveraging detailed heart rate zone data, we aim to predict marathon finishing times, highlighting the role of heart rate dynamics in performance estimation. By combining biometric data with meteorological factors, this study demonstrates how AI and basic predictive models can enhance sports training, improve performance predictions, and lay the groundwork for more personalised fitness plans.

DATASET

In this paper, we use the Berlin marathon dataset, which includes data about the weather and runners for each marathon year from 1974 until 2019. The dataset is publicly available on Kaggle under the CC0 license. Weather data consists of the following information: year, precipitation, sunshine, clouds, atmospheric pressure, average temperature, max temperature, and min temperature. Runner data is anonymised and includes gender, age group, and time. By combining these two data sources, this dataset allows for a detailed analysis of how weather conditions such as temperature, precipitation, sunshine, and atmospheric pressure might influence marathon runner performance over the years. The data can be used to explore patterns and correlations between environmental factors and finishing times, offering insights into the impact of different conditions on marathon performances over the years.

Since the dataset includes data from when technology was not as advanced as today, it has been supplemented with heart rate zone data from all participants, which ideally would have been tracked using smartwatches. These values were simulated using the Karvonen formula (Karvonen et al., 1957), which utilises heart rate reserve (HRR) and personalised percentages to define the boundaries of each heart rate zone and calculate the duration of time spent in each zone. These metrics can now be monitored and tracked using modern smartwatch technology. In future work, smartwatch data could be directly integrated into the neural network models, providing even more accurate and real-time data for performance analysis. In this study, however, the data was simulated very precisely to closely match the values that would have been obtained using such wearable devices.

The Karvonen formula for calculating maximum heart rate (Karvonen et al., 1957):

$$HR_{max} = 220 - age$$

where HR_{max} is the maximum heart rate based on age. The calculation of heart rate reserve (HRR) (Karvonen et al., 1957):

$$HRR = HR_{max} - HR_{rest}$$

where HR_{rest} is the resting heart rate. Finally, to calculate the boundaries of heart rate zones based on maximum heart rate (HRmax), resting heart rate (HR_{rest}), and the desired intensity (expressed as a percentage), we used the formula (Karvonen et al., 1957):

$$Running\ Zone = (HR_{max} - HR_{rest}) \times \%\text{Intensity} + HR_{rest}$$

where %Intensity represents the desired level of physical activity intensity.

For each runner, the maximum heart rate was estimated based on their age, while the resting heart rate was randomly generated within the range of 50 to 90 beats per minute to simulate realistic conditions for each participant. The boundaries for each of the five heart rate zones (from Zone 1 – Very

Light, to Zone 5 – Maximum) were then calculated using the difference between the maximum and resting heart rates (heart rate reserve, HRR).

Simulated data for the time spent in each heart rate zone were generated by assigning random percentages of time spent in each of the five intensity levels. Specifically, the time spent in each zone was randomly allocated within ranges that corresponded to realistic values based on physical exertion during the race (Zone 1: 15-25%, Zone 2: 30-50%, Zone 3: 20-35%, Zone 4: 5-10%, Zone 5: 0.5-1%).

These simulated values allowed for the analysis of performance based on heart rate dynamics and the prediction of race finishing times. While it would ideally be based on actual data collected through smartwatches, these simulated values closely mimic the actual heart rate dynamics of runners and were used for training the model. A sample of the new dataset is presented in Table 1.

Table 1 Sample of the Berlin marathon dataset with weather and runner data

ID	**116219**	**839581**	**707422**	**187499**	**80504**
YEAR	1992	2019	2015	1998	1990
GENDER	1	1	1	1	1
AGE	18	30	20	30	40
TIME	3:30:18	2:56:39	4:13:58	3:30:06	3:00:00
TIME_ss	12618	10599	15238	12606	10800
AVG_TEMP_C	13.9	14	10.1	13.8	15.2
MAX_TEMP_C	22.7	16.4	16.9	19.8	18.5
MIN_TEMP_C	6.7	11.2	5.2	10.3	11.2
PRECIP_mm	0	8	0	0	0.5
ATMOS_PRESS_mbar	1011.1	993.62	1021.4	1026.8	1000.4
SUNSHINE_hrs	10.2	0.7	8.93	9.9	0.1
CLOUD_hrs	0.3	6.8	2	1	7.3
resting_hr	61	71	77	53	74
max_hr	202	190	200	190	180
Zone 1 (Very Light) Time	2569	1972	2854	2801	2679
Zone 2 (Light) Time	6146	5031	6981	6101	4240
Zone 3 (Moderate) Time	2771	2856	3907	2896	3154
Zone 4 (Hard) Time	1010	636	1418	722	669
Zone 5 (Maximum) Time	122	104	78	86	58

