Win prediction in fighting games from temporal input sequences: a proof of concept with Street Fighter 6

Marcel Ricardo Goto
http://lattes.cnpq.br/6225327196769122

ABSTRACT: Outcome prediction in esports has been widely explored in strategy games, MOBAs, and first-person shooters, usually based on match statistics or game-state variables. Fighting games, however, remain less explored, although they constitute a particularly suitable domain for performance analysis based on player behavior. The present study introduces a proof of concept for predicting round outcomes in a popular fighting game, Street Fighter 6, from temporal sequences of commands (inputs) executed by the player. The aim was to assess whether input patterns contain enough information to estimate the probability of victory without relying on images, audio, health bars, or other contextual game variables. A total of 128 rounds from matches played by the author were analyzed, each represented as an ordered sequence of discrete events and classified using a Gated Recurrent Unit (GRU) recurrent neural network. The model correctly classified 75.0% of rounds, with 83.7% sensitivity for wins and 69.6% specificity for losses. Won rounds showed a higher mean predicted probability of victory than lost rounds (72.0% versus 33.9%), with a statistically significant difference. In addition, the time during which the estimated probability remained above 50% was associated with the round outcome. The results indicate that command sequences in fighting games provide relevant signal for outcome prediction and suggest future applications in automated analysis and performance feedback.

Keywords: digital games. Esports. Recurrent neural networks. Outcome prediction. Machine learning.

1. INTRODUCTION

Esports provide an especially suitable context for the computational study of competitive performance, as they combine structured contests, high event frequency, precise digital recording, and substantial diversity of strategies and motor skills. The availability of high-resolution data has motivated research on outcome prediction, skill modeling, and training-support systems (FORMOSA et al., 2022; RAETZE; STAEDTER; HÜLLMANN, 2025).

A substantial part of this literature focuses on predicting match winners. In real-time strategy games, Álvarez-Caballero et al. (2017) showed that outcomes can be predicted before the end of the match from features of the game state. In MOBAs, studies have explored recurrent networks, pre-game variables, player statistics, player-character experience, feature engineering, and confidence calibration (SILVA; PAPPA; CHAIMOWICZ, 2018; JUNIOR; CAMPELO, 2023; DO et al., 2021; HITAR-GARCÍA; MORÁN-FERNÁNDEZ; BOLÓN-CANEDO, 2023; CHUNG, 2020). In Dota 2, studies using decision trees, team-fight models, and live prediction reinforce the predictive value of intermediate events and combat interactions (YANGIBAEV; MATTIEV; MOKWENA, 2025; KE et al., 2022; HODGE et al., 2021).

Other competitive genres have also been analyzed through similar methods. In shooters, studies have investigated outcome prediction, explainable models, granular action data, and limitations imposed by small datasets or restricted API access (ŠVEC, 2022; BROMS; NORDANSJO, 2024; GARCÍA-MÉNDEZ; ARRIBA-PÉREZ, 2025; XENOPOULOS; SILVA, 2022; NIELAND, 2025). Taken together, these works broaden the scope of esports prediction beyond MOBAs and RTS games, but they remain largely based on match statistics, game events, or state variables.

Beyond high-level game statistics, there is a line of research focused on modeling skill and performance from low-level behavioral signals. Buckley, Chen and Knowles (2013) showed that keyboard and mouse inputs can classify the player's experience level in an FPS; Smith and Nayar (2016) applied topic models to streams of controller commands; Guglielmo et al. (2023) demonstrated that commands issued in the first seconds of Tetris predict later performance; and Avontuur, Spronck and Van Zaanen (2013) associated skill in StarCraft II with indicators derived from actions recorded by the game, such as execution speed, command volume, and early patterns of unit control.

This behavioral approach is particularly relevant to fighting games, in which inputs are discrete and direct-action events. Unlike RTS, FPS, and MOBAs, in which many commands are mediated by cursor navigation, aiming, selection, positioning, or unit management, in fighting games each direction, pressed button, and input duration can directly influence the executed action. A timing inaccuracy of a single frame, approximately 16.7 milliseconds in games running at 60 frames per second, often determines whether a command will be recognized or canceled, punished, or converted into advantage. This direct relationship between input, timing, and mechanical consequence makes fighting games especially suitable for analysis based on input sequences.

Nevertheless, predictive studies on fighting games have still relied mainly on variables that describe the game state or accumulated consequences of both players' actions. Chulajata et al. (2024), for example, approached prediction in Super Street Fighter II Turbo using the time series of health bars, that is, a variable that already expresses accumulated consequences of actions. Minami et al. (2024), in turn, indicate that these games can also be analyzed by predictive models using neurophysiological signals. Such studies show the feasibility of prediction in this genre, but they leave open the possibility of investigating the player's observable behavior as the primary source of information.

