

Beyond the Meta: Leveraging Game Design Parameters for Patch-Agnostic Esport Analytics

Alan Pedrassoli Chitayat, Florian Block, James Walker, Anders Drachen

University of York, York, UK

{alan.pchitayat, florian.block, james.walker, anders.drachen}@york.ac.uk

Abstract

Esport games comprise a sizeable fraction of the global games market, and is the fastest growing segment in games. This has given rise to the domain of esports analytics, which uses telemetry data from games to inform players, coaches, broadcasters and other stakeholders. Compared to traditional sports, esport titles change rapidly, in terms of mechanics as well as rules. Due to these frequent changes to the parameters of the game, esport analytics models can have a short life-span, a problem which is largely ignored within the literature. This paper extracts information from game design (i.e. patch notes) and utilises clustering techniques to propose a new form of character representation; namely Clustered Character Representation (CCR). As a case study, a neural network model is trained to predict the number of kills in a Dota 2 match utilising CCR. The performance of this model is then evaluated against two distinct baselines, including conventional techniques. Not only did the model significantly outperform the baselines in terms of accuracy (85% AUC), but the model also maintains the accuracy in two newer iterations of the game that introduced one new character and a brand new character type. These changes introduced to the design of the game would typically break conventional techniques that are commonly used within the literature. Therefore, the proposed methodology for representing characters can increase the life-span of machine learning models as well as contribute to a higher performance when compared to traditional techniques typically employed within the literature.

Introduction

Esport titles, such as *League of Legends* and *Dota 2*, have amassed both large audiences and player-bases (Newzoo 2022; Petrovskaya and Zendle 2020). Due to the competitive nature of the genre, the player community often develop so called “*metas*” as explained by Kokkinakis et al. (2021). According to the author, *metas* are naturally discovered and developed strategies for optimum ways of playing the game that are focused in determining competitive advantage available within the current parameters of the game design. As a result, game developers are constantly updating and changing the rules of the game in order to balance the game, prevent game stagnation as well as to maintain player interest. This is done through releasing *patches* to the game, which

can alter the rules and the parameters of the game (Summerville, Cook, and Steenhuisen 2016). Those changes can be manifested by introducing new content (such as new characters), changing existing content (such as altering the duration or cost of an in-game ability), removing old content (such as discarding excising mechanics or rules) or a combination of them (such as replacing an ability for an existing character to a new ability with unique traits).

As the patch changes are generally made with the intent to re-balance the game, they often incur changes to the meta, forcing competitive players to play differently and creating new game mechanics interactions. As outlined by Demediuk et al. (2021), the way in which a game is played, and the current rules set could impact machine learning models and other forms of data analysis. This could be caused both by differences in player decision making, as well as by the new environmental state (i.e. the changes to the game’s parameters themselves). As a consequence, much of the esport literature is done focusing on a limited periods of time, to reduce the impact of those changes (Demediuk et al. 2019; Katona et al. 2019; Pedrassoli Chitayat et al. 2020; Tot et al. 2021). However, as the game continues to change, the performance of such models may suffer, and models may need to be re-trained to support a new architecture.

One example that can be observed in the esport literature comes from models that include the characters present in a match. Typically, characters are represented by their unique Character IDs (which are arbitrarily assigned numerical identifiers), which most commonly undergo a one-hot encoding or a similar variation (Summerville, Cook, and Steenhuisen 2016; Makarov et al. 2018; Katona et al. 2019; Vigiato and Bezemer 2020; Demediuk et al. 2021; Ringer et al. 2023). In this encoding, the unique identifier of a character is represented as a vector class. This means that if a game only has four characters, with IDs 1 through 4 respectively, character 1 can be presented as the vector $(1,0,0,0)$, while character 2 would be $(0,1,0,0)$, etc... This can lead to an architectural problem where, as the number of Character IDs increases so does the number of dimensions needed to encode them. Using the same example, if a new character is then introduced with Character ID 5, the vector needed to represent all previous character would need an additional dimension. Thus character 1 would then be represented as the vector $(1,0,0,0,0)$. A machine learning model developed

prior to the introduction of the 5th character would be architected to support a 4-dimensional input vector for each character, and thus not support a 5-dimensional input needed given the new game design parameters. This could lead the model to be unusable due to the technical constraints of how it was trained and applied.

Additionally, characters within esports games usually have unique abilities and traits which allows them to be played in different ways (Demediuk et al. 2019, 2021). If a character is re-designed in a way in which several of their original abilities or traits are altered, the way in which they are played could significantly change from patch to patch. Those changes would not be encompassed by the Character IDs. Using the same example as above, if character 1 is designed to pose several supportive traits (such as healing or otherwise improving the in-game performance of allies) but then is re-designed to have several offensive abilities, the way in which the character is played could be changed significantly. However, as the character ID would remain the same, a model that uses the ID could have its performance impacted, as it had been trained in a different state that does not account for the new changes to the character. In these hypothetical scenarios, while the model would still be able to produce results, it would be unclear if the results are reliable due to the uncertainty of the impact of the change to the game's environment.

Thus, this paper conjectures that changes to game design parameters may be interpreted in three ways:

- Breaking changes - These are fundamental changes to the design of the game which would require a change in model architecture in order to produce any results.
- Impactful changes - These do not incur changes in architecture, however have a substantial impact in model performance.
- Unimpactful changes - These do not alter the state of the game significantly enough to affect the performance of previously trained models.

While breaking changes are usually trivial to identify, due to their severe disruption to the application of models, differentiating between impactful and unimpactful changes may require analysis. Furthermore, if either breaking or impactful changes are identified, a previously trained model may need to undergo the training process again to account for the new parameters. This can be a cumbersome process, especially as esports titles typically change rapidly and abruptly (Summerville, Cook, and Steenhuisen 2016; Kokkinakis et al. 2021).