NEURAL NETWORK

Neural networks are computational models designed to emulate the functioning of brain neurons. Over the years, this technology has garnered sustained interest within the artificial intelligence community (Cary & Upham, 1992).

Fig. 1 shows that a deep learning neural network consists of input, hidden and output layers. Deep learning neural networks can have many layers (hence "deep") that can learn from vast amounts of data. Deep learning neural networks are named for consisting of multiple layers that enable them to learn from vast amounts of data.

After simulating the heart rate zone data and gathering the weather and runner data, we utilised a neural network approach to analyse the impact of these variables on marathon performance. Neural networks are particularly well-suited for this complex, multi-dimensional data analysis, as they can capture non-linear relationships between inputs (e.g., weather conditions, heart rate dynamics) and outputs (e.g., marathon finishing times).

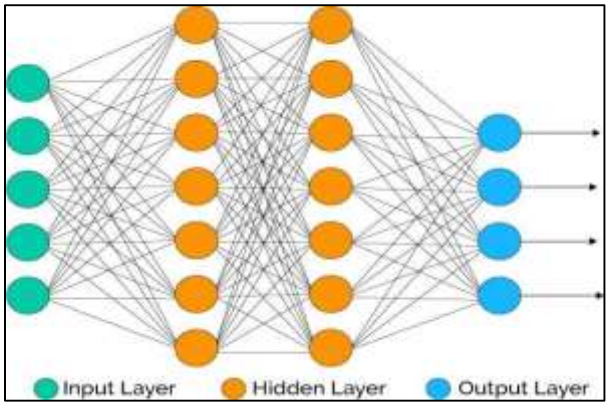

Fig. 1 Neural network layers

This study used a supervised learning neural network with multiple hidden layers. The input features included environmental factors like temperature, precipitation, and atmospheric pressure, as well as each runner's simulated heart rate data. The network output was the predicted race time, evaluated against actual race results.

The neural network was trained using a subset of the data, and performance was evaluated using metrics such as Mean Absolute Error (MAE) and R^2. We employed the Adam optimiser for efficient training, and dropout layers were added to prevent overfitting. The model learned to identify patterns in how heart rate zone distribution interacted with weather conditions to influence race performance.

By integrating this neural network approach, we gained more profound insights into the factors influencing marathon performance and provided more accurate predictions for future races. In future work, we aim to incorporate more advanced neural network techniques, such as recurrent networks, to account for temporal dependencies and better model the evolving dynamics of marathon races over time.

EXPERIMENTS AND RESULTS

We conducted five network tests, simulating the time spent in each heart rate zone. During these tests, we sequentially fed values for Zones 1 through 5 into the network. Our findings indicate that the time spent in each zone significantly impacts race outcomes. Fig. 2 shows the actual time of the first observed participant alongside the predicted split times generated by the neural network.