The present work is situated in this gap. By analyzing input sequences, the study dialogues with the skill-modeling approach discussed by Avontuur, Spronck and Van Zaanen (2013), but shifts this logic to a genre in which the precision of each input has greater impact on performance. In the literature consulted and in searches conducted in online academic indexes, no previous studies were found that exclusively use player input sequences to predict victory or defeat in rounds of fighting games.

To explore this possibility, a proof of concept was conducted for outcome prediction in rounds of Street Fighter 6, a competitive one-on-one fighting game, from temporal sequences of commands from a single player. The contribution of the study is to adapt to the context of fighting games an agenda already consolidated in RTS, MOBA, and FPS research. In doing so, it seeks to offer an initial basis for automated analysis and performance feedback tools, investigating whether command sequences contain enough information to predict victory or defeat in a round.

2. METHOD

The data consisted of the author's inputs, using a single character ("Ryu"), extracted from replay videos of online matches. The unit of analysis was the individual round, labeled binarily as victory or defeat. Each round was stored as a .csv file and represented as a temporal sequence of discrete events, defined by direction, pressed buttons, and event duration in frames. No additional contextual game data were used, such as health bars, character position, opponent actions, or the visual state of the match. The training set consisted of 2,200 rounds, with sequences of mean length 250.8 events, standard deviation 88.0, minimum 47, and maximum 696 events. Directions were encoded as discrete categories, including NEUTRAL, LEFT, RIGHT, UP, DOWN, and diagonals UP-LEFT, UP-RIGHT, DOWN-LEFT, and DOWN-RIGHT. Buttons included LP, MP, HP, LK, MK, and HK (light, medium, and heavy punches and kicks). Input durations were capped at 99 frames and normalized by division by this value before entering the model.

Before training, an independent sample of 128 rounds was separated and not used in any stage of model adjustment. This sample was used only for subsequent evaluation, in order to test the classifier's performance on rounds to which it had not previously been exposed. The remaining set was used for training and internal validation, with a validation split of 10%.

A recurrent neural classifier based on a bidirectional GRU was trained. The model used two recurrent layers, hidden size of 128 units, dropout of 0.2, and final aggregation using the lastk_mean mode. Training was carried out for 20 epochs, with batch size of 32, learning rate of 0.001, weight decay of 0.0001, and random seed 420, running on GPU via CUDA.

For each round, the model generated an estimated probability of victory for the player, referred to as p(win). For binary classification, a threshold of 0.50 was adopted as reference: values above this point were classified as victory, and values equal to or below it were classified as defeat. An adjusted threshold to maximize the F1-score in internal validation was also computed, resulting in a cutoff point of 0.4955.

The main evaluation metrics were accuracy, precision, recall, F1-score, AUC, and confusion matrix. In addition to final classification, probabilistic and temporal indicators derived from the p(win) curve throughout the round were computed, as shown in Table 1.

Table 1 - Indicators derived from the p(win) curve

Indicator	Description
final p(win)	Estimated probability of victory at the end of the round, after 100% of events.
p(win) @50%	Estimated probability after 50% of events.
p(win) @95%	Estimated probability after 95% of events.
time with p(win) > 50%	Proportion of the round during which the estimated probability remained above 0.50.
area	Accumulated area of the p(win) curve, used as an aggregate measure of probabilistic tendency throughout the round.
Mean and median p(win)	Descriptive statistics compared between won and lost rounds.

The inclusion of these indicators aimed to evaluate not only whether the model correctly classified the round outcome, but also whether its probabilistic trajectory distinguished won and lost rounds over time. In particular, "time with p(win) > 50%" was used as a measure of the persistence of states classified by the model as favorable to victory.

Comparisons between won and lost rounds were performed using two-tailed between-group difference tests applied to the metrics derived from p(win).

As this was a computational study based on the author's inputs, no third-party data were collected or analyzed. The scope was limited to modeling command sequences and their respective outcomes.

3. RESULTS AND DISCUSSION

The independent evaluation was performed on 128 rounds previously separated from the dataset before model training. Of this sample, 49 rounds corresponded to wins and 79 to losses. Using final p(win) > 50% as the classification criterion, the model correctly classified 96 of the 128 rounds, resulting in an overall accuracy of 75.0%.