This paper builds on some of the methodology suggested in the literature (Demediuk et al. 2019; Viggiano and Bezeemer 2020), to generate a novel character representation. This form of representing characters utilises patch specific game design data, which is then clustered to allow for a fixed and reusable notation that can be readily applied to future models. The Clustered Character Representation (CCR) method holds meaningful information about the character's capabilities and it is sensitive to changes introduced in patches. This has been done using *Dota 2* - a popular Multiplayer Online Battle Arena (MOBA) esports title - which contains over 100

unique characters. Each playable character has a range of unique abilities - which are active skills that can be used by players during play with in-game effects, such as causing damage or healing allies - and stats, including "Intelligence", "Agility", "Strength", etc... *Dota 2* is a team based game, in which two teams of 5 players each attempt to destroy the main building in the opponent's team base. This game has been chosen as a focus in this study due to its popularity within academia (Katona et al. 2019; Pedrassoli Chitayat et al. 2020; Tot et al. 2021; Semenov et al. 2017; Hodge et al. 2019; Agarwala and Pearce 2014; Ringer et al. 2023; Makarov et al. 2018), large complexity and abundant access to data. While the resources made available within this paper are designed for *Dota 2*, a similar methodology could readily be applied to other MOBA titles - such as *League of Legends* - with minimal alterations, as well as advising the development of similar techniques for other esports genres.

CCR is then tested and evaluated in a case-study, which simulates a hypothetical future work within the esports literature. Three versions of a Neural Network (NN) are trained using professional games of *Dota 2* from patches 7.27 to 7.33. These models attempt to predict the number of kills (also referred to as the score) for each team at the end of the match. It is important to note that these NNs bear no direct contribution within the paper other than as an evaluation metric, however kill prediction was selected as a case study as it is a promising, under-explored area of the literature (Schubert, Drachen, and Mahlmann 2016; Katona et al. 2019; Ringer et al. 2023). Thus this serves as an illustration possible use case within future research, and outline the capabilities of this methodology. (NN1) was used as a simple base-line, where only the match duration was used as an input. No additional features were included in this baseline, therefore this network would have no way of determining or modeling the characters present in the game. (NN2) was utilised as an additional control, where Character IDs (one-hot encoded) were used to represent the characters present in each team, as well as the duration of the match. This network controls for the standard encoding typically used in the literature, which holds information about the characters that have been selected in the match. Lastly, (NN3) was trained using the characters through CCR, as well as the match duration.

Towards addressing the impact of game changes in machine learning models and other forms of data analytics within esports, which poses a problem of short life-span, this study provides 3 major contributions:

- A feature set of character traits is compiled through the literature and game-design data made available by the game's publisher.
- A novel way of representing characters (CCR) is proposed and validated through performance, outlining how it is sensitive to patch specific context as well as reliant to fundamental changes to the core game environment.
- Access to CCR, including standardized format, centroids, clustered abilities and characters are made freely available for use for future research in the field¹.

¹<https://github.com/ChitaAPC/Dota2CCR>

Related Work

Within the esports literature many studies which aim to apply data analysis and machine learning techniques to esports data can be found. Those studies investigate a range of problems and applications, from classifying complex patterns (Demediuk et al. 2019), performing micro/macro predictions about the game state (Katona et al. 2019; Hodge et al. 2019) or provide advisory tools to tournament organisers (Tot et al. 2021), esports audiences (Pedrassoli Chitayat et al. 2020), and players (Summerville, Cook, and Steenhuisen 2016). The application of machine learning and other forms of analytics is vast within the literature. In this section, this paper will summarize and evaluate different examples of applications, with a focus on how characters were utilised and the potential impact of game changes to the work proposed, including any steps taken to mitigate this impact.

Summerville, Cook, and Steenhuisen (2016) provides an early example of the application of machine learning techniques. In this paper, Summerville, Cook, and Steenhuisen utilises both Bayes Nets (BNs) and Long Short-Term Memory Recurrent Neural Networks (LSTM RNNs) to investigate the character selection part of the game, also known as the “*Draft Phase*”. Using those techniques, the paper attempts to predict the characters which professional players choose to play for a given match. In this work, the authors outline the significant impact of patches to machine learning models, due to changes to the core game mechanics that alter the environment severely and abruptly. In this example, the characters were represented by One-Hot encoding as described previously. This means that, as identified by Summerville, Cook, and Steenhuisen, their model validity is no longer guaranteed as the parameters of the game change in patches. Additionally the model would be unable to support the addition of new characters.

Demediuk et al. (2019) introduces clustering techniques to automatically identify player roles (which are play styles adopted by players within a team strategy) within *Dota 2*. These roles are community determined and are not labelled within the game, but instead relate to how a player may play (for example) in a “*support*” role, aiming to enhance their teammates rather than themselves, while other players may play in a “*carry*” role, which aims at accumulating in-game resources to improve themselves while being supported by their teammates. Therefore, the study focused on player performance - i.e. player decision making and how it impacts the game. Later the authors extended this work to measure the player performance (i.e. how well a player is performing in a particular match) based on their detected roles (Demediuk et al. 2021). In this extended work, the authors utilised archetype analysis - a form of unsupervised learning similar to clustering - in the Key Performance Indicators (KPIs) available in the game to provide insights into how well players are performing within their detected role. In both cases the authors utilised the order in which abilities were upgraded by the players as a key indicator of role detection (in addition to other metrics such as spatio-temporal data). This outlines a clear importance on the abilities of the characters, and how they may dictate play-style, however, it is important to note that as patches change abilities, their

effect could also change. This could lead to a limiting the life-span of such model, as abilities are only represented by their ordering within the character. To mitigate that impact, the clustering model would likely need to be re-trained for each character when *impactful* changes are present as identified by the authors.

Similarly, Makarov et al. (2018) proposed a model for *Dota 2* (as well as *Counter Strike*) that utilises roles to generate a probabilistic prediction of the winner of the match. Within the *Dota 2* model, Makarov et al. utilised the knowledge of professional player role (as opposed to automatically detecting it through performance) in addition to other features, such as aggregate metrics of damage and resources to formulate its prediction. This model achieved predictions ranging from 73% to 90% AUC depending on the input features used in the function. Amongst the input features, the authors used the characters present in the match as one of the function’s parameters. While the exact form in which characters are used is not stated, it is conjectured that those undergo One-Hot encoded or similar. This is conjectured as this categorical value would need to be encoded into a numerical representation in order to be utilised in the function. For this reason, this function may also be subject to impact of changes in the game. However, it is unclear how much weight characters have within the function, and the use of other metrics may obfuscate or otherwise help minimize the impact of such data.

