Research Article

Engineering: Open Access

ISSN: 2993-8643

Judgion: A Frist Approach to Applying Machine Learning Techniques to Judge Mixed Martial Arts Bouts

Iván Ontiveros*

Independent Researcher Alicante, Spain

***Corresponding Author**
Iván Ontiveros, Independent Researcher Alicante, Spain.

Submitted: 2025, Oct 13; **Accepted:** 2025, Nov 14; **Published:** 2025, Nov 20

Citation: Ontiveros, I. (2025). Judgion: A Frist Approach to Applying Machine Learning Techniques to Judge Mixed Martial Arts Bouts. *Eng OA, 3*(11), 01-07.

Abstract

*In this work, I present **Judgion**, a family of six lightweight neural models for round-level MMA scoring from round-by-round statistics. The models are trained on a 501-round corpus that includes real UFC rounds and a set of theoretical rounds crafted to encode key judging criteria that real rounds may not be enough to capture. Judgion's goal is not to predict how judges would score a fight, but to give their own scores following the judging criteria, learning the rules and applying them to real rounds. Judgion was evaluated live on UFC 317–319. Most model scorecards were defensible under the criteria; 7/52 rounds (≈ 13.5%) were flagged as misjudged on qualitative review. At the model–round level, 23/312 decisions (≈ 7.4%) were flagged. Given the complexity and subjectivity of judging and the coarseness of publicly available data, these results are promising and highlight a clear path forward: with richer inputs and continued refinement, automated round scoring can become substantially more reliable.*

Keywords: Mixed Martial Arts, Judgion, Machine Learning,Neural Networks

1. Introduction

Judging is an integral part of most combat sports. Even in disciplines where a fighter can finish their opponent, such as boxing or Mixed Martial Arts (MMA), a very relevant percentage of fights reach their round limit without a finish, and the official result is decided by the judges. Therefore, proper fight scoring becomes crucial, and these decisions frequently spark discussions and debate due to the subjective nature of judging. In MMA, judging has historically been contentious due to the sport's relative youth and its multi-phase, heterogeneous exchanges (striking, wrestling, grappling and the phases that occur when transitioning between them) that judges must interpret in real time. In the UFC specifically, between 2020 and 2024 there were 2,600 bouts, of which 1,298 went the distance, meaning nearly half of fights were decided by the judges [1].

Artificial Intelligence (AI) presents itself as a solution which has the potential to eliminate human subjectivity and bias, promising a more objective way to judge bouts and achieving even better results than human judges. This work serves as a first approach to applying AI to judge MMA fights, using Machine Learning (ML) techniques to train models that are able to 'learn' the intangibles of fighting and properly judge rounds objectively. The goal of this work is not to replace human judges or to achieve better results than humans would, but to introduce an alternative that could be very well used as a tool in the near future.

2. State of the Art

The majority of MMA automation research projects use the UFC official statistics website, as it is the bigger repository available, with the most complete information and the longest continuous record of statistics. Most machine learning approaches have used the information stored in the website, even for different purposes from those of this work. For example, created a dataset using all the statistics available for individual fighters to create models that are able to predict fight outcomes, after processing the data extracted from the website [2]. Rather than predicting fight outcomes, this work focuses on judging real rounds that have already occurred.

Related research exists; the most notable example is ***JudgeAI***, which trains random forests with time-series cross-validation on UFC rounds from 2011–2020, using the official result as the ground-truth for the model [3]. JudgeAI reports about 80% round-score accuracy (and 83% round-winner accuracy) when predicting the judges' scoring (both correct winner and correct score). Because of the nature of the dataset and the training method, JudgeAI is not an effort to create a model that learns how rounds should be scored, but rather an effort to predict how would human judges score those rounds. Therefore, its results should be interpreted as mathematical probabilites about which score is more likely to happen in a given round.

Academic research shows that humans are biased towards certain statistics over others despite supposedly scor- ing following a clear and objective criteria. Feldman, 2020 used logistic regressions on official bout statistics (UFC/Strikeforce/WEC, 2000–2015) to model the probability of a fighter winning a round relying purely on the statistics, and found that, particularly in close rounds, judges tend to overvalue striking numbers over grappling success, even neglecting metrics such as submission attempts, despite the judging criteria giving equal value to both effective striking and grappling. This suggests that treating official round results as ground truth may encode human biases and can be insufficient for training models intended to score rounds strictly by the criteria [4].