Table 2 - Confusion matrix in the independent sample

	Prediction: loss	Prediction: win	Total
Actual: loss	55	24	79
Actual: win	8	41	49
Total	63	65	128

Sensitivity for identifying wins was 83.7%, indicating that the model correctly recognized 41 of the 49 won rounds. Specificity for losses was 69.6%, with 55 of the 79 losses correctly classified. Precision for the "win" class was 63.1%, whereas the negative predictive value was 87.3%, indicating that rounds classified as losses by the model had a high probability of corresponding to actual losses.

Table 3 - Classification metrics in the independent sample

Metric	Value
Accuracy	75.0%
Sensitivity for wins	83.7%
Specificity for losses	69.6%
Precision for wins	63.1%
Negative predictive value	87.3%
F1-score for wins	71.9%

The estimated probabilities generated by the model also clearly differentiated won and lost rounds. The mean final p(win) in wins was 72.0%, whereas in losses it was 33.9%. The difference was even more marked in the medians: 80.2% in wins and 29.6% in losses. The comparison between groups indicated a statistically significant difference for final p(win) ($p = 8.37 \times 10^{-14}$).

Table 4 - Estimated probability of victory by actual outcome

Indicator	Wins	Losses
Mean final p(win)	72.0%	33.9%
Median final p(win)	80.2%	29.6%
Standard deviation	22.1	28.6

Temporal indicators derived from the p(win) trajectory throughout the round also distinguished wins and losses. The estimated probability halfway through the round (p(win) @50%) differed between groups ($p = 4.98 \times 10^{-4}$), as did the estimated probability at 95% of the round (p(win) @95%; $p = 1.98 \times 10^{-5}$). The accumulated area of the probability curve was also distinct between wins and losses ($p = 5.41 \times 10^{-10}$).

The most interpretable temporal indicator was the time during which p(win) remained above 50%. This measure showed a strong association with round outcome ($p = 6.38 \times 10^{-10}$). Rounds in which the estimated probability remained above 50% for less than 10% of the time were won in only 8.6%

of cases. In contrast, when p(win) remained above 50% for 40% to 49% of the round, the proportion of wins was 75.0%.

Table 5 - Proportion of wins by time with p(win) > 50%

Time of round with p(win) > 50%	n	Wins	Losses
0–9%	35	8.6%	91.4%
10–19%	29	24.1%	75.9%
20–29%	9	33.3%	66.7%
30–39%	17	47.1%	52.9%
40–49%	20	75.0%	25.0%
50–59%	10	70.0%	30.0%
60–69%	6	66.7%	33.3%
70–79%	1	100.0%	0.0%
80–89%	1	100.0%	0.0%

Accuracy also varied according to round duration. In rounds up to 1,499 frames, the model reached 92.3% accuracy; between 1,500 and 1,999 frames, 88.5%; between 2,000 and 2,499 frames, 75.0%; and between 2,500 and 2,999 frames, 76.9%. In longer rounds, between 3,500 and 4,499 frames, accuracy fell to 40.0%, although these ranges contained a small number of cases. Thus, the trend suggests better model performance in short and medium-length rounds, but results for long rounds should be interpreted with caution.

Taken together, the results indicate that command sequences contain relevant predictive signal about round outcomes. The model not only classified final outcomes above chance, but also produced a coherent probabilistic trajectory: won rounds showed higher final p(win) values, greater persistence above the 50% threshold, and greater temporal separation from lost rounds. In future work, this trajectory may be related to the literature on psychological momentum in esports: White and Romano (2020) used recurrent networks to model momentum and tilt in League of Legends, suggesting that temporal variations in performance may carry information about transient player states. In the present study, performance drops in long rounds should be interpreted only as an exploratory hypothesis, possibly related to loss of momentum, fatigue, or greater situational complexity, and not as direct evidence of psychological state.

4. FINAL CONSIDERATIONS

This study presented a proof of concept for predicting outcomes in Street Fighter 6 rounds from temporal sequences of player commands. The results indicate that low-level inputs, consisting of direction, pressed buttons, and command duration, contain enough information to estimate round outcome above chance. In the independent sample, previously separated from training, the model achieved 75.0% accuracy, with 83.7% sensitivity for wins and 69.6% specificity for losses.

In addition to final classification, the probabilistic trajectory produced by the model showed a consistent relationship with the actual outcome. Won rounds had substantially higher mean and median p(win) values than lost rounds, and the time during which the estimated probability remained above 50% was also strongly associated with victory. This result suggests that the model not only distinguishes the final outcome, but also captures temporal command patterns related to the development of the round.