Another example is the work by Hodge et al. (2019), in which the authors developed a Random Forest model to predicting the winner of a *Dota 2* match. In Hodge et al. (2019), the use of match-state is used to predict the outcome of the game. The model achieved accuracy ranges varying from 70% to 90% depending on the in-game time of the prediction. This high level accuracy was achieved without including any character selection information, instead utilising only aggregate values from the match of features such as resource distribution (i.e. gold and experience), score (kills), and other in-game metrics. The use of these aggregate features and lack of information about the characters can mitigate the impact of patches, for the potential cost of contextual information. For example, a match which contains several characters that benefit from fighting in both teams could be interpreted differently from a match which contains several characters that benefit from avoiding fights and instead focusing on gathering/denying resources. Furthermore, it is important to note that even though characters are not included in the model directly, the current meta could still have consequential impacts in the model. This means that, if the meta when the model was trained revolved around team-fights and encounters, then more kills are to be expected. A change in the meta that pivots away from confrontations (or vice-versa) could then have an impact in the performance of the model.

By contrast, Semenov et al. (2017) outlines a comparison of performance of different techniques for predicting the winner of a match using data from character selection alone. Across the many methodology evaluated by the authors, it is clear that accuracy decreases as skill proficiency increases (i.e. the higher the player rank, the harder it is to

predict the winner based on character selection alone). The performance of the different methodology ranged from approximately 71% (Factorization Machines classifier for normal skill level) to approximately 64% (Naive Bayes for very high skill matches). In this evaluation of models, the authors encoded the character present through a variation of One-Hot encoding, where all characters present for the Radiant team were encoded as 1, while the characters present in the Dire team were encoded as -1 and 0 otherwise. Thus this approach is subjected to breaking changes when new characters are released (as previously explained in this paper) and would also be subjected to changes in the meta, as explained in the literature (Kokkinakis et al. 2021; Summerville, Cook, and Steenhuisen 2016).

Katona et al. (2019) introduced a different approach to handling esports characters. In this study Katona et al. utilised a predictive deep neural network, which utilised shared weights in the inputs which pertain to a single player/character within the dataset to predict the chances of a character being killed within five seconds. This approach allowed for greater flexibility than separating the characters, as the same weights of a network could be trained multiple times across all players for a given match. Later this work was extended by Ringer et al. (2023), in which a similar approach was taken to handle the characters and their associated data. Comparably to previously mentioned works, both Katona et al. and Ringer et al. represent the character selection through a One-Hot encoding of the Character IDs, similarly arising the risk of breaking changes as new content is released. Additionally, information related to the abilities of the characters were also utilised, indicating another example where key importance to the abilities can be found within the literature.

Older work in the literature by Agarwala and Pearce (2014) has attempted to utilise character selection data to identify strategies in order to predict the outcome. In the work by Agarwala and Pearce, the use of aggregate performance metrics for the characters themselves (across all players) were clustered, in order to determine similarities. This work differs from other works in the literature, as it focuses on the characters overall performance across the whole population, as opposed to individual player performance, such as (Demediuk et al. 2019; Makarov et al. 2018), or simply in the presence of the character in the match as a binary value such as (Semenov et al. 2017). According to Agarwala and Pearce (2014), using aggregate metrics of the population was insufficient to meaningfully represent the state of the game and detect strategies in order to produce accurate predictions. However, should more refined data be used, similar approaches could be beneficial to addressing the problem of changes to the game.

Later, Viggiano and Bezemer (2020) has taken the idea suggested by Agarwala and Pearce further. By focusing on individual player past performance with a character (i.e. wins/losses), in addition to some game design parameter about the character for a given patch (e.g. max health, attribute, attribute gain, etc.), as well as several other features, Viggiano and Bezemer proposed a XGBoost model to predict the winner of the match given the character selection.

By performing feature analysis, the authors concluded that, in addition to individual player's previous experience with their chosen character, the character's "raw attribute" (i.e. the game design parameters of the character excluding any player decision making) are the most significant features in predicting the outcome of a game within their model. In this work, the authors collected over 55 thousand professional games from the years of 2012 to 2020. This data expanded many patches, thus, was subjected to many core game design changes. The authors addressed this by parsing through the game's change logs (patch notes), which contains detailed information about the current game state at any given date. Thus the current iteration of game for any given match could be contextualised. However, the model was limited on character attributes without including any information about character abilities, which other points of literature suggest bear significant importance to how a character is played (Demediuk et al. 2019, 2021; Makarov et al. 2018; Katona et al. 2019; Ringer et al. 2023). Furthermore, Viggiano and Bezemer suggest that more data and a further breakdown of additional features may improve the performance further. Additionally, the author also included the characters present in a match using through a One-Hot encoding, making it similarly susceptible to breaking changes as new characters are introduced.

Overall, several points can be extracted from the literature. It is clear that changes in game design can impact the performance of models (Summerville, Cook, and Steenhuisen 2016; Viggiano and Bezemer 2020). It is also clear that high accuracy and proficiency can be achieved in models within the game iteration used during training, be it to predict the overall outcome (Hodge et al. 2019) or to perform micro prediction about in-game events (Ringer et al. 2023). High emphasis in the importance of character attributes and stats can also be observed (Viggiano and Bezemer 2020), as well as the importance of abilities, which can dictate how characters are played (Demediuk et al. 2019, 2021; Makarov et al. 2018; Katona et al. 2019; Ringer et al. 2023). It is also clear that character selection is a very common feature utilised in the literature (Summerville, Cook, and Steenhuisen 2016; Makarov et al. 2018; Katona et al. 2019; Demediuk et al. 2019; Viggiano and Bezemer 2020; Demediuk et al. 2021; Ringer et al. 2023). However, the typical approach to including character selection subjects models to breaking changes when new characters are introduced, or to impactful changes through updates to existing game design parameters. This could lead to the need of frequent retraining models, which can be a cumbersome process. Despite its wide spread utilisation, very little work has been done to formulate a consistent form of character representation that is patch-aware, to allow for changes in the game design parameters to be included in the analysis while supporting the addition of new characters. This paper aims to address this gap, by building from (Demediuk et al. 2019) and (Viggiano and Bezemer 2020) to introduce a methodology that allows researchers to contextualise game design parameters in a robust way that can support new characters without compromising the flow of data. Through analysing patch notes, it is possible to support a more rich environ-

ment for representing characters (CCR) that includes data from both, their attributes and their abilities.