Therefore, training models that actually *learn* how to score rounds requires:

- **Better datasets:** Even though the UFC website is the most complete repository, it tracks a limited set of statistics in a simplistic manner. For example, both a jab and a high kick are tracked as a significant strike to the head, which could cause a bottleneck in the model performance.
- **Robust labelling:** Almost all current approaches use the official result as the ground-truth, which is justifiable since it's the closest to an objective way of labelling. Nevertheless, it's still using the result chose by three human judges who must score each round in real time - a process prone to error even for qualified judges. There are plenty of documented widely disputed decisions; thus, treating the official result as ground truth is a pragmatic but ultimately naive strategy.
- **Rule-informed models:** Training a model over real rounds can be enough for it to perform well, but since judging in MMA is done under unified rules, it would be beneficial if the model gets information about the criteria it's supposed to use. This could be especially relevant to resolve close rounds where even a human would struggle to make a decision.

3. Method

Judgion is composed by a collection of neural networks that have been trained using a personalised dataset based on the information stored in the official UFC website. Each model has been trained with the same data, using different architectures and approaches. The goal for these judges is to be able to properly score MMA rounds, mimicking the way real judges do it.

3.1. Training Dataset

The training dataset used consists of 501 JSON dictionaries, each one storing the statistics for a MMA round. This dataset can be divided in two groups:

- **Real Rounds:** 469 have been collected from the UFC website, translating the information to a JSON file.
- **Theoretical Rounds:** 32 of the rounds have been manually created as fictional rounds, with the purpose of giving information to the models about the judging criteria.

Each JSON file in the dataset is structured as follows:

- Red fighter stats
- Blue fighter stats
- Label

3.1.1. Fighter Stats

Both fighters involved in the round have their own *statistics dictionary*, following the same structure. The red corner fighter is always associated with the first dictionary, and the blue corner fighter is associated with the second. For real rounds, the majority of the content in these dictionaries are extracted from the UFC website via web scraping using Python. A total of 22 statistics are extracted this way:

- Knockdowns
- Significant strikes. This is a dictionary in itself, which includes 14 statistics related to significant strikes.
- Total significant strikes.
- Significant strikes to the head, body and legs.
- Significant strikes at distance, during clinch and at the ground.

For each statistic, two different values are tracked: Both attempted and landed strikes.

- Strikes. The total number of strikes is stored as a dictionary which includes two values: Both attempted and landed strikes.
- Takedowns. This statistic is tracked exactly as the previous one, tracking both attempted and landed takedowns.
- Submission attempts
- Reversals
- Control time, in seconds.

As it's been stated, this information is quite limited. Model performance is limited by how appropiate and broad the data is, so extending the dataset is one of the best ways to gain accuracy in scoring. As a first effort, I've included an extra stat in the dataset, **cuts**. A cut is an open injury caused by an impact, either intentional or not, which causes split skin and bleeding, usually in the face or head. The requirements used for considering what a cut is are simple: A visible break of the epidermal layer, and external bleeding caused by the open wound. Both requirements are needed for a cut to be considered; other injuries that cause bleeding, such as nosebleeds, have not been considered as cuts.

3.1.2. Labeling

Correct labeling is the key to achieve good model performance. In this study, official result has played a big part in the label decision, but it has not been the only factor taken into account. Expert and analysts opinions, and the general public thoughts and discussions, have also been carefully revised using different websites, social media and websites, such as:

1. ***MMADecisions.com*** is a website that tracks UFC decisions, with judges and MMA media scorecards [5]. Journalists featured on this website as MMA media are usually well known for their substantial coverage experience and dedication to the sport. In spite of that, they are not qualified judges, so they are not necessarily experts in judging and they may not understand or apply the judging criteria, so their opinion should not be considered as ground truth either, but their scores have been used as auxiliary evidence: An agreement between judges and media usually suggests high confidence in the result; big disagreements usually meant the round needed further review. People who enter this website can also submit their scorecards, so the opinion of the general public could be obtained out of this website, but the accuracy of these scorecards is very susceptible to *self-selection bias*: when people disagree with a scorecard, they are more prone to submit their disenting score, whereas people who agree are less prone to bother, causing imbalance and less accuracy in these scorecards, so public scorecards here were not used as a primary source for labeling.
2. ***Verdict*:** Verdict is a mobile platform that collects live user-submitted round scores, being very useful to understand the general fan perception of a round [6]. Live round scores have been considered, but the scorecards on Verdict are susceptible to *non-expert input and popularity bias* (fighters with larger fan bases tend to receive more favorable scores). Since everyone can submit a scorecard at any time, people who may not have watched the round or may not understand the criteria can submit scorecards who have the same weight as more dedicated fans who try to objectively score fights. Therefore, Verdict has been used as a secondary source, taken into account in specific contested rounds.

- **Social media:** Technical analyses by experts were consulted when available, particularly for close rounds. Highly respected members of the community, including former MMA fighters and champions, coaches or analysts upload content where rounds are analysed and dissected. This type of content have been taken into account during labeling, especially in close fights, but as detailed breakdowns are usually only produced for high-profile bouts, its influence on the overall dataset was limited.

Finally, the author's personal opinion has inevitably played a part in the labeling. It's important to note that the author does not necessarily agree with all labels in the dataset: Personal agreement with the result has not been taken into account when creating the dataset. Big effort was made to minimize subjectivity, but some residual bias is unavoidable. This work is focused solely on getting the best model performance possible.

3.1.3. Theoretical Rounds

Model performance can be improved if information about the criteria is given to it during the training. To achieve this, theoretical rounds have been added to the dataset. Each added round is associated with a certain part of the official judging criteria Association of Boxing Commissions and Combative Sports, 2019 [7]. The information given to the models is the following:

- When neither fighter produces effective offense, but there is a significant difference in aggressiveness (output numbers), the offensive fighter should be granted the round over the defensive one. Four rounds are added for this criterion.
- Effective offense always outweighs aggressiveness. If a fighter has a significant advantage in volume, but isn't effective, while the opponents lands more (even if the difference is small), the round should always be granted to the effective fighter. Six rounds are associated with this criterion.
- Focusing on striking, we can deduce that significant strikes have a higher weight in judging than strikes. Per the rules: *Legal blows that have immediate or cumulative impact with the potential to contribute towards the end of the match with the IMMEDIATE weighing in more heavily than the cumulative impact*. Six rounds have been added for this criterion.
- Focusing on grappling, the rules say: *Successful execution of takedowns, submission attempts, reversals and the achievement of advantageous positions that produce immediate or cumulative impact with the potential to contribute to the end of the match, with the IMMEDIATE weighing more heavily than the cumulative impact*. This means that Effective Grappling (such as: Getting takedowns/reversals or trying submissions) should always outscore control time. This criterion has been included with six new rounds.
- Per the rules, a round should ALWAYS be given a 10-8 score if: *one fighter has dominated the action of the round, had duration of the domination and also impacted their opponent with either effective strikes or effective grappling maneuvers that have diminished the abilities of their opponent*. Dominance occurs when *the losing fighter is forced to continually defend, with no counters or reaction taken when openings present themselves*. This means that a round where one fighter's effective offense is big enough and his rival's is minimal or non-existent, the round should be scored as a 10-8. Six rounds have been added related to this criterion.
- Per the rules: *A judge shall assess if a fighter impacts their opponent significantly in the round, even though they may not have dominated the action*. Impact can be seen as effectiveness in damaging the opponent, such as visual impact (cuts, swelling. . .), or being close to the finish (knockdowns, submission attempts...). So, if a fighter has not dominated a round, but has had a lot more success impacting his opponent, he should be given the round with a 10-8 score. The final four theoretical rounds cover this criterion.

The structure of the JSON dictionaries for the theoretical rounds is identical to that of real rounds. Featured values in them have been manually set to emulate rounds where the information in each piece of the criteria is represented sufficiently, enabling the model to learn decision boundaries that may be underrepresented in real data.

