Universidade Estadual de Campinas
Instituto de Computação

Maisa Silva

Graph-based Spatiotemporal Data Analysis for Passing Difficulty Classification in Football

Análise de Dados Espaço-temporais Baseada em Grafos para Classificação de Dificuldade de Passe no Futebol

CAMPINAS
2024

Maisa Silva

Graph-based Spatiotemporal Data Analysis for Passing Difficulty Classification in Football

Análise de Dados Espaço-temporais Baseada em Grafos para Classificação de Dificuldade de Passe no Futebol

Dissertação apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Computer Science.

Supervisor/Orientador: Prof. Dr. Ricardo da Silva Torres
Co-supervisor/Coorientador: Dr. Allan da Silva Pinto

Este exemplar corresponde à versão final da Dissertação defendida por Maisa Silva e orientada pelo Prof. Dr. Ricardo da Silva Torres.

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

Si38g

Silva, Maisa, 1986-
Graph-based spatiotemporal data analysis for passing difficulty classification in football / Maisa Silva. – Campinas, SP : [s.n.], 2024.

Orientador: Ricardo da Silva Torres.
Coorientador: Allan da Silva Pinto.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Representações dos grafos. 2. Futebol. 3. Redes temporais. I. Torres, Ricardo da Silva, 1977-. II. Pinto, Allan da Silva, 1984-. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Informações Complementares

Título em outro idioma: Análise de dados espaço-temporais baseada em grafos para classificação de dificuldade de passe no futebol
Palavras-chave em inglês:
Representations of graphs
Football
Temporal networks
Área de concentração: Ciência da Computação
Titulação: Mestra em Ciência da Computação
Banca examinadora:
Allan da Silva Pinto [Coorientador]
Milton Shoiti Misuta
Bruno Luiz de Souza Bedo
Data de defesa: 26-01-2024
Programa de Pós-Graduação: Ciência da Computação

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-0994-2881
- Currículo Lattes do autor: http://lattes.cnpq.br/6794119441625711

Universidade Estadual de Campinas
Instituto de Computação

Maisa Silva

Graph-based Spatiotemporal Data Analysis for Passing Difficulty Classification in Football

Análise de Dados Espaço-temporais Baseada em Grafos para Classificação de Dificuldade de Passe no Futebol

Banca Examinadora:

- Dr. Allan da Silva Pinto
  LNLS/CNPEM
- Prof. Dr. Milton Shoiti Misuta
  FCA - UNICAMP
- Prof. Dr. Bruno Luiz de Souza Bedo
  EEFE - USP

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 26 de janeiro de 2024

Dedicatória

Aos meus queridos sobrinhos, Lorenzo Rizzioli e Vinicius Rizzioli. Que este esforço dedicado à busca do conhecimento possa inspirar e iluminar o caminho de vocês, assim como a presença de vocês enriquece a minha vida.

O passe é devoção; o drible, inspiração. (Armando Nogueira)

Acknowledgements

I thank God, the source of strength and inspiration, for guiding my steps throughout this academic journey. To my beloved parents, Hiêda de Fátima Mania Silva and Walter Silva, I express my infinite gratitude for their unconditional love, tireless support, and the values that shaped my character.

To my respected supervisor, Prof. Ricardo da Silva Torres. I am grateful for the solid guidance, patience, and intellectual stimulus that were essential for the development of this dissertation. I extend my sincere appreciation to the co-supervisor, Dr. Allan da Silva Pinto, as well as Prof. Paulo Roberto Pereira Santiago and Dr. Murilo Merlin for their valuable contributions and insights that enriched this work.

I cannot fail to express my recognition to the dedicated professors of the Institute of Computing at Unicamp, whose expertise and commitment provided a stimulating and enriching academic environment.

To all who, in some way, contributed to this achievement; thank you very much. This work is the result of the collective effort of many, and each of you has a special place in my journey. I sincerely thank you for being part of this significant chapter in my academic life.

This work was partially funded by São Paulo Research Foundation (FAPESP) – Grants #2019/17729-0, #2016/50250-1, and #2017/20945-0. This work received support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Financing Code 001.