Methodology

This section describes the steps taken to collect and process the data used. As explained above, the methodology proposed focuses on the *Dota 2* title, as a popular esports game with a wealth of academic research, and large player-base/audience.

Data Collection

In order to conduct this study, two distinct datasets were collected. Firstly patch data was collected, which consists of data about game design parameters for characters, including of the abilities and attributes from patches 7.27 to 7.33. Secondly, a match history dataset was compiled which consists of the character selection data of all competitive matches for these patches, as well as the duration of the match, the number of kills on each side and the result of the match.

Both datasets were collected through the OpenDota API² and its associated public table of constants³. OpenDota is a free platform that offers in-depth statistics and break down of *Dota 2* public matches, including those in professional tournaments and events. This platform has been commonly used by other works in the literature for data collection (Hodge et al. 2019; Demediuk et al. 2019, 2021).

As OpenDota’s table of constants is stored in a Git repository, it is possible to access older versions of the tables, referring to previous game patches, through the history of individual files or folders. The first dataset was compiled by parsing through three JSON files retrieved from this repository (“hero_abilities.json”, “abilities.json” and “heroes.json”).

The second dataset was compiled through OpenDota’s SQL query feature available on their platform. The data collected included:

- MatchID - a unique identifier for the match
- Patch - the patch in which the match was played in (such as '7.27')
- Duration - The duration of the match in seconds
- KillsR - The number of kills obtained by the Radiant team
- KillsD - The number of kills obtained by the Dire team
- HeroX - The hero played by player X, where X is a number from 0 to 9. Players 0-4 correspond to the Radiant team, and 5-9 correspond to the Dire Team
- RadiantWin - A binary variable containing 1 if Radiant won the game and 0 if Radiant lost (note it is not possible for a match to end in a draw in *Dota 2*)

Only professional and premium matches were collected from patches 7.27 to 7.33, which included data from *Jun 2020* to present. Games which did not conclude in a natural state (for example if a player has abandoned, or if there had been a server error) were not collected. This lead to a total

²<https://docs.opendota.com/>

³<https://github.com/odota/dotaconstants/tree/master/build>

| | | | | |
|----------------|--------|--------|-------|--------|
| Patch | 7.27 | 7.28 | 7.29 | 7.30 |
| Matches | 9,822 | 4,666 | 5,492 | 9,807 |
| Patch | 7.31 | 7.32 | 7.33 | Total |
| Matches | 13,476 | 16,915 | 2,235 | 61,254 |

Table 1: Matches distribution per patch

of 61,254 matches from the 7 different patches. (note that at the time of writing, the latest patch (7.33) is still active, and therefore fewer matches are available, as demonstrated in Table 1.

Data Processing

In order to perform clustering on the game design data, the JSON files retrieved from the table of constants were processed. Firstly “hero_abilities.json” contains the name of each individual character and their associated abilities. This file was parsed to compile a list of the relevant character abilities, which allowed abilities from non-playable entities, such as neutral characters, to be discarded. Using the name of each individual ability as a key, “abilities.json” was then parsed to generate a CSV file containing the properties of every character ability for each of the relevant patches. This was achieved by compiling a script that normalise and standardise the way data is represented.

Several inconsistencies were detected in the way data was stored by the game between patches and abilities. For example, an ability with the property of “movement_speed_slow” of 30 and another ability of “movement_speed_bonus” of -30 both have the same effect of reduce the movement speed of the target by 30 units. Through manually comparing the properties names and values and cross referencing to the game’s wiki, a script that standardised the properties was compiled. While this approach is subject to errors, gathering data is limited by the way in which data is made available by the game’s publishers. However, extracting data using this methodology allows for more in-depth analysis than previous techniques which generally ignore what abilities do and their associated properties (Demediuk et al. 2019; Katona et al. 2019). Once all of the properties for every ability was compiled, info on the characters attributes - as described in the literature (Viggiato and Bezemer 2020) - was extracted from the “heroes.json” file and amended to the relevant abilities for the heroes. Thus, each entry in the newly compiled CSV file contained data for each ability for each character as well as data about the character itself, which contextualises its use in the game. This process has been illustrated in Figure 1 and the CSV files have been made available in the resources listed in the Introduction section for future use.

Cluster Analysis

Once all of the character abilities were processed and the CSV files were compiled it was possible to utilise clustering algorithms on the data to produce CCR. As this study aims to provide longevity to machine learning models in esports analytics and support the addition of new unseen data, K-Means (Humaira and Rasyidah 2020) was utilised as it