3.2. Trained Models

In this initial version for Judgion, six different models have been trained. All models are neural networks, whose architecture, input preprocessing and training hyperparameters vary between each other. The training dataset corresponds to the one described in the previous section; however, it was additionally ***mirrored*** to enhance model robustness. Specifically, each round was included twice: once in its original form, and once with the both fighter statistics swapped, while adjusting the round label accordingly. This augmentation ensures that models do not implicitly learn positional biases associated with the fighter's corner (red vs. blue) and instead focus on the substantive fight statistics, sustained by the objective fact that, if a round which was won by one fighter were to occur in exactly the same way but with fighters reversed, the winner must necessarily be the other fighter.
Training procedure has had some common characteristics for all models.

- The metric used has been ***accuracy.***
- ***Categorical cross-entropy*** as the loss function.
- Identical output layer for all models: Four neurons with softmax activation, commonly used for multi-class classification.
- No dedicated validation or test split applied, as the focus was on model exploration rather than evaluation. In order to tune the models' architecture and hyperparameters, Grid Search technique was implemented and applied.

3.2.1. The Diagonal

The Diagonal is the first and most *naive* Judgion model. It consists of a single hidden layer with the same dimensionality as the input. Each hidden neuron is connected exclusively to its corresponding input feature and thus acts as a scalar weight on that statistic — a structure referred to as a *Diagonal Layer*. A schematic of the architecture is shown in Figure 1. The connection between the hidden and output layers is dense (fully connected).

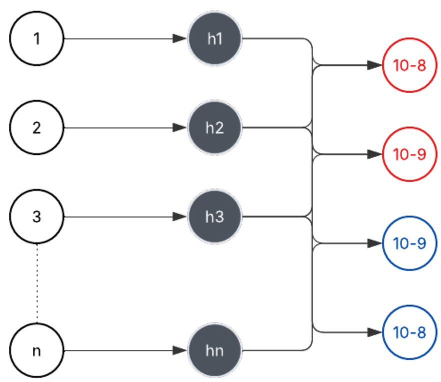

Figure 1: The Diagonal Model Architecture

This model was trained with the Adam optimizer. Its aim is to approximate how a human might judge a round using only aggregate statistics by assigning a distinct importance to each feature and then producing a decision from the resulting representation. Algebraically, this corresponds to multiplying the statistics vector by a diagonal matrix whose entries are learned during training. This approach motivates the model's name, *The Diagonal*. On the training dataset, the model achieved 86% accuracy.

3.2.2. The Cross

The Cross is, together with *The Diagonal*, the only model that employs a *Diagonal Layer*. Figure 2 depicts its architecture: two hidden layers—the diagonal layer followed by a dense hidden layer with ReLU activation. Both layers share the same number of neurons.

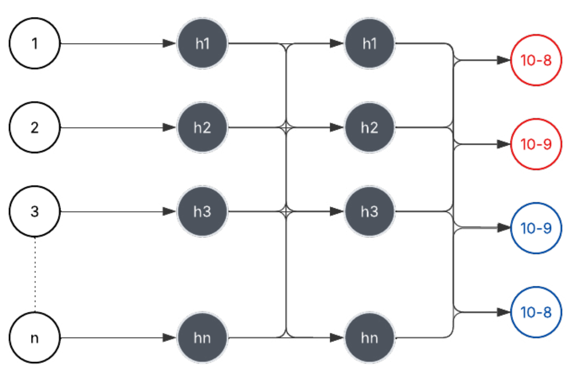

Figure 2: The Cross-Model Architecture

Adam was again the chosen optimizer. The objective is to augment *The Diagonal* by introducing an additional hidden layer capable of learning cross-feature interactions beyond the diagonal structure. As an incremental extension of *The Diagonal*, the model is named *The Cross*, reflecting the superposition of two diagonals. On the training dataset, it achieved 93% accuracy.

3.2.3. The Traditional

The Traditional is the first Judgion model that uses only fully connected (dense) layers. It comprises two hidden layers with ReLU activation and is trained with the Adam optimizer. As these are standard choices in machine learning, the model is termed *The Traditional*. It is also the only model that achieved 100% accuracy on the training set.