The main contribution of this work is to demonstrate the feasibility of applying recurrent neural networks to the analysis of fighting games using exclusively input sequences, without relying on images, audio, health bars, physiological information, or complex internal game statistics. Whereas approaches based on visual state or variables such as the health bar observe consequences already accumulated during the interaction, the present study directly investigates the commands executed by the player. In doing so, it brings research on outcome prediction in esports closer to an execution-centered approach, especially relevant for genres in which temporal precision, command sequencing, and motor response are central components of performance.

As a limitation, the dataset analyzed here remains restricted in both size and diversity of players and characters. In addition, victory and defeat are objective but incomplete labels: a round may be decided by contextual factors not directly captured in the commands, such as opponent decisions, situational reading, matchup, or specific variations in the game state. Therefore, the results should be interpreted as initial evidence of feasibility, not as a definitive measure of skill or overall performance.

Future work may expand the dataset, test generalization across players and characters, compare recurrent architectures with attention-based models or Transformers, and investigate explainability methods capable of indicating which command patterns contribute most to the estimated probability of victory. This step is particularly important because recurrent models may function as black boxes; in line with García-Méndez and Arriba-Pérez (2025), XAI techniques such as SHAP or temporal attribution methods could indicate which command sequences precede abrupt increases or decreases in p(win). It will also be relevant to evaluate model robustness in comebacks and unexpected victories, a problem highlighted by Minami et al. (2024), as well as to explore applications in real-time support tools. In this regard, the literature on live companion tools, such as Wang et al. (2024), suggests the possibility of integrating models of this type into performance feedback systems, capable of indicating to the player when their current command pattern diverges from sequences associated with victory.

Conflict of interest: The author declares no conflict of interest.

Funding: This research received no funding.

REFERENCES

ÁLVAREZ-CABALLERO, Antonio; MERELO, J. J.; GARCÍA SÁNCHEZ, Pablo; FERNÁNDEZ-ARES, Antonio. Early Prediction of the Winner in StarCraft Matches. In: Proceedings of the 9th International Joint Conference on Computational Intelligence - Volume 1: IJCCI. [S. l.]: SCITEPRESS, 2017. p. 401-406. DOI: 10.5220/0006587304010406.

AVONTUUR, Tetske; SPRONCK, Pieter; VAN ZAANEN, Menno. Player skill modeling in Starcraft II. In: Proceedings of the Ninth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE'13). [S. l.]: AAAI Press, 2013. p. 2-8.

BROMS, E.; NORDANSJO, W. Predicting Counter-Strike Matches Using Machine Learning Models. 2024. Dissertation – Lund University School of Economics and Management, Lund, 2024. Available at: https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9145457&fileOId=9145459.

BUCKLEY, David; CHEN, Ke; KNOWLES, Joshua. Predicting skill from gameplay input to a first-person shooter. In: 2013 IEEE Conference on Computational Intelligence in Games (CIG). Niagara Falls: IEEE, 2013. p. 1-8. DOI: 10.1109/CIG.2013.6633655.

CHULAJATA, K.; WU, S.; SCALZO, F.; CHA, E. S. Predicting Outcomes in Video Games with Long Short Term Memory Networks. arXiv, 2024. Available at: https://arxiv.org/abs/2402.15923.

CHUNG, K.-S. A Confidence-Calibrated MOBA Game Winner Predictor. In: 2020 IEEE Conference on Games (CoG). [S. l.]: IEEE, 2020. DOI: 10.1109/COG47356.2020.9231878.

DO, Tiffany D.; WANG, Seong Ioi; YU, Dylan S.; MCMILLIAN, Matthew G.; MCMAHAN, Ryan P. Using Machine Learning to Predict Game Outcomes Based on Player-Champion Experience in League of Legends. In: The 16th International Conference on the Foundations of Digital Games (FDG 2021). New York: ACM, 2021. DOI: 10.1145/3472538.3472579.

FORMOSA, Jessica et al. Definitions of Esports: A Systematic Review and Thematic Analysis. Proceedings of the ACM on Human-Computer Interaction, New York, v. 6, CHI PLAY, art. 227, p. 1-45, 2022. DOI: 10.1145/3549490.

GARCÍA-MÉNDEZ, Silvia; ARRIBA-PÉREZ, Francisco de. Explainable e-sports win prediction through Machine Learning classification in streaming. Entertainment Computing, v. 55, art. 101027, 2025. DOI: 10.1016/j.entcom.2025.101027.