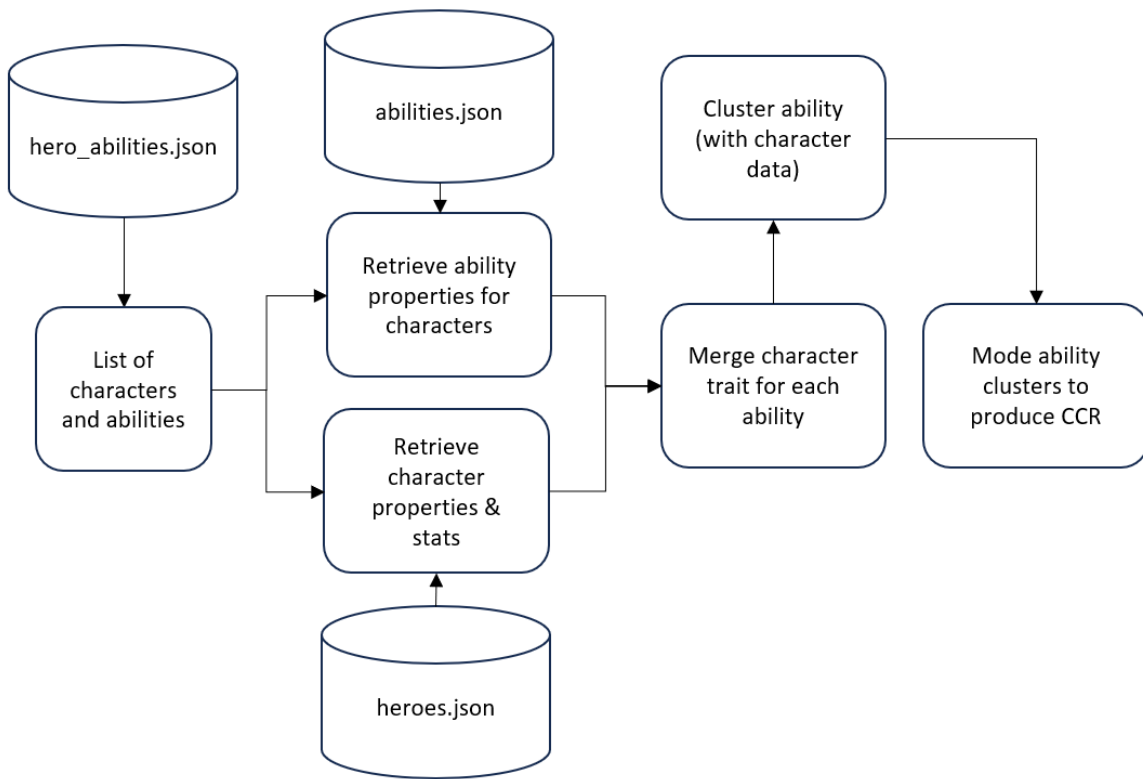


Figure 1: A depiction of how data was processed to produce CCR

can be readily used without altering the labeling or number of clusters once the centroids are found. This means that once the centroids are found, the model would not need to be retrained with unseen patches. Instead, the data from new patches could simply be predicted with the existing centroids, therefore clusters do not change when adding new data. This approach would then contribute to reducing the impact of breaking changes, as the value of K would not need to change unless the new data leads to a new concentration of clusters (i.e. an impactful change). In this case analysis of the cluster distribution - such as that described in this section - would reveal the need for new clustered. However, data from several patches (7.27 to 7.31) are used to generate the clusters to reduce the risk of impactful changes being present. (Note that data from patches 7.32 and 7.33 were excluded from cluster training, as those are used as test datasets in the later stages).

In this study, K-Means was used and the “Sum of the Squared Error” (SSE) was used to evaluate the split of the data and identify an optimum value for K (Humaira and Rasyidah 2020). Figure 2 outlines the elbow plot generated with different values for K . As the figure outlines, the elbow lies roughly between $K=40$ and $K=75$. The exact value for K was then selected using the highest “Silhouette Score” (Humaira and Rasyidah 2020) within this range, which provided $K=68$.

Once clustering was performed with the ability data, the labels obtained could then be used to represent characters.

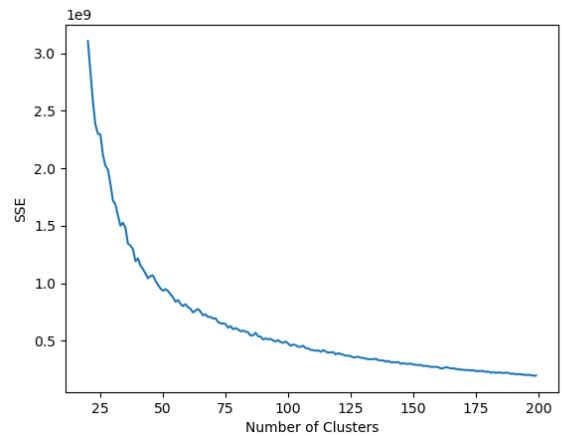


Figure 2: The Elbow Plot for the K-Means clustering of the character’s abilities and traits together

As one character contains multiple abilities, the cluster for each ability was encoding into a one-hot-encoded format. This means that any single ability was then represented as a K -long vectors of zeroes, except for the i ’s dimension, where i is the cluster number for that particular ability. Subsequently, one character can then be represented by adding all of the K -long vectors of their abilities. This allows one

character to be represented as a K-long vector, containing the modes for each of their ability clusters as depicted in Figure 1. Unlike the One-Hot encoding of Character IDs, however, CCR would not be subjected to breaking changes when new characters are introduced, as the value for K would remain the same.

For demonstrations purpose, this approach can be applied to a fictional game, where $K=3$. In this fictional game “Character X” has three abilities, of clusters 0, 0 and 2 respectively. This means that Character X can be represented as the addition of the vectors (1, 0, 0); (1, 0, 0) and (0, 0, 1). Through this technique, Character X can then be represented as a single K-dimensional vector of (2, 0, 1) - where the mode of *cluster 0* is two, *cluster 1* is zero and *cluster 2* is one. Conversely, a more complex character with 5 abilities of clusters 0, 0, 1, 1 and 2 respectively could be represented as the single vector (2, 2, 1). This simplistic, yet powerful approach allows for complex patterns to emerge from the data while still maintaining the meaning of each cluster as identified by the data. Furthermore, it maintains the size of the vector needed to represent any characters to K, even when new abilities or new characters are introduced by future patches.

Predictive Neural Network

Three NN are used to help evaluate CCR. This simulates a potential use-case within future research, with two baselines. The case-study of predicting the number of kills at the end of a match was selected, as this is a under-explored area in the literature with some of the potential highlighted by existing research (Schubert, Drachen, and Mahlmann 2016; Viggiano and Bezemer 2020; Tot et al. 2021). This is also a similar use-case as to predicting the winner of a game, which contains a wealth of knowledge (Viggiano and Bezemer 2020; Agarwala and Pearce 2014; Hodge et al. 2019; Makarov et al. 2018; Summerville, Cook, and Steenhuisen 2016). However, as the inputs for the models are highly controlled to enable validation, a comparison of performance between these NN and other predictive models would not be suitable. For this reason, the kills prediction use-case was selected instead of win prediction to enable and incentivise future research in the subject.