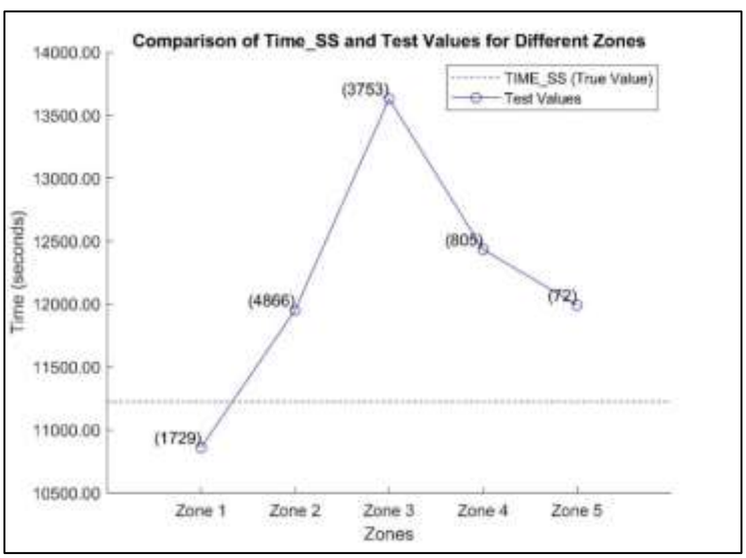

Fig. 2 Comparison of Time_SS and test values for different zones for the first participant

The exact measured time of the first observed participant was 11,225 seconds, which was the total time spent across all zones. We tested the predicted race time by feeding individual zone times into the neural network. For Test 1, we input the time spent in Zone 1, which was 1,729 seconds. The network predicted a split time of 10,860.29 seconds. When the participant spent 4,866 seconds in Zone 2, the predicted split time increased to 11,951.50 seconds, and so on.

We further tested by varying the time spent in Zone 1. Increasing the time in Zone 1 from 1,729 to 3,000 seconds resulted in a predicted race time of 14,396.04 seconds, while decreasing it from 1,729 to 600 seconds led to a predicted time of 7,719.56 seconds. This demonstrated that the race time would decrease when more time was spent in higher-intensity zones.

Fig. 3 presents the actual race time for the second observed participant, along with the corresponding predicted split times from the neural network.

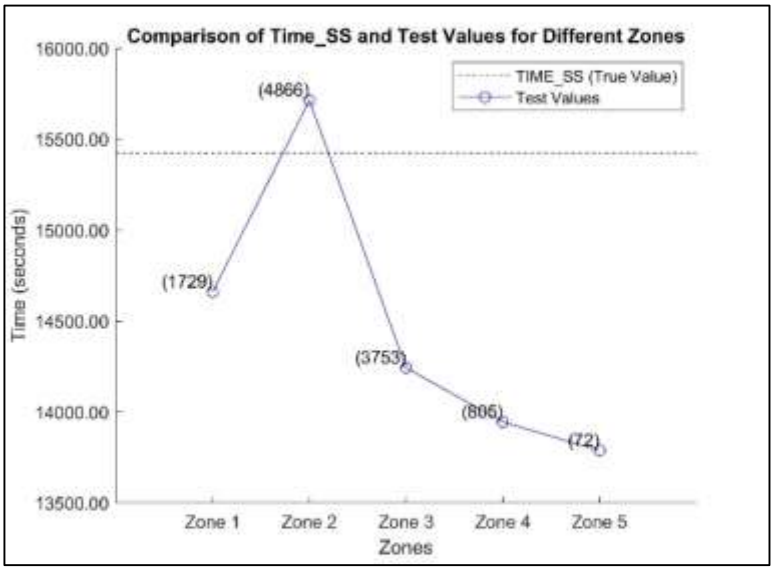

Fig. 3 Comparison of Time_SS and test values for different zones for the second participant

The exact measured time of the second observed participant was 15,424 seconds. For 4,700 seconds spent in Zone 1, the neural network predicted a race time of 19,356.73 seconds. In contrast, for 600 seconds spent in Zone 1, the expected time decreased to 7,951.07 seconds. This demonstrates that increasing the running time in Zone 1 results in a slower race finish while reducing time spent in Zone 1 leads to a faster finish.

These results underscore the significant impact of time spent in different heart rate zones on overall marathon performance. The same results can be obtained for the half marathon by using a dataset from the half marathon. Spending more time in lower heart rate zones typically reflects a moderate-intensity race, while more time in higher heart rate zones indicates more tremendous effort and intensity. Understanding the distribution of effort across these zones can offer valuable insights into optimising training, enhancing performance, and assessing a runner's capabilities and progress.

The neural network successfully identified a relationship between the final race time and the time spent in different heart rate zones. Satisfactory prediction results were achieved for each zone, with the model considering other parameters such as age, gender, and temperature. By varying the time spent in Zone 1 (while keeping other values fixed), we were able to predict potential changes in the final race time, showing whether the runners could improve or worsen their performance by adjusting their time in this zone.

