Deep-Racing: An Embedded Deep Neural Network (EDNN) Model to Predict the Winning Strategy in Formula One Racing

Syeda Sitara Wishal Fatima* and Jennifer Johrendt

***Abstract*—This paper presents an embedded deep neural network model to predict the driver rank and the optimum pitstop strategy. In formula one racing, the race strategy is critical to determine optimal pitstops and finish the race in the best possible position. Considering a system with only one racing car, the pitstop can be decided just by looking at the degradation of the tires. But in reality, the formula one environment is more complex, and multiple probabilistic factors (like safety car phases, opponent strategy, and overtaking) influence the pitstop decision. Deep-Racing is a prediction and decision algorithm for formula one racing cars that uses neural networks with embedding layers. The algorithm is developed after carefully reviewing formula one racing and appropriate statistical modeling techniques, which can be trained for pre-race and real-time predictions during the race using the data from previous laps. Deep-Racing has the potential to help team principals and race engineers to decide the optimized strategy for making pitstops. It is trained on the data from seasons 2015-2022. This project is the first to utilize an embedded layer in motorsport racing predictions, and the results show an improvement in predictive accuracy compared with the previously available literature. This paper significantly expands the previous research in this field and proposes trends in the data available from the latest seasons.**

***Index Terms*—Neural networks applications, formula one racing, embedded deep neural networks, predictive analysis**

I. Introduction

Formula one racing has the world's most extensive viewing of all the circuit motorsport, with a combined viewership of 445 million in 2021 [1, 2]. Formula one cars are a marvel of engineering and have been considered among the fastest-designed vehicles. Formula one, or simply F1, is sanctioned by Fédération Internationale de l'Automobile (FIA), which sets the rules and regulations for the races, including the car design, frequency of pitstops, and choice of tires. Multiple teams, including Ferrari, Mercedes, and McLaren, have participated in the world championship every year since its inaugural season in 1950, involving a series of Grand Prix [3]. Grand Prix is a race performed on closed road circuits of varying lengths and complexities across various parts of the world, including Monaco, Abu Dhabi, Singapore, Monza, and Silverstone. F1 racing dominates the betting sites where sportsbooks and fans are interested to know the prediction of the winner before the race day. The goal of a team in F1 racing is simple – be the first one to cross the chequered flag. The race result depends not only on the driver or the car's pace but also, on multiple real-time stochastic elements such as the timing and duration of pitstops, tire degradation, and weather conditions. Although the driver's skills and the car's speed are the major factors impacting the race results, these two points are highly subjective. The other critical aspect of F1 racing is the pitstop strategy which is positively correlated with the race results [4]. In F1 racing, the drivers make pitstops during the race for multiple reasons, including tire replacement, refueling, or repairing broken parts, if possible. After the 2009 season, car refueling was skipped due to safety concerns, and now pitstops have become a synonym for tire strategy.

This paper presents a data-driven Embedded Deep Neural Network (EDNN) design to predict the race results in the form of driver rank and optimal pitstop strategy in real-time, which can be helpful for team principals and engineers to make critical pitstop decisions for winning. The pitstop strategy includes deciding when and how many pitstops should be made in the given racing conditions by predicting the driver's racing position in every lap.

II. Background

Formula one car are one of the fastest racing cars approaching the speed of 300km/h [5]. However, for entering and remaining in the pitlane, the car's speed is limited to 60km/h [6]. Also, the race car remains stationary during the pitstop, which adds to the total time to complete the race. Hence, every team must optimize the pitstops as much as possible. Formula one car are one of the fastest racing cars approaching the speed of 300km/h [5]. However, for entering and remaining in the pitlane, the car's speed is limited to 60km/h [6]. Also, the vehicle remains stationary during the pitstop, which adds to the total time to complete the race. Hence, every team must optimize the pitstops as much as possible.

The teams want to make a pitstop to make up for tire degradation during the race and have more speed. The fresh tires give the car more speed resulting in more effortless overtaking maneuvers, especially when the rival is racing on an older set of tires—this sudden pace increases and balances the time lost in making a pitstop. Naturally, every team wants to make the pitstop at the right time, and balancing the pitstop timing, frequency, and duration becomes the central tenant of the winning race strategy. Typically, there are three categories of compound tires used in F1 racing Hard, Medium, and Soft. The lap time of the car changes with the change in the type of tire used. Fig. 1 shows the graph for lap time and the lap number for these three types of tires.