Firstly (NN1) was trained, which used only the match duration and no additional information. This provided a baseline for comparison, as all models include the match duration. This was included as the literature suggests that duration bears a significant impact in the outcome of the game, both in terms of result and in terms of score (Viggiano and Bezemer 2020). Secondly (NN2) was trained which contained both match duration and character selection data. The character selection data was represented by encoding the Character IDs into a one-hot encoding as a standard technique in the literature (Semenov et al. 2017; Ringer et al. 2023; Pedrassoli Chitayat et al. 2020). Lastly, (NN3) was trained using only the duration and the character selection data as represented by CCR. In order to train this model, whole line-ups - rather than individual characters - were used. This was achieved by replicating the steps for transforming an ability vector into a character vector as depicted

| Train | Validation | Test | Test (7.32) | Test (7.33) |
|--------|------------|-------|-------------|-------------|
| 27,171 | 6,793 | 8,491 | 16,564 | 2,235 |

Table 2: Number of matches per split used for training and evaluating the neural networks from the total 61,254 matches in the dataset

in the previous section. The individual character vectors for a team’s selection is added together collapsing it into one K-long vector. This means that vectors for abilities, characters and team compositions can all be represented in the same number of dimensions, which expresses the modes for each of the clusters.

All matches for patches 7.32 & 7.33 were excluded from the training set to be used as special test comparisons. Patch 7.32 introduced a brand new character, *Muerta*⁴. Patch 7.33 had several major changes to the game, including the way in which characters primary attributes are handled. In this patch, a new type of primary attribute was created that makes no attribute the primary. This means that - where before, characters were classed as one of “*Strength*”, “*Agility*” and “*Intelligence*”, as of the latest patch characters can also be “*Universal*”⁵, which means no attribute is their primary attribute. The data collected for the two patches serve as a illustration of breaking and impactful changes. Additionally, the remaining match data (patch 7.27 to 7.31) were split using a (64/16/20)% split into a training, test and validation datasets. Table 2 outlines the number of matches per each of the different data splits.

As an additional control, all networks were trained using the same architecture with the exception of the input layer, as described above. The networks consist of 6 hidden layers with (1024, 512, 128, 64, 32, 8) neurons respectively, with the “Sigmoid” activation function. These model predicts the number of kills for each team simultaneously (i.e. two outputs neurons, one for each team). Because of the simplistic architecture and the amount of data, batching was not used (i.e. batch size of 1) and the model was trained for 100 epochs, which was sufficient to outline the trends, detailed in the Results and Discussion Sessions. The networks were trained with an *Adam* optimizer and learning rate was set to $1e-4$. No dropout layers or regularizer were used. Figure 3 Outlines the train and validations loss for each of the three networks per epoch. Further work to this architecture could produce more reliable results, however a comprehensive study on this use-case is beyond the scope of this paper.

The performance of the predictions were then compared. This includes the training, validation and test performances for patches from 7.27 to 7.31, which outlines the performance within representative data. Additionally, the performance of predictions for games in patch 7.32 and 7.33, which introduced the new character and new character types were also compared. In order to make those predictions possible for (NN2), the maximum value used for the Character ID encoded included the Character ID for *Muerta*. This was

⁴<https://www.dota2.com/hero/muerta>

⁵<https://dota2.fandom.com/wiki/Universal>

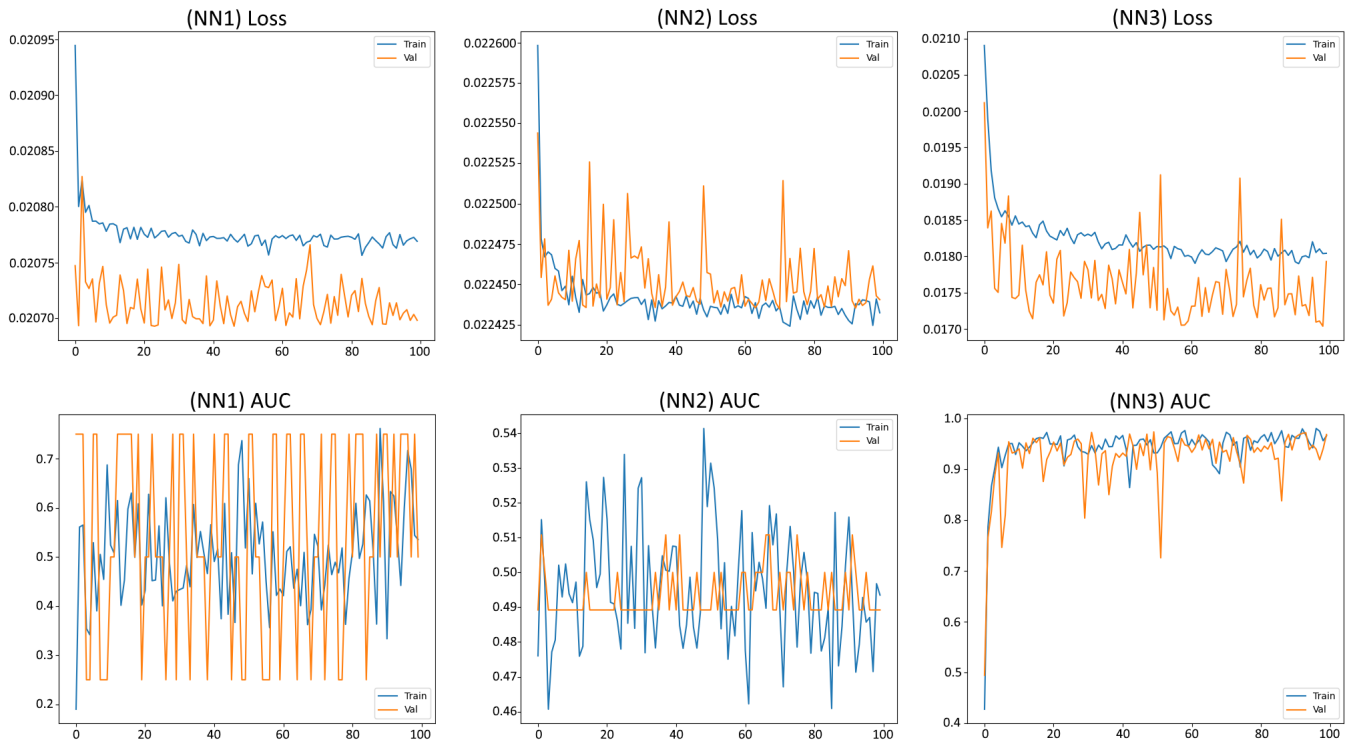


Figure 3: Training and validation loss and AUC graph per epoch for each neural network.