We observed that increasing the time spent in Zone 1, the least demanding zone led to a longer race time. Conversely, decreasing the time in Zone 1 resulted in a faster finish. This occurs because the runner would spend more time in higher-intensity zones (Zones 2, 3, 4, and 5), which require tremendous effort and contribute to a quicker finish**.**

CONCLUSION

Based on the results of this study, it is clear that the time spent in different heart rate zones plays a significant role in determining a runner's final race time. These findings emphasise the importance of managing race intensity effectively to optimise marathon and half marathon performance. The experiment demonstrated that modifying the time spent in specific heart rate zones can significantly influence overall race time, with longer durations in lower-intensity zones leading to slower times, while reducing time in these zones could result in faster finishes.

The results also highlighted that individual characteristics, such as age, gender, physical conditioning, fitness level, and external factors like temperature, precipitation, and atmospheric pressure, contribute to the final race time. This underscores the need for personalised training and race strategies, as the effect of spending time in different heart rate zones can vary widely from one runner to another.

By applying neural networks, this study successfully analysed a combination of marathon data and heart rate zone data to predict race performance. These insights suggest how runners can optimise their strategies for the best possible time. However, given the complexity of the factors affecting marathon performance, further research is needed, which could involve analysing larger datasets, exploring advanced neural network techniques, and considering additional variables that may influence race outcomes.

REFERENCES

1. Akay, M., Zayid, E.I.M., Aktürk, E., & George, J. (2011). Artificial neural network-based model for predicting VO2max from a submaximal exercise test. *Expert Systems with Applications, 38*, 2007-2010. https://doi.org/10.1016/j.eswa.2010.07.135.
2. Cary, P., & Upham, W. (1992). An evaluation of neural networks. Geophysics, 11, 45-47. https://doi.org/10.1190/1.1436904.
3. Dracopoulos, D. (2017). A better predictor of marathon race times based on neural networks. In Proceedings of the International Conference on Computational Science and Its Applications (pp. 293-299). https://doi.org/10.1007/978-3-319-71078-5_25.
4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
5. Karvonen, M. J., Kentala, E., & Mustala, O. (1957). The effects of training on heart rate: A longitudinal study. *Annales Medicinae Experimentalis et Biologiae Fenniae*, 35(3), 307-315. PMID: 13470504.
6. Lerebourg, L., Saboul, D., Clémençon, M., & Coquart, J. (2022). Prediction of Marathon Performance using Artificial Intelligence. *International Journal of Sports Medicine*. https://doi.org/10.1055/a-1993-2371.
7. Riegel, P.S. (1981). Athletic records and human endurance: A time-vs.-distance equation describing world-record performances may be used to compare the relative endurance capabilities of various groups of people. American Scientist, 69(3), 285-290. http://www.jstor.org/stable/27850427.
8. Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education.

PRIMENA NEURONSKIH MREŽA U POLUMARATONU I MARATONU

Dubljanin Milan, Savić Milan, Dejanović Milan

Sažetak*: Veštačka inteligencija je najbrže rastuća oblast računarskih nauka. Njena upotreba je sve veća, a još uvek nije u potpunosti istražena. Korišćenjem dubokog učenja i skupova podataka, veštačke neuronske mreže mogu biti obučene da pomognu korisnicima u svakodnevnom životu. Ovaj rad ima za cilj da napravi neuronske mreže koje će pomoći u postizanju najboljeg vremena u polumaratonima i maratonima. Mnogi faktori utiču na vreme trke, kao što su temperatura, padavine, sunčevi zraci ili atmosferski pritisak, pol, starost, fizičke pripreme i trenutno stanje trkača. Skup podataka prikupljen sa pametnih satova biće korišćen za određivanje stanja trkača, kao i skup podataka sa Berlinskog maratona. Analizom ova dva skupa podataka, neuronske mreže će predvideti da li je postignuto najboljeg vreme.*

Ključne reči*: neuronske mreže, veštačka inteligencija, skup podataka, pametni satovi, maraton*