3.2.4. The Decayed

The Decayed is the first Judgion model to use input normalization. Instead of feeding raw counts, each feature is rescaled to lie in the [0, 1] range, with a simple per-feature max rule: for each statistic s, all the rounds in the dataset are scanned, in order to find the round r with the largest value M for each statistic s. Then, every round's value v for that statistic is divided by the maximum value M , i.e., v is replaced with v/M . In this way, the round that had the maximum value becomes 1, and all other rounds remain proportional. For example, if the maximum "head strikes landed" across the dataset is $M = 140$, a round with $v = 20$ maps to $20/140 \approx 0.14$ (= 1/7).

Architecturally, the network has two hidden layers—the first with a sigmoid activation and the second with ReLU—and it is trained with the AdamW optimizer, and it relied heavily on weight decay for its training (hence the name *The Decayed*). On the training dataset, the model achieved ∼ 87% accuracy.

3.2.5. The Handler

The Handler uses three hidden layers without input normalization. The activations are, in order, sigmoid (layer 1), ReLU (layer 2), and ReLU (layer 3), and the model is trained with the Adam optimizer. The naming reflects the role of the first sigmoid layer in "handling" the raw, unnormalized inputs by compressing their dynamic range and handling extreme values; the subsequent ReLU layers then sparsify the representation for the final decision. On the training dataset, the model achieved 95% accuracy.

3.2.6. The Pyramid

The Pyramid is the last and most complex Judgion model. Like *The Decayed*, it applies input normalization (per-feature max scaling to [0, 1], as described above). It uses three hidden layers with, respectively, sigmoid, GELU, and ReLU activations, and is trained with the AdamW optimizer. The model's name reflects its width profile: the first hidden layer contains 360 units (the largest among all Judgion models), the second 160, and the third 45; it then culminates in a 4-unit output layer (as in the other models). On the training dataset, the model achieved 93% accuracy.

4. Results

As noted in the labeling section, the principal challenge in evaluating Judgion models is the absence of a definitive *ground truth*. Although judging is subject to an objective set of rules, their application inevitably involves subjectivity. Consequently, the applied testing strategy emphasizes whether model outputs are defensible under the criteria rather than treating any single outcome as uniquely correct. In practice: a prediction is deemed as acceptable if it can be reasonably justified under the official criteria.

Since the training data spans events up until UFC 316, live tests were made during the UFC 317, UFC 318 and UFC 319 events. Tables 1, 2 and 3 summarize the official decisions alongside the corresponding Judgion predictions.

Bout	Official	Diagonal	Cross	Traditional	Decayed	Handler	Pyramid
Diniz vs Hines	29–28 (×3)	30–27	30–27	29–28	30–27	30–27	29–28
Araujo vs Cortez	27–30 (×3)	26–30	26–30	26–30	26–30	26–30	26–30
Talbott vs Lima	29–28 (×3)	29–28	30–27	30–27	30–27	29–28	30–27
Dariush vs Moicano	29–28 (×3)	29–28	29–28	29–27	29–28	29–28	29–28
Royval vs Van	28–29 (×2), 27–30	28–29	28–29	27–30	28–29	28–29	27–30

Table 1: UFC 317: Official Decisions and Judgion Scorecards by Model

Bout	Official	Diagonal	Cross	Traditional	Decayed	Handler	Pyramid
Prado vs Veretennikov	28–29 (×2), 29–28	28–29	29–28	28–29	29–28	28–29	29–28
Allen vs Vettori	30–27 (×2), 29–28	30–27	29–28	29–28	29–28	28–29	28–29
Oliveira vs Phillips	29–28 (×3)	29–27	29–28	29–28	29–27	29–27	29–27
Johnson vs Zellhuber	29–28 (×3)	30–27	30–27	30–27	30–27	29–28	30–27
Freire vs Ige	29–28 (×3)	30–27	29–28	29–28	30–27	29–28	30–27
Holland vs Rodriguez	28–29 (×3)	29–28	29–28	29–28	28–28	29–28	27–29
Costa vs Kopylov	30–27 (×2), 29–28	30–26	30–26	30–26	30–26	30–26	30–26
Holloway vs Poirier	49–46 (×2), 48–47	49–46	49–46	50–45	49–46	50–45	49–46