Resumo

Diversas pesquisas têm aplicado modelos matemáticos para descrever os padrões de partidas de futebol, avaliar seus componentes e fornecer informações cruciais para a tomada de decisão de técnicos e equipes, tanto para planejar estratégias para partidas quanto definir procedimentos para treino. Um dos eventos mais comuns em uma partida de futebol é o passe, que consiste na transmissão da bola de um jogador para outro. Muitas métricas são estudadas em relação ao passe, como a frequência entre dois jogadores, o número de passes bem-sucedidos, a precisão do passe, entre outras. No entanto, a dificuldade do passe ainda é um tema pouco explorado na literatura científica. Nosso estudo apresenta uma abordagem para caracterizar passes em diferentes graus de dificuldade: fácil, médio e difícil. Avaliamos quais representações adotadas pelos elementos do grafo (vértices, arestas e medidas de rede complexa) são mais adequadas para a modelagem do problema. Em seguida, comparamos o impacto de um maior detalhamento temporal no desempenho da classificação do passe. Por fim, avaliamos se a fusão de vetores de características influencia o desempenho da classificação do passe. Os resultados mais destacados incluem a constatação de que grafos contendo apenas o time que possui a posse da bola não são uma boa representação para o problema. Além disso, a melhor representação de grafos para resolução temporal binária apresenta arestas para jogadores a até 5 metros de distância, com o grau sendo a melhor medida de rede complexa. Nesse caso, a acurácia balanceada atingiu 61%. Para a avaliação multinível, a melhor resolução temporal foi de 20 frames com uma configuração de grafo bipartido, na qual as arestas representam a interferência do oponente em relação ao alvo, e a melhor medida foi a *closeness*; esse conjunto de vetores de características alcançou uma acurácia de 65%, o mesmo valor do índice de concordância dos dois técnicos rotuladores. A combinação de vetores de características, no entanto, não apresentou um ganho significativo na acurácia balanceada, atingindo o valor de 66%. Estes resultados sugerem que os melhores vetores de características não provêm visões complementares para a tarefa de classificação.

Abstract

Several research studies have applied mathematical models to describe the patterns of football matches, assess their components, and provide crucial information for coaches and teams, both for planning strategies for matches and defining procedures for training sessions. One of the most common events in a football match is the pass, which involves the transmission of the ball from one player to another. Many metrics are studied concerning the pass, such as the frequency between two players, the number of successful passes, and pass accuracy, among others. However, the assessment of pass difficulty is still a relatively unexplored topic in the scientific literature. Our study presents an approach to characterize passes at different difficulty levels: easy, medium, and difficult. We evaluate which graph elements (vertices, edges, and complex network measurements) are the most suitable for modeling the problem. Next, we compare the impact of increased temporal granularity on pass classification performance. Finally, we assess whether the fusion of feature vectors influences pass classification performance. The most promising results include the finding that graphs containing only the team in possession of the ball are not suitable representations for the problem. Additionally, the best graph representation for binary temporal resolution includes edges to players up to five meters away, with degree being the best complex network measure. In this case, balanced accuracy reached 61%. For multilevel evaluation, the best temporal resolution was 20 frames with a bipartite graph configuration, in which edges represent opponent interference concerning the target, and the best measure was closeness. This feature set achieved an accuracy of 65%, the same value as the agreement index of the two annotator coaches. However, the combination of features had not led to significant gains in balanced accuracy, reaching a value of 66%. This result suggests that the best features do not provide complementary views for the classification task.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

In football, a collective sport, a multitude of spatiotemporal events occurs across different temporal and spatial scales. Take, for instance, the pass – an event in which the ball undergoes a positional shift, i.e., a translation in space occurring within a specific time interval. The interactions among players, both within and between teams, play a crucial role in determining the success or failure of actions. Effectively modeling and representing the properties associated with these interactions, to support performance analysis, is a complex and challenging task [2, 41].

Passes are one of the most frequent events in football. Most teams execute more than 500 passes in a single game [24]. Notably, successful teams in tournaments like the 2010 World Cup demonstrated a significantly higher number of passes [33]. Existing literature on football passes has identified certain statistics, such as the ratio of short to long passes [40] and passing accuracy (defined as the ratio of completed passes to total passes) [13]. These insights underscore the importance of evaluating and enhancing passing strategies, making it a key focus for both individual players and teams [24].

Early investigations into passing in football primarily centered on statistical analysis, such as the frequency, density, and sequence of events. Notably, Chassy [12] found that passing density (the number of passes per minute of possession) and precision (the ratio of completed passes to total passes) serve as highly accurate predictors (99.85%) of a team's ability to maintain ball possession and generate shooting opportunities (94.92%). Additionally, Gyarmat et al. [28] discovered that homogeneous passing structures (sequences of players between passes) tend to be associated with success.

More recently, advancements in player tracking technology, providing accurate positional data for all players on the field, have facilitated the introduction of novel variables and spatiotemporal analyses in football research [11, 20]. Horton et al. [30] introduced an efficient yet computationally expensive $O(n^2)$ algorithm for calculating dominant regions. They extracted features from these dominant regions and utilized them to classify pass quality, achieving an accuracy of 85.8%. Similarly, Merlin [36] evaluated pass difficulty using spatiotemporal data. Key variables considered included the presence of opponents between the passing player and the receiver player at the final pass moment, as well as the density (5m) at the initial pass moment and the density (5m) at the final pass moment.