In modern-day F1 racing, where the race positions are changed with the 10th or even 100th of a second, the teams

Manuscript received November 2, 2022; revised November 24, 2022; accepted February 20, 2023.

Syeda Sitara Wishal Fatima and Jennifer Johrendt are with the Department of Mechanical, Automotive and Materials Engineering, University of Windsor, Canada.

*Correspondence: fatima92@uwindsor.ca (S.S.W.F.)

doi: 10.18178/ijml.2023.13.3.1135

try to save every millisecond spent in the pitlane, trying not to lose the advantage over other players. A recent example of the importance of pitstop strategy is evident from Lando Norris winning third place in a McLaren at the 2022 Italian Grand Prix. While having only the seventh fastest lap time, the right pitstop strategy made him compete with strong teams like Ferrari and Mercedes, securing more wins throughout the season. Fig. 2 shows the impact of the pitstop strategy, including the lap to perform the pitstop and the type of tire transition. For example, it can be inferred that switching to hard tires after the first pitstop results in shorter lap time.

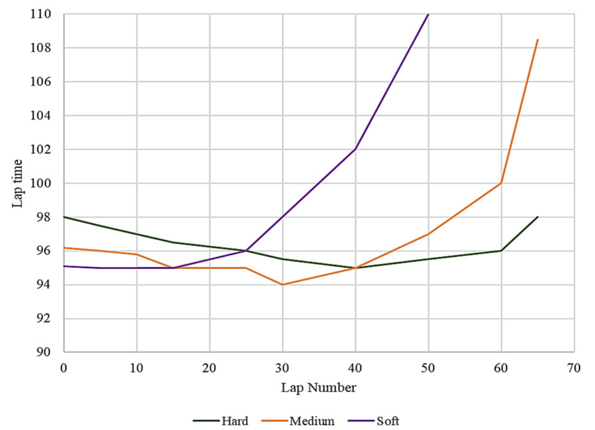

Fig. 1. Lap time for hard, medium, and soft compound tires for different lap times for f1 racing seasons 2015-2022. The data is taken from the source cited in section III-A.

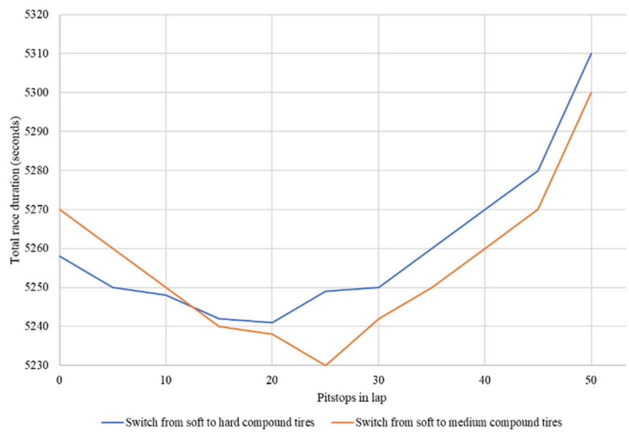

Fig. 2. The race duration as a function of pitstop lap for two types of tire strategies [4].

A. Formula One Racing Terminologies and Rules

It is a frequent scenario in the F1 world to witness a car hitting the barriers because of multiple factors such as under-steering, over-steering, or lock-up. So, when there is any accident on a track, a full course yellow (FCY) phase is deployed where a physical or a virtual safety car is deployed, and the drivers must follow specific rules. During the virtual safety car (VSC) phases, drivers must reduce their speed, lengthening the lap time by 140% [4]. In the safety car (SC) phase, a real car drives ahead of all the cars on the track, increasing the lap time by 160%, as the drivers are not allowed to overtake this car [4]. This FCY scenario makes an ideal point in the race to make a pitstop, as the time taken to enter and exit the pitlane is the same as passing the track under a safety car. Therefore, the choice of pitstops changes dramatically under an FCY phase and impacts the race results in the most. The reason for stating the safety car procedure here is to use the FCY phase to decide the optimum lap for a pitstop. The importance of the FCY phase on race results is evident from the championship-deciding race 2021 Abu Dhabi Grand Prix, where the choice of pitstop before and during safety car changed the race results. This FCY phase is discussed in detail by Heilmeier and Thomaser *et al.* [7].