| NN | Test | Test (7.32) | Test (7.33) |
|-------|------|-------------|-------------|
| (NN1) | 0.50 | 0.50 | 0.50 |
| (NN2) | 0.50 | 0.49 | 0.49 |
| (NN3) | 0.85 | 0.85 | 0.86 |

Table 3: AUC of the kills prediction for each dataset

made with the sole purpose of ensuring the architecture supports the additional dimension. It is important to note, however, that this dimension would - necessarily - always have zero in the input vector for all training data. Therefore this approach would not be suitable for actually training models with real world applications and it is limited to exactly one additional character (thus, any future characters released would not be supported by the (NN2)’s architecture).

Results

The results obtained for all networks can be found in Table 3. The test dataset contains unseen matches from patches 7.27 to 7.31 - which are the same patches used in training dataset. Conversely, all matches from Patch 7.32 and 7.33 have been entirely excluded from the training dataset. Patch 7.32 introduced a new character, while Patch 7.33 (the currently active patch as of time of writing) introduced some significant changes to the game design, such as a new character type, as previously discussed.

Furthermore, both (NN1) and (NN2) typically produced the same prediction value for both outputs. This means that

for most matches, the prediction for the number of Radiant kills was exactly equal to the prediction for the number of Dire kills. Between the approximately 27 thousand matches in all test datasets (Test, Patch 7.32 & 7.33), (NN1) only produced different predictions for 351 matches while (NN2) produced different prediction 403 of the matches. On those cases (for both networks), the difference in kills only varied by one between teams. This was not the case for (NN3), which only predicted the same number of kills in approximately 3 thousand matches (11.1% of the dataset).

Discussion

As outlined by Figure 3, all three models had a sharp decrease on the loss followed by a plateau within the first 25 epochs. This suggests that all three models reached their convergent point within a few epochs. It is also noteworthy that the train and validation losses did not deviate significantly at any stage for any of the three neural networks. This suggests that none of the models overfit to the training data. Table 3 also suggests that there was no overfitting in any of the models, as the overall performance of each network was consistent from that observed with its respective training/validation accuracy at the 100th epoch.

Furthermore, both (NN1) and (NN2) produced very similar results throughout all test datasets. Thus, it is conjectured that (NN2) trained in a way to put very little significance for the lineup vectors, instead having more significant weights around the duration feature. This brings to question the use of lineup vectors as input features. Par-

ticularly when comparing similar performance in literature models observed when comparing the works of Hodge et al. (2019) - which does not include any character selection data - and Summerville, Cook, and Steenhuisen (2016) - which does - as well as the comparatively lower accuracy obtained by (Semenov et al. 2017) which only uses character IDs to (Viggiato and Bezemer 2020) which includes additional information.

Moreover, both (NN1) and (NN2) produced a relatively flat training AUC graph, despite the drop in loss at the early epochs. This may indicate that the models trained to produce a consistent prediction that minimizes the loss without detecting any meaningful patterns. In the other hand, (NN3)'s AUC graph produces a sharp rise in AUC, which is consistent with the drop in the loss. In addition to this, it significantly outperformed both baseline models (NN1 & NN2), pointing to the network detecting meaningful patterns in input data.

When comparing the performance of the three networks across the Test, Patches 7.32 and 7.33 datasets, it is clear that (NN3) produced consistent results despite the introduction of the new character and the change in character primary attributes observed in the two new patches. No observable difference could be detected in (NN1), which is expected, as it was not trained with any character information. Although there was a small drop in performance for (NN2), this could fall within the margin of error, particularly as the predicted values for this model resembles that of (NN1). Similarly, while there was a small increase in accuracy for (NN3) in Patch 7.33, that could also fall within the margin of error, particularly because there are fewer matches in that patch as the current live patch at the time of writing. However, as (NN3) maintained a high degree of accuracy through all patches, that is indicative that the CCR may hold significant underlying principles of the game that is able to contribute to the prediction.

It is important to note, however, that significant changes to the core game design - beyond that of the patches present in this study - could impact the cluster themselves and change the density distribution of the clusters. In this case new clustering may be required, which would similarly cause breaking changes to (NN3) - as the number of clusters or the meaning of a individual cluster may change. Thus, while this approach is a step towards more robust machine learning models, it is not completely devoid of risks of future impacts. Nevertheless, it reduces the risk of breaking changes, which provide improvements over the existing methodology (One-Hot encoding of Character IDs).

Lastly, the advantages observed in this study (maintaining high performance, supporting the addition of any number of characters) suggest that the CCR can increase the longevity of models developed in future research. Additionally, the results presented here suggests that this approach can more easily represent underlying patterns in the data, as observed by the differences in results between (NN2) and (NN3).

Conclusion

In conclusion, this paper outlined how the use of character selection data is a common feature for a range of ma-

chine learning models within the esports analytics literature. The impact risk of changes in the game design has also been covered in the literature, which often reduces the lifespan of models, limiting them to a single or few iterations of the game, while iterations are released rapidly. Despite that, very little work has been done to support changes to the game design parameters. In the current literature that relies on character selection data, no model could be found that decisively support the addition of a new character without needing to change its architecture (i.e. the introduction of new characters would incur breaking changes). Through clustering characters abilities and attributes, as extracted from the literature, this paper proposes a novel character representation (CCR) that both encapsulate the iteration of the game for any given patch and support the inclusion of future characters without the need for a change in architecture.