Systems that enable player tracking throughout entire football matches have also provided the scientific community with an extensive dataset that opened the opportunity

for understanding the technical and tactical factors that influence a team's success based on data [25]. Machine learning algorithms play a crucial role in extracting patterns and knowledge from these datasets, providing valuable insights for decision-making. These algorithms find applications in sports, assisting in developing training plans, in-game tactical decisions, and more [23].

This study aims to assess graph representations that most effectively capture the complexity of pass difficulty and the metrics that can be derived for this specific purpose. Since a pass occurs in a time interval defined by an initial timestamp and final timestamp, our investigation also delves into the influence of multiscale temporal features in the classification results. Furthermore, we explore whether the combination of features has a discernible impact on accuracy.

1.1 Motivation and Objectives

The unpredictable nature of football poses a contemporary challenge in constructing predictive statistical models for evidence-based decision-making. With the advent of technology and abundant available data, Machine Learning (ML) has emerged as a crucial strategy to aid team staff in decision-making. However, as ML models hinge on the availability of datasets, further studies have been delving into the analysis of the data needed as input for meaningful and accurate predictions [44].

Complex Network measurements provide a suitable avenue for such analyses. Complex Networks have been successfully employed to model match dynamics in various applications, such as analyzing goal-scoring passing networks [35], characterizing player contributions to the team [1], and identifying complex tactical patterns [49]. Typically, players are represented as vertices, and edges denote interactions between them. Rodrigues [46, 47], for example, proposed the 'Opponent-Aware Graph,' representing the interference of opposing players in the possibility of a pass based on player location data. Complex network measures were used to characterize match events and player roles.

Certain studies focus on specific aspects of football, such as Merlin's work [36, 37], which assessed pass difficulty through spatiotemporal variables. The study demonstrated that the density of opposing players near those involved in the pass influences its difficulty. We can conceptualize the interaction/proximity of opposing players as a 2-mode dataset, with two distinct sets of data – the first being players passing the ball, and the second being opposing players. A suitable graph technique for representing this is the bipartite graph, which we evaluate in our study.

To the best of our knowledge, no study in the literature has investigated the use of graph-based representations in the assessment of pass exchange difficulty based on spatiotemporal data. This is a literature gap that we bridge in this work. Our research aims to evaluate graph representations that best describe pass difficulty and the metrics that can be extracted for this purpose. As a pass is an event with an initial timestamp and a final timestamp, our study also assesses the impact of a multilevel temporal analysis on classification. Additionally, we investigate whether the combination of features impacts classification accuracy results.

1.2 Hypothesis and Research Questions

Considering the works available on the literature, a notable gap in studies that integrate both the spatial dynamics of players on the field and the temporal aspects of the game, which dictate the flow of actions during matches. As passing constitutes the most common event in football and embodies player interactions, our study addresses two primary hypotheses: *(1) Temporal graphs and their associated complex-network measurements effectively model pass difficulty classification; and (2) Detailed temporal representations yield improved classification accuracy.* These hypotheses prompt the following investigations:

- **Best Graph Representations:** Exploring the landscape of pass difficulty, one of our central research questions revolves around the identification of the most appropriate graph representations. We seek to unravel which configurations and structures within these representations excel at capturing the intricate patterns and nuances that define pass difficulty in a football match. By addressing this question, we aim to uncover the most effective approaches to transform the dynamic interplay of players and the ball into insightful graph-based models.
- **Impact of Multilevel Temporal Representations:** We aim to gauge the influence of multilevel temporal representations in comparison to binary-level representations. Specifically, we seek to understand how considering a range of temporal resolutions affects the accuracy of our pass difficulty analysis. This question helps us explore the value of time-based features in pass difficulty classification and extends our understanding of temporal dynamics in football gameplay.
- **Feature Fusion:** We explore the synergy of features. We are curious about whether combining effective features can result in enhanced prediction performance.

In summary, we address the following research questions:

1. Which graph representations are the most suitable for encoding patterns related to pass difficulty? This question is associated with the following sub-questions:
    - Which vertex representations are suitable for assessing pass difficulty?
    - Which edge representations are suitable for assessing pass difficulty?
2. Which graph measures are suitable for encoding patterns associated with pass difficulty?
3. Which classification system is suitable to determine the level of pass difficulty?
4. Do multilevel temporal representations contain crucial information that enhances the accuracy of pass difficulty analysis?
5. Would the combination of features improve the pass difficulty classification accuracy and provide deeper insights into the underlying patterns?