B. Related Work

Little literature is available on the predictive analytics of motor racing. A few publications are available, taking National Association for Stock Car Auto Racing (NASCAR) as a case study, and even though it is different from F1 racing, the algorithm efficiency can be estimated for F1 racing. Various publications use other target metrics and models. Delen and Cogdell *et al.* propose that the classification models work better in predicting the winners than the regression-based models using the NCAA football data [8]. Tulabandhula and Rudin [9] propose multiple machine learning algorithms, including Support Vector Regression and Least Absolute Shrinkage and Selection Operator (LASSO), to predict the tire changing time in NASCAR considering the real-time data. They have described the complexities of race data pre-processing in motorsport prediction, which can also be applied to formula one racing [9]. Graves and Reese *et al.* [10] propose a stochastic model for results forecasting in NASCAR racing. Pfitzner, Barry, and Tracy D. Rishel [11] and Allender [12] propose the impact of multiple features on the race outcome and propose solutions to identify essential features like the pace of the car, the speed of the car in the qualifying session, and starting position. A few papers consider the data only specific to formula one racing. Stoppels [13] suggests a race prediction algorithm using historical data, and Stergiousdis [14] uses famous Machine Learning algorithms to predict the result position of a driver in F1 racing. Aversa and Cabantous *et al.* analyze the cause of Ferrari's failed race strategy using a decision support system for the 2010 season [15]. Another notable work in this domain is by Liu and Fotouhi, where they use a hybrid model with neural networks and a Monte Carlo algorithm to predict the car's performance under a particular energy management technique [16].

This project aims to use real-time data to choose a pitstop strategy that will result in the highest scores earned and to independently predict the ranks of every driver. In this context, the work of Heilmeier and Thomaser *et al.* [7] and T. Tulabandhula and C. Rudin [9] comes closer. The inspiration for pitstop strategy design has been taken from Heilmeier and Thomaser *et al.* [7], and the data pre-processing and feature selection criteria are inspired by the works of Tulabandhula and Rudin [9]. However, none utilize an embedded deep-learning algorithm to evaluate model performance. Certain publications prove the improved accuracy of predictive models using the embedded layers using real-time data [17–19].

C. Embeddings in Neural Networks

The use of embeddings in neural networks is getting famous due to their applications in dimensionality reduction [20]. The embedding maps the higher dimensional feature set into a lower dimensional vector space. Using an embedded layer within the neural network captures the semantics of the input feature set and provides a transformed space as an output where the semantically similar features are placed together [21]. Embeddings are learned as a part of the neural network, which is very helpful, especially when the training dataset is sparse. This embedded layer can be considered a hidden layer that can be combined with any other direct features or more hidden layers, and the last layer of the DNN will be the loss function.

III. METHODOLOGY

The main idea behind the winning strategy prediction is the high correlation between the finishing position and the pitstop strategy. The model utilizes two neural networks with embeddings, thus called Embedded Deep Neural Network (EDNN). Both EDNNs use two different target metrics.

1) The rank position of the driver
2) The optimum lap to make a pitstop to finish the race on top ranks

The designed system uses two EDNNs, one for each target metric, and the final output is shown as a combination of both predictive metrics to propose the winning strategy for the race. The first EDNN uses the data for all the drivers across all laps to predict the driver's rank, and the second EDNN uses the per-lap data of the race to predict the optimum lap for the pitstop.

The first EDNN considers the racer's recent history, which includes the number of races won, the maximum speed achieved by the driver on that track, and the car's speed. For the second target metric, the impact of the tire change on the rank needs to be evaluated to decide the optimum pitstop lap. This second problem is more complex as the choice of tires varies dramatically with the track conditions, weather, type of tires used in the race start, and what kind of tires had the fastest lap in the qualifying round.