By evaluating CCR through a case-study, this paper simulated a potential use case for future research. Through comparing three predictive neural networks (NN1-3), this paper concludes that this CCR can add longevity to future models, while providing more information about the character selection than the standard methodology currently employed within the literature. This paper has made the resources available to use for future research, including the unclustered data; the centroids used; and the clustered representations of abilities and characters for *Dota 2*.

Ethical Statement

In esports analytics research, a considerable amount of publicly available data is utilized. Such data is usually shared via APIs developed from the publisher of the game in question. OpenDOTA collects a significant amount of data from players of *Dota 2* via querying the API system provided by the publisher of the game, Valve. All matches of *Dota 2* played are recorded and available via the game's API, which can include recording of gamer tags. No device information or similar is made available through the API, and neither are private chat logs. When you install *Dota 2*, you agree to the data collection and sharing through the End User Licence Agreement (EULA) presented when installing Steam, the platform through which *Dota 2* is installed. While performance data is always recorded as agreed by the EULA, gamer tag is not collected unless explicitly opted in by players to agree to have this data made publicly available.

The data used in the current study does not include personally identifiable information (PII). Gamer tag information was not included from the dataset, as these can sometimes be used across multiple online profile accounts and thus potentially be used to identify people. Collectively, we consider this use of data within reasonable ethical use under research ethics norms and expectations.

Acknowledgements

This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) EP/S022325/1.

References

- Agarwala, A.; and Pearce, M. 2014. Learning Dota 2 team compositions. <http://cs229.stanford.edu/proj2014/Atish%20Agarwala,%20Michael%20Pearce,%20Learning%20Dota%202%20Team%20Compositions.pdf>. Accessed: 2023-08-10.
- Demediuk, S.; Kokkinakis, A.; Patra, M. S.; Robertson, J.; Kirman, B.; Coates, A.; Chitayat, A.; Hook, J.; Nolle, I.; Olarewaju, O.; et al. 2021. Performance index: A new way to compare players. In *2021 MIT Sloan Sports Analytics Conference. IEEE*.
- Demediuk, S.; York, P.; Drachen, A.; Walker, J. A.; and Block, F. 2019. Role identification for accurate analysis in dota 2. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 15, 130–138.
- Hodge, V. J.; Devlin, S. M.; Sephton, N. J.; Block, F. O.; Cowling, P. I.; and Drachen, A. 2019. Win Prediction in Multi-Player Esports: Live Professional Match Prediction. *IEEE Transactions on Games*.
- Humaira, H.; and Rasyidah, R. 2020. Determining the appropriate cluster number using elbow method for k-means algorithm. In *Proceedings of the 2nd Workshop on Multi-disciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia*.
- Katona, A.; Spick, R.; Hodge, V. J.; Demediuk, S.; Block, F.; Drachen, A.; and Walker, J. A. 2019. Time to die: Death prediction in dota 2 using deep learning. In *2019 IEEE Conference on Games (CoG)*, 1–8. IEEE.
- Kokkinakis, A.; York, P.; Patra, M.; Robertson, J.; Kirman, B.; Coates, A.; Pedrassoli Chitayat, A.; Demediuk, S. P.; Drachen, A.; Hook, J. D.; et al. 2021. Metagaming and metagames in Esports. *International Journal of Esports*.
- Makarov, I.; Savostyanov, D.; Litvyakov, B.; and Ignatov, D. I. 2018. Predicting winning team and probabilistic ratings in “Dota 2” and “Counter-Strike: Global Offensive” video games. In *Analysis of Images, Social Networks and Texts: 6th International Conference, AIST 2017, Moscow, Russia, July 27–29, 2017, Revised Selected Papers 6*, 183–196. Springer.
- Newzoo. 2022. Newzoo’s Global Esports & Live Streaming Market Report 2022 — Free Version. <https://newzoo.com/resources/trend-reports/newzoo-global-esports-live-streaming-market-report-2022-free-version>. Accessed: 2022-02-09.
- Pedrassoli Chitayat, A.; Kokkinakis, A.; Patra, S.; Demediuk, S.; Robertson, J.; Olarewaju, O.; Ursu, M.; Kirmann, B.; Hook, J.; Block, F.; et al. 2020. WARDS: Modelling the Worth of Vision in MOBA’s. In *Intelligent Computing: Proceedings of the 2020 Computing Conference, Volume 2*, 63–81. Springer.
- Petrovskaya, E.; and Zendle, D. 2020. The battle pass: A mixed-methods investigation into a growing type of video game monetisation. *OSF Preprints, Sep*.
- Ringer, C.; Missaoui, S.; Hodge, V. J.; Chitayat, A. P.; Kokkinakis, A.; Patra, S.; Demediuk, S.; Munoz, A. C.; Olarewaju, O.; Ursu, M.; et al. 2023. Time to Die 2: Improved in-game death prediction in Dota 2. *Machine Learning with Applications*, 100466.
- Schubert, M.; Drachen, A.; and Mahlmann, T. 2016. Esports Analytics Through Encounter Detection. In *10th MIT Sloan Sports Analytics Conference*.
- Semenov, A.; Romov, P.; Korolev, S.; Yashkov, D.; and Neklyudov, K. 2017. Performance of machine learning algorithms in predicting game outcome from drafts in dota 2. In *Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7-9, 2016, Revised Selected Papers 5*, 26–37. Springer.
- Summerville, A.; Cook, M.; and Steenhuisen, B. 2016. Draft-analysis of the ancients: predicting draft picks in dota 2 using machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 12, 100–106.
- Tot, M.; Conserva, M.; Chitayat, A. P.; Kokkinakis, A.; Patra, S.; Demediuk, S.; Munoz, A. C.; Olarewaju, O.; Ursu, M.; Kirmann, B.; et al. 2021. What Are You Looking At? Team Fight Prediction Through Player Camera. In *2021 IEEE Conference on Games (CoG)*, 1–8. IEEE.
- Viggiato, M.; and Bezemer, C.-P. 2020. Trouncing in dota 2: An investigation of blowout matches. In *Proceedings of the AAAI conference on artificial intelligence and interactive digital entertainment*, volume 16, 294–300.