1.3 Thesis Organization

In Chapter 2, we introduce the fundamental concepts for the construction of graphs and measures of complex networks that can be extracted from them; an overview of classifiers and statistical learning methods, as well as works related to the topic are also provided. Next, in Chapter 3, we present our proposed framework for classifying passing difficulty through spatiotemporal data. It describes the choices made in the implementation of the framework. Chapter 4 presents a detailed description of the materials used and the methodology employed in the conducted research. The purpose of that chapter is to provide readers with an overview of the approach taken to address the research questions and to ensure the transparency and reproducibility of the study. That chapter details the materials, data collection procedures, and analytical methods. Its content may guide future researchers interested in replicating or extending our study. Chapter 5 presents the results of the study and provides a comprehensive discussion of the findings in light of the research questions. The purpose of that chapter is to present the empirical outcomes and to engage in a meaningful interpretation of the data, thereby addressing the research questions and contributing to the broader body of knowledge in the field. Chapter 6 provides a summary of our main contributions and presents directions for future work.

Chapter 2

Background Concepts and Related Work

This chapter provides the necessary background information and concepts that are relevant to understanding our study. This chapter also reviews previous research or related work in the field, highlighting key findings, methodologies, and gaps or limitations in the existing literature.

2.1 Background Concepts

In this section, we lay the groundwork for our study by introducing three foundational concepts: graphs, complex network measurements, and classifiers. These concepts serve as the cornerstones of our research, providing the essential framework upon which our investigation is built.

2.1.1 Type of Graphs

In graph theory, a graph is a representation of elements (vertex) and their connections (edge). Similar to networking science, networking is a representation of real elements (nodes) and their relations (link). In this section, we present two types of graphs [3].

Classical Graph

Definition 1 *A classical graph* [1]*, denoted as* $G = (V, E)$*, comprises two sets: a set of vertices* V*, and a set of edges,* E*, where* E *is a subset of* $V \times V$*. In this representation, each vertex can be uniquely identified by an integer,* $i = 1, 2, \ldots, N$*, and an undirected edge* $(v_i, v_j) \in E$ *connects vertices* v_i *to* v_j *[34]. For a visual example, refer to Figure 2.1, which illustrates a classical graph representing a football match frame.*

[1] Several studies in the literature refer to one-mode graphs. The term 'modes' pertains to the diversity of referenced entity types.

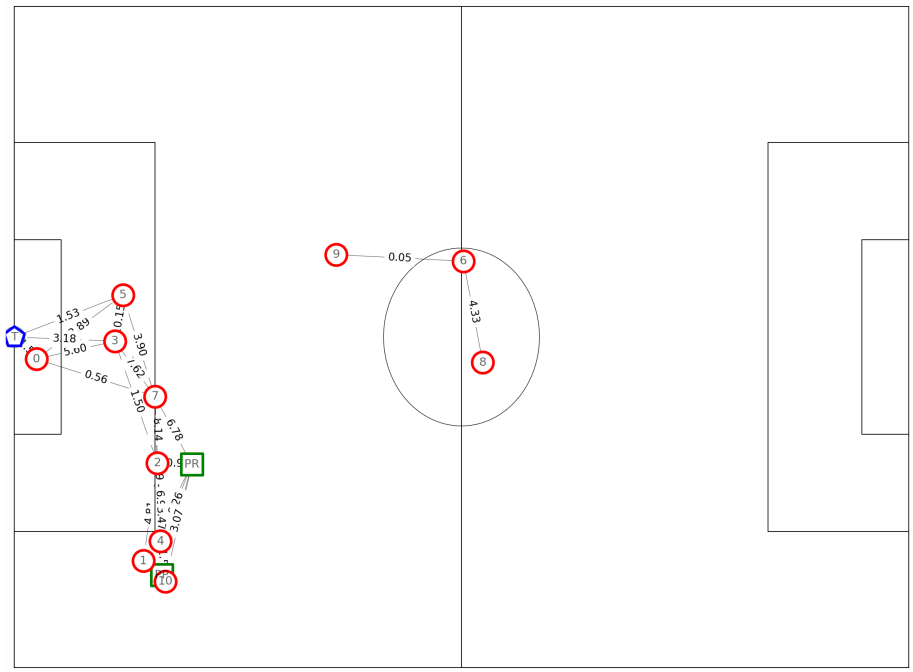


Figure 2.1: This classical network example illustrates a frame from a football match that represents passing actions. In this visualization, green square vertices represent the players executing passes, blue pentagon represent their intended targets, while red circle vertices represent the defensive team. The edges connecting these vertices are weighted based on the calculated distances between the corresponding players, and the weights reflect the spatial relationships among the vertices.

Bipartite Graph

Definition 2 *A bipartite graph* [2]*, typically denoted as* $G = (U, V, E)$*, can be defined by a triplet consisting of two disjoint sets of vertices, denoted as* U *and* V*, and a set of edges,* E*, which forms a subset of the Cartesian product* $U \times V$*. Within each subset* U *and* V*, the vertices can be uniquely identified by an integer* $i = 1, 2, \ldots, N$*. In this context, an undirected edge* $(u