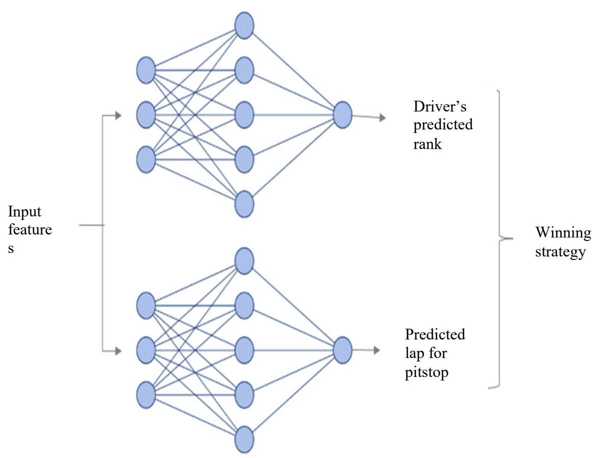

Fig. 3. The schematic overview of proposed EDNN model.

The teams can make one, two, three, or more pitstops depending on the racing conditions. More than three pitstops usually occur in the rain, where the teams switch between wet, intermediate, and hard tires depending on the weather conditions and the race opponent's strategy. This algorithm considers only one, two, and three pitstop strategies, so the model does not learn the behavior of wet weather races. Prediction of lap changes depending upon which pitstop strategy the team is following. If the team chooses one pitstop strategy, the output neuron will only perform one prediction. Hence, the pitstop strategy is added as an input feature to the model, where another feature evaluates the number of pitstops remaining. The total number of pitstops remaining and the race progress significantly impact the predicted lap for a pitstop.

As summarized in Section II, published papers use Machine Learning algorithms and Artificial Neural Networks to predict multiple metrics related to motorsport racing, like pitstop duration. The published literature also suggests that Neural Networks have better performance in terms of accuracy and robustness when compared with other supervised algorithms [11]. Hence, this project utilizes Deep Neural Networks to achieve a better predictive design.

The winning strategy is evaluated in three stages.

1) Predict the rank of the driver in a race using the historical data of that driver and the other participating drivers.
2) The pitstops prediction, i.e., in which lap the pitstop should be made for the winning strategy—the pitstop prediction changes with the number of desired pitstops and the race progress.
3) Combine the predicted rank and pitstop lap to evaluate the winning strategy.

One neural network predicts the rank of the racer using historical data, and the second neural network determines if the pitstop is required. Here, the decision to make a pitstop or not and the choice of the compound tire are treated as independent variables. The choice of two different networks is made because of the different feature sets for both metrics. It also provides the flexibility to tune hyperparameters specific to each target metric. Fig. 3 shows the conceptual diagram of the designed neural network. The amount of data available to every F1 team comes from multiple sensors in their racing car. Most teams use this data and the Monte Carlo algorithm to decide the real-time strategy. However, since this kind of data is not publicly available, it can impact the model's predictive accuracy. The following sections explore the applications of embedded deep neural networks in a predictive framework for accuracy improvement. In this project, the data pre-processing requires detailed domain knowledge; otherwise, the input feature set can be impacted by Simpson's paradox and make an important feature appear less critical [11].

The designed algorithm can be utilized in two ways. Firstly, to perform the predictions before the race, using historical data. These predictions can be made to assess the racing conditions and engage the fans for a better racing experience. Secondly, this model can also be trained for real-time predictions.

A. Training Dataset

The data has been taken from 2015 to 2022, combined from multiple resources. The data for races, including race lap times, qualifying lap times, starting positions, pole positions, and pitstop duration, has been obtained from Ergast API [22]. Except for the tire choice, the complete database is available on GitHub with an open-source license (https://github.com/TUMFTM/f1-timing-database). The dataset has the data for 258 races with a total of 169525 laps.

The number of accidents and failures per driver and season is taken from an online motorsport statistics site [23].

Out of all the data points, in 5213 laps, the driver enters the pitlane, i.e., 3.07% of all the laps. It suggests a data imbalance for both classes, i.e., pitstop or no pitstop. Similarly, tires were changed in only 4855 laps, making it 2.9% of the total laps used in training. It has been taken into consideration while performing the model training and validations.

B. Data Filtering

The raw data contains the data for all the drivers throughout the races. Since F1 racing is a dynamic sport and many events such as rain or crashes happen, which can be treated as outliers and eliminated from the raw data. This training data for this model uses the following filters:

1) The pitstop strategy changes if it rains before or during the race. Not only do more than three pitstops happen, but also, in such a situation, winning the race mainly depends on the driver and their sheer luck. Therefore, the data for wet season races have been excluded as the model should not learn from such unusual cases.
2) According to FIA regulations, race points are only awarded to the top ten positions. So, the data for the top ten drivers possess a winning strategy, and since the drivers below rank ten are not awarded any points, their data has been eliminated in training.
3) Those cases are eliminated where a driver cannot qualify or finish the race due to damage or malfunctioning in the racing car.
4) If any driver breaches the rules set by FIA, he is assigned a penalty to wait for some time at the pitstop. Such cases are also filtered from the raw data, which prevents the model from learning from the longer pitstops.