Table 2: UFC 318: Official Decisions and Judgion Scorecards by Model

Bout	Official	Diagonal	Cross	Traditional	Decayed	Handler	Pyramid
Klose vs Barboza	29–28 ($\times$3)	28–29	28–29	28–29	28–29	29–28	27–30
MVP vs Cannonier	29–28 ($\times$3)	28–29	28–29	28–29	28–29	29–28	28–29
Du Plessis vs Chimaev	44–50 ($\times$3)	44–50	45–50	44–50	44–50	43–50	44–50

Table 3: UFC 319: Official Decisions and Judgion Scorecards by Model

Overall, the results are largely consistent with the scoring criteria, although a few bouts show systematic divergences from the official scorecards and merit closer inspection:

- **Araujo vs Cortez (R3).** All Judgion models scored Round 3 as 10–8 for Cortez, whereas no official scorecard contained a 10–8. In that round, Cortez largely neutralized Araujo's offense, maintained initiative, and delivered sustained, effective offense with clear impact. Under the criteria (effective striking/grappling, then dominance and duration), a 10–8 is defensible.
- **Johnson vs Zellhuber (R3).** Five of six models scored Round 3 for Johnson. While Johnson clearly won Rounds 1 and 2, he reduced output in Round 3, which all three officials awarded to Zellhuber. The round was numerically close, but the qualitative assessment favors Zellhuber; this is likely a *misjudged* round by the models.
- **Holland vs Rodriguez.** For this fight, fans and experts scorecards were widely dispersed. None of the Judgion models matched the unanimous official 29–28 for Rodriguez. Round 1 is arguably scoreable for either fighter depending on how visible damage is weighted; Round 2, while officially 10–9 Rodriguez, presents a reasonable case for 10–8 given the degree of control and impact Rodriguez applied in the duration of the round. Therefore, Judgion cards are *arguable* under the criteria despite the obvious disparity with the official scoring.
- **Freire vs Ige (R3).** Three out of six judges scored this round for Freire. While Freire dominated the first two rounds, the momentum swinged in the third round, where Ige was able to create some effective striking and pressure, even inflicting big damage with a knee to the temple. In spite of that, Freire was able to achieve some control time for himself, but it was hardly enough for him to take the round; therefore, this round's judging was *questionable* for half of the Judgion models.
- **Costa vs Kopylov (R1).** All Judgion cards scored a 10–8 in Round 1 (overall 30–26). Although Costa clearly won the round, the combination of dominance and opponent inactivity did not rise to a strong 10–8 threshold; therefore, this round is marked as *misjudged*.
- **Klose vs Barboza.** Five of six models scored the bout for Barboza despite unanimous 29–28 Klose officially. Both Rounds 1 and 3 are reasonably debatable under the criteria, so the model cards remain broadly justifiable even though they diverge from the official result.
- **MVP vs Cannonier (R2).** Five of six models gave Round 2 to Cannonier (10–9). This appears *misjudged*: despite favorable counts for Cannonier, the effective striking with notorious impact in that round came from MVP.

There are also isolated outliers where the ensemble mostly aligned with the officials, but one or two model cards warrant attention:

- **Dariush vs Moicano (R2).** *The Traditional* produced a 10–8 score for Dariush in Round 2. While Dariush won the round, the dominance/duration thresholds do not support a 10–8; this round was labeled as *misjudged*.
- **Allen vs Vettori.** *The Handler* and *The Pyramid* scored the bout 28–29 for Vettori, contrary to the officials. While Round 2 could be argued in favour of Vettori, Round 3 was unanimously scored in favour of Allen in all official, experts and fans scorecards; the models *misjudged* this round.
- **Klose vs Barboza.** *The Pyramid* produced a 30–27 for Barboza. Given Klose's damage, striking output and pace control in Round 2, a sweep for Barboza is not supported; this 30–27 is labeled as a *mistake*.