GUGLIELMO, Gianluca; KLINCEWICZ, Michal; HUIS IN 'T VELD, Elisabeth; SPRONCK, Pieter. Predicting Tetris Performance Using Early Keystrokes. In: Proceedings of the 18th International Conference on the Foundations of Digital Games (FDG '23). New York: ACM, 2023. art. 46, p. 1-4. DOI: 10.1145/3582437.3587184.

HITAR-GARCÍA, J. A.; MORÁN-FERNÁNDEZ, L.; BOLÓN-CANEDO, V. Machine Learning Methods for Predicting League of Legends Game Outcome. IEEE Transactions on Games, v. 15, n. 2, p. 171-181, 2023. DOI: 10.1109/TG.2022.3153086.

HODGE, V. J.; DEVLIN, S.; SEPHTON, N.; BLOCK, F.; COWLING, P. I.; DRACHEN, A. Win Prediction in Multiplayer Esports: Live Professional Match Prediction. IEEE Transactions on Games, v. 13, n. 4, p. 368-379, 2021. DOI: 10.1109/TG.2019.2948469.

JUNIOR, Jailson; CAMPELO, Cláudio. League of Legends: Real-Time Result Prediction. In: Anais do XVI Congresso Brasileiro de Inteligência Computacional. [S. l.]: SBIC, 2023. p. 1-8. DOI: 10.21528/CBIC2023-161.

KE, C. H. et al. DOTA 2 match prediction through deep learning team fight models. In: 2022 IEEE Conference on Games (CoG). Beijing: IEEE, 2022. p. 96-103. DOI: 10.1109/CoG51982.2022.9893647.

MINAMI, Sorato; KOYAMA, Haruki; WATANABE, Ken; SAIJO, Naoki; KASHINO, Makio. Prediction of esports competition outcomes using EEG data from expert players. Computers in Human Behavior, v. 160, art. 108351, 2024. DOI: 10.1016/j.chb.2024.108351.

NIELAND, Jesse. The Possibilities of Applying Machine Learning to a Limited Esports Dataset. 2025. Undergraduate thesis (Bachelor in Data Science and Artificial Intelligence) – LIACS, Leiden University, Leiden, 2025.

RAETZE, Sebastian; STAEDTER, Christian; HÜLLMANN, Joschka Andreas. Taking aim at research on esports teams: a systematic literature review and cross-disciplinary future agenda. Team Performance Management: An International Journal, v. 31, n. 5/6, p. 364-389, 2025. DOI: 10.1108/TPM-09-2024-0109.

SILVA, Antonio Luis Cardoso; PAPPA, Gisele Lobo; CHAIMOWICZ, Luiz. Continuous Outcome Prediction of League of Legends Competitive Matches Using Recurrent Neural Networks. In: SBC – Proceedings of SBGames 2018: Computing Track – Short Papers. Foz do Iguaçu: SBC, 2018. p. 639-642. ISSN 2179-2259.

SMITH, Brian A.; NAYAR, Shree K. Mining Controller Inputs to Understand Gameplay. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16). New York: ACM, 2016. p. 157-168. DOI: 10.1145/2984511.2984543.

ŠVEC, O. Predicting Counter-Strike Game Outcomes with Machine Learning. 2022. Undergraduate thesis – Czech Technical University in Prague, Prague, 2022.

WANG, Letian; DORMANN, Claire; WALLNER, Günter. A Feature Comparison Study of Live Companion Tools for Esports Games. In: INTERNATIONAL CONFERENCE ON THE

FOUNDATIONS OF DIGITAL GAMES (FDG), 19., 2024, Worcester. Proceedings of the 19th International Conference on the Foundations of Digital Games. New York: ACM, 2024. 11 p. DOI: 10.1145/3649921.3650004.

WHITE, Alfonso; ROMANO, Daniela M. Scalable Psychological Momentum Forecasting in Esports. In: WORKSHOP SUM ’20: STATE-BASED USER MODELLING, THE THIRTEENTH ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING (WSDM ’20), 2020, Houston. Workshop SUM ’20: State-based User Modelling, The Thirteenth ACM International Conference on Web Search and Data Mining. New York: ACM, 2020. 8 p.

XENOPOULOS, Peter; SILVA, Cláudio T. ESTA: An Esports Trajectory and Action Dataset. arXiv, 2022. Available at: https://arxiv.org/abs/2209.09861.

YANGIBAEV, S.; MATTIEV, J.; MOKWENA, S. DotA 2 Match Outcome Prediction System Using Decision Tree Ensemble Algorithms. Big Data and Cognitive Computing, v. 9, n. 12, art. 302, 2025. DOI: 10.3390/bdcc9120302.