C. Feature Engineering

The input feature set is passed through data imputations to avoid any problem related to algorithm behavior. The mean values replace the missing values of the numerical features. Also, the data for the starting two laps of the races have been eliminated to avoid edge cases.

A description for the input feature set is provided in a public repository (https://github.com/SitaraWishal/An-Embedded-Deep-Neural-Network-EDNN-Model-to-Predict-the-Winning-Strategy-in-Formula-One-Racing). There are total 36 input features used as raw data. Table I shows the average normalized life of compound tires used in this project as an input feature.

TABLE I: COMPARISON OF AVERAGE TIRE AGE BY SEASONS [4]

Season	Average age of hard compound	Average age of medium compound	Average age of soft compound
<2019	-	33.01%	31.60%
>=2019	38.80%	34.90%	32.00%
Overall	38.80%	34.10%	31.80%

Categorical features are one-hot encoded and split into training, validation, and test sets. Ten folds cross-validation has been applied to avoid overfitting or skewing in the training data. The importance of each feature against both target metrics has been evaluated using Random Forest Regressor. The top 80% of features have been kept for each neural network training using their importance scores. Using the pair-wise correlation of the input features, in the feature pairs for which the correlation is greater than 85%, only one feature from the pair is kept, depending on its importance score. It prevents the EDNN from learning any bias of information in the data.

D. Stint Lengths

The time taken to complete each race is a sum of multiple time factors. Considering the total race time equation given by A. Heilmeier, A. Thomaser, M. Graf, and J. Betz [7]:

$$t_{lap}(l) = t_{base} + t_{tire} + t_{fuel} + t_{car} + t_{driver} + t_{grid} + t_{pit} \quad (1)$$

Where t_{lap} is the time taken to complete one lap, l, t_{base} is the minimum time to complete one lap on a track under ideal circumstances, t_{tire} is the time added as a delay due to tire degradation, t_{fuel} is due to stalling caused by the mass of the fuel, and as the race progresses, more fuel is burned out, and t_{fuel} becomes lower. t_{car} and t_{driver} are subject to driver skills and car's ability. t_{grid} is the time lost due to not being in a pole position, i.e. for a driver starting the race at a pole position, the t_{grid} would be zero. Lastly, by t_{pit} are the times lost in the pit lane.

Optimum stint lengths are taken as the lap conditions where the lap time is minimum for a one-driver-race pair. If we consider (1), all the other time delays remain the same for a particular condition, and the only variable is time delay due to tire degradation. Heilmeier and Thomaser *et al.* [7] suggests that if we consider a linear tire degradation model, then the minimum duration for the driver's stint lengths can be taken as a mixed-integer quadratic optimization problem (MIQP). So, the minimum race time can be calculated as a function of the total number of laps l_{tot} as:

$$min\, t_{race}(l_{tot}) \triangleq min \sum_{l=1}^{l_{tot}} t_{tire}(l) \quad (2)$$

E. Deep-Racing Embedded Deep Neural Network Architecture

A deep learning-based algorithm has been designed to predict the ranks of the drivers and the optimum pitstop lap for any race. The input feature set has the data for all the laps for every participating driver per race. The data per lap is necessary to understand the dynamics of pitstop decisions, but it makes the input feature set sparser. In this environment, a typical neural network does not perform well [24]. However, the literature review does suggest improved predictive accuracy using embeddings in the neural network [17–19]. The feature set is divided into a 90-10 split, where the top 7 features are fed directly into the neural network, whereas the other 20 features are fed to an embedding layer first. Fig. 4. shows the structure of one of the two neural networks. The embedding layer is just another layer of size d, combined with other hidden layers and features. The dropout layer is used to turn off on maximum feature limit. The final output layer is the binary cross-entropy loss, which must be minimized. The output of each embedding layer provides a lower dimensional vector space. The rank prediction is a regression problem, and the pitstop decision is a multi-class classification prediction problem predicting the lap number with maximum probability and using the race progress and the number of remaining pitstops as the leading indicators. Table II shows the hyperparameters for each EDNN.