Despite the disparity seen in some of these results with the official scorecards, it's proven that the majority of the scores produced by the models are actually justifiable under the current criteria. By the numbers, across these three events we evaluated 52 rounds in total. Of these, 7 rounds were flagged as *misjudged* under the criteria (Johnson–Zellhuber R3, Freire–Ige R3, Costa–Kopylov R1, MVP–Cannonier R2, plus the isolated model outliers: Dariush–Moicano R2 by *The Traditional*, Allen–Vettori R3 by *The Handler* and *The Pyramid*, and Klose–Barboza R2 by *The Pyramid*). This corresponds to $7/52 \approx 13.5\%$ of rounds. At the lower level, these cases account for 23 flagged decisions out of 312 reviewed, (7.4% approximately). It is also important to consider that some of the misjudged rounds, such as Cannonier vs MVP Round 2, are just very difficult rounds to score just with the tracked statistics. If the data were more in-depth about what's happening in the round, models would have an easier time judging them; when a head kick, an overhand and a jab are all counted as '*one significant strike to the head*', models' scores can't be 100% reliable.

5. Conclusions

This work introduced *Judgion*, a family of neural network models for round-level MMA scoring using round-by-round statistics. Six total models were presented, each one with its unique approach, and all achieving high training accuracy, and in out-of-sample, live tests (UFC 317–319) their scorecards were mostly defensible under the current criteria. Given the inherent subjectivity of judging, the obtained results in said tests indicate that Judgion's performance is *promising*, while there's still work to be done to further improve the automated scorecards. MMA judging is a very complex problem where two constraints dominate: the lack of *ground truth* and the *coarseness* of the available data, as many distinct actions are collapsed into the same statistical bin (i.e. both a jab and a head kick being counted as a *significant strike to the head*), obscuring impact and context. The evaluation done through this work emphasized whether a model's score can be *reasonably justified* under the criteria rather than strict agreement with a single source. Put differently: if the data were more granular about *what* happened and *how much it mattered*, models would have an easier time judging consistently.

While there is room for improvement, the present results support the claim that Judgion's performance is promising. For future work, the main goal will be improving the models' performance. That will be not only about better training techniques, architectures or hyperparameters; it will be an effort to obtain a better, more in-depth dataset that allows models to learn beyond the available data. While the addition of cuts as a statistic and the introduction of *theoretical rounds* is a good step in the right direction, there's more to be done in the future to achieve models that actually know how to score MMA bouts [8].

5.1. Releases and Live Results

Judgion's project is publicly available at *Judgion Github*. The public repository includes the main scripts used during the development of Judgion, the set of *theoretical rounds* used during training, and some example rounds for testing. I additionally included graphs extracted from the training dataset I used for the models.

The current release is a **beta** version (labeled as *v*0.0.0). Judgion will be actively updated in the future; the first planned milestone is scheduled for **January 2026.**

For dissemination, live scorecards during UFC events will be posted live on X (Twitter) and Instagram, at @judgion and @ judgionmma, respectively. The first UFC event that will be covered by Judgion will be UFC 320.

When citing this work, please reference the repository as shown in Ontiveros [2025]. For correspondence: jud- gion@gmail.com.

References

1. FightMatrix. Ufc fight outcomes by year. 2025. Accessed: 025-08-21.
2. Yin, J. (2024, June). Data-Driven MMA Outcome Prediction Enhanced by Fighter Styles: A Machine Learning Approach. In *2024 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)* (pp. 346-351). IEEE.
3. Nate Latshaw. Introducing judgeai: A machine learning algorithm that predicts judges' scores by round in the ufc. 2021. Accessed: 2025-07-30.
4. Feldman, T. (2020). The way of the fight: An analysis of MMA judging. *Journal of Applied Sport Management, 12*(2), 5.
5. MMADecisions. Mma decisions. 2025.
6. Verdict MMA. Verdict mma — live scoring, picks, and community 2025.
7. Association of Boxing Commissions and Combative Sports. Unified rules of mixed martial arts. Technical report, ABC, 2019.
8. Iván Ontiveros. Judgion: A First Approach to Applying Machine Learning Techniques To Judge Mixed Martial Arts Bouts, 2025