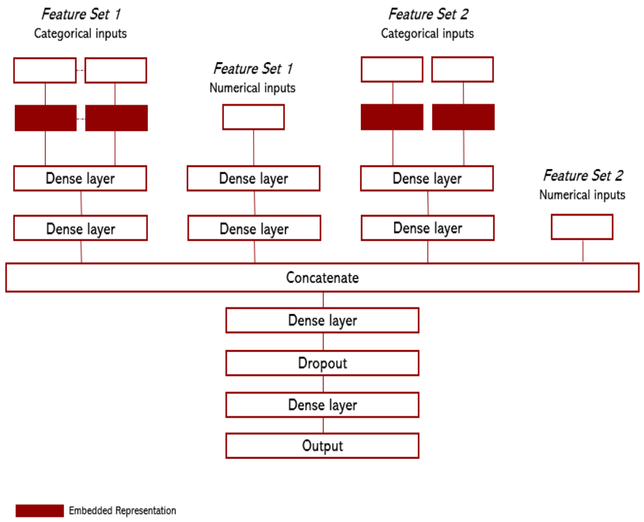

Fig. 4. A conceptual diagram of EDNN model.

F. Model Training

The model is trained using python 3.6 using the Keras and TensorFlow frameworks, and it is tested on Amazon Web Services (AWS) SageMaker with the Linux operating system [25]. The designed model is trained on historical data available. The winning strategy is shown by integrating the output of both neural networks and feeding it to a simple optimization problem. The model shows the predictions of the driver's rank for upcoming laps, using the previous lap's data simultaneously with the pitstop prediction.

IV. Results and Discussions

A. Performance Parameters

In this paper, the performance of the designed model is evaluated using root mean square error (RMSE) and R-squared (R2) error for the driver ranking prediction. As the pitstop predictions have a bias for both cases (pitstop vs. non-pitstop), precision (p) and recall (r) are better indicators of the predictive quality.

A confusion matrix has been shown, and the performance has been evaluated using a harmonic mean of precision (p) and recall (r) have been used, shown as the F1 score in (3).

$$F1\ score = 2.\frac{p.r}{p+r} \tag{3}$$

B. Choice of Hyperparameters for Deep Learning Network

Table II shows the hyperparameters for the rank and pitstop predictions EDNNs. The choice of activation functions has been made using the available literature. The rest of the parameters have been selected as the best case from the hit-and-trial method.

C. Results Analysis

For the first EDNN model, the model's performance has been evaluated against RMSE and R2 scores. The model shows an RMSE of 2.51 on the training dataset and 2.05 on the test data set. R2 score is 0.42 for the training dataset and 0.39 for the test dataset.

TABLE II: Choice of Hyperparameters for EDNN

Hyperparameters for driver's rank prediction	
Hyperparameter	**Value**
Number of hidden layers	15
Number of neurons per layer	150
Activation function for hidden layers	ReLU
Activation function for output layer	Sigmoid
L2 regularization	0.005
Optimizer	Root mean squared propagation
Loss function	Binary cross entropy
Training batch size	2048
Learning rate	0.001
Epochs	150
Hyperparameters for pitstop prediction	
Hyperparameter	Value
Number of hidden layers	10
Number of neurons per layer	120
Activation function for hidden layers	ReLU
Activation function for output layer	Sigmoid
L2 regularization	0.0005
Optimizer	Nadam
Loss function	Binary cross entropy
Training batch size	2048
Learning rate	0.001
Epochs	150

The predicted score is 93% correlated with the actual rank of the drivers. Fig. 5 shows a scatter plot for actual and predicted driver ranks.

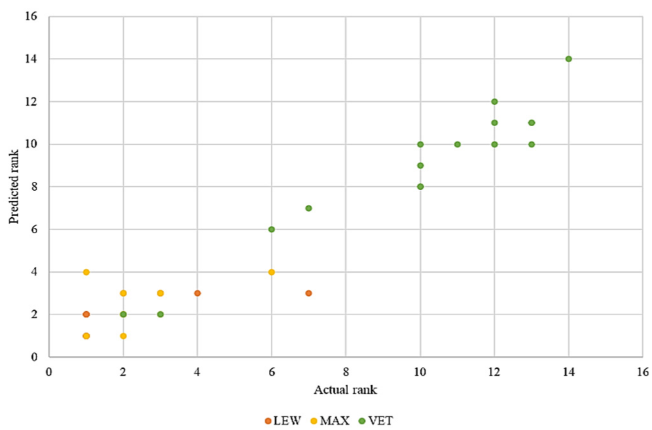

Fig. 5. A scatter plot for actual vs predicted driver rank for 2020 season of formula one racing. Driver abbreviations: LEW—Lewis Hamilton, MAX—Max Verstappen, VET— Sebastian Vettel.

The confusion matrix in Table III summarizes the results for the pitstop prediction metric. There were 1986 false positives, which is lesser than with the hybrid model suggested by Heilmeier and Thomaser *et al.* [7], and 544 pitstops happened without the algorithm detecting them. It can be caused by multiple factors, including car retirement or when the team principals experiment with different strategies on their two available cars.

$$precision = p = \frac{TP}{TP+FP} = \frac{2588}{2588+1986} = 0.56$$

$$recall = r = \frac{TP}{TP+FN} = \frac{2588}{2588+544} = 0.83$$

$$F1\ score = 2.\frac{p.r}{p+r} = 2.\frac{0.56 * 0.83}{0.56 + 0.83} = 0.67$$

Overall, the precision of the Deep-Racing algorithm is 0.56 and the F1 score is 0.67.

TABLE III: CONFUSION MATRIX FOR PITSTOP PREDICTIONS

		Predictions	
		No pitstop	**Pitstop**
True value	**No pitstop**	69765	1986
	Pitstop	544	2588

V. FURTHER RESEARCH

The focus of this paper has been to use Embeddings with Deep Neural Networks for formula one race predictions. The racing cars, drivers, and race regulations vary from season to season, so the deep-racing algorithm cannot provide a perfect response to specific situations, for example, racing under a safety car phase. The predictive accuracy in such conditions can be increased by considering a more exhaustive feature set. For instance, every formula one team has an enormous amount of data coming directly from the sensors deployed on the racing car, which can be very helpful in designing a better tire degradation model. Also, in real-world applications, the racing strategy changes dramatically concerning the nearest opponents of the driver. In such cases, a feature like sector times can be helpful to increase the predictive accuracy of pitstops. Merging the two outputs from the EDNN into a single network can improve the prediction quality as well as the computational efficiency of the deep-racing algorithm.

The designed system can be used to predict the results of the first race of the new season using only the previous season's data for better fan engagement. The racing car's performance can also be improved by retraining the deep-racing model with the data from the racing car sensors and predicting the car's performance parameters like drag force. This framework of predictions can be used to predict the race results in other motorsports like NASCAR and Formula E.

VI. CONCLUSION

In this paper, deep neural network embeddings for formula one racing predictions have been investigated. Overall, the embeddings show improved performance on regression and classification problems compared to the previously published algorithm designs. The model's predictive accuracy is improved by using a subset of features through embeddings and with other features for deep neural network training. For the classification model, 1986 false positives and a total of 544 false negatives happened. The error scores for the regression problem also show significant improvement for the current dataset. The results can be improved by more extensive hyperparameter tuning for the deep neural network and testing different embeddings.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Author Syeda Sitara Wishal Fatima contributed to the architecture of the proposed framework, collected the data, performed the coding, and was responsible for the paper's write-up. Author Jennifer Johrendt contributed ideas to the model development and reviewed and revised the manuscript.

ACKNOWLEDGMENT

We would like to thank Ergast Motor Racing Developer API and Motorsport Stats teams for contributing to the formula one open-source database and providing motor-racing insights.

REFERENCES

[1] F. Staff, *The Most Popular Motorsports in the World*, Frontstretch, October 12, 2022.

[2] Statista Research Department. *Formula One-Statistics & Facts*. Statista. [Online]. Available: https://www.statista.com/topics/3899/motor-sports/#topicHeader__wrapper

[3] B. Tendu, "Ranking the 15 most expensive sports in the world right now," *SportsBrief-Sport News*., Aug. 21, 2022.

[4] A. Heilmeier, M. Graf, J. Betz, and M. Lienkamp, "Application of monte carlo methods to consider probabilistic effects in a race simulation for circuit motorsport. Appl. Sci. 2020, 10, 4229," *Applied Sciences*, vol. 10, no. 17, p. 5745, Aug. 2020, doi: 10.3390/app10175745.

[5] A. Duxbury and J. Holding, "How fast is an F1 car? Top speeds of F1, IndyCar, MotoGP and more," *AutoSport*, Nov. 23, 2022.

[6] WEC FIA World Endurance Championship. Pits- General conditions. [Online]. Available: https://www.fiawec.com/en/pits-general-conditions/116#:~:text=Speed%20in%20the%20pit%20lane,the%20time%20of%20the%20infringement

[7] A. Heilmeier, A. Thomaser, M. Graf, and J. Betz, "Virtual strategy engineer: using artificial neural networks for making race strategy decisions in circuit motorsport," *Applied Sciences*, vol. 10, no. 21, p. 7805, Nov. 2020.

[8] D. Delen, D. Cogdell, and N. Kasap, "A comparative analysis of data mining methods in predicting NCAA bowl outcomes," *International Journal of Forecasting*, vol. 28, no. 2, pp. 543–552, 2012.

[9] T. Tulabandhula and C. Rudin, "Tire changes, fresh air, and yellow flags: challenges in predictive analytics for professional racing," *Big Data*, vol. 2, no. 2, pp. 97–112, Jun. 2014, doi: 10.1089/big.2014.0018.

[10] T. Graves, C. S. Reese, and M. Fitzgerald, "Hierarchical models for permutations: Analysis of auto racing results," *Journal of the American Statistical Association*, vol. 98, no. 462, pp. 282–291, 2003.

[11] P. Barry and T. D. Rishel, "Do Reliable Predictors Exist for the Outcomes of NASCAR Races?" *The Sport Journal*, vol.8, no. 2, 2005.

[12] M. Allender, "Predicting the outcome of NASCAR races: The role of driver experience," *Journal of Business & Economics Research (JBER)*, vol. 6, no. 3, Mar. 2008, doi: 10.19030/jber.v6i3.2403.

[13] E. Stoppels, "Predicting race results using artificial neural networks," Master's thesis, University of Twente, 2017.

[14] A. Stergioudis, "F1 predictor – Machine-learning based F1 race prediction engine," *F1 Predictor*, 2022.

[15] P. Aversa, L. Cabantous, and S. Haefliger, "When decision support systems fail: Insights for strategic information systems from Formula 1," *The Journal of Strategic Information Systems*, vol. 27, no. 3, pp. 221-236, 2018.

[16] X. Liu and A. Fotouhi, "Formula-E race strategy development using artificial neural networks and Monte Carlo tree search," *Neural Computing and Applications*, vol. 32, no. 18, pp. 15191–15207, 2020.

[17] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, p. 106181, May 2020, doi: 10.1016/j.asoc.2020.106181.

[18] S. Tharmakulasingam, B. T. Phung, and E. Ambikairajah, "Enabling deep learning on embedded systems for IoT sensor data analytics: Opportunities and challenges," in *Proc. IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, Dec. 2018, pp. 1–5.

[19] R. Mu and X. Zeng, "A review of deep learning research," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 4, Apr. 2019, doi: 10.3837/tiis.2019.04.001.

[20] K Fernandes, D Chicco, and J. S. Cardoso, "Supervised deep learning embeddings for the prediction of cervical cancer diagnosis," *Peer J Computer Science*, vol. 4, May 2018.
[21] Google Machine Learning Education. (2022). Embeddings: Obtaining Embeddings. [Online]. Available: https://developers.google.com/machine-learning/crash-course/embeddings/video-lecture
[22] C. Newell. (2020). Ergast Motor Racing Developer API. Ergast Developer API. [Online]. Available: http://ergast.com/mrd
[23] Motorsport stats. *FIA Championships Results & Statistics.* [Online]. Available: https://fiaresultsandstatistics.motorsportstats.com/results
[24] J. Bissmark and O. Wärnling, 'The sparse data problem within classification algorithms: The effect of sparse data on the Naïve Bayes algorithm," Dissertation, 2017.
[25] keras-team. (2019). Keras: Deep Learning for humans. GitHub. [Online]. Available: https://github.com/keras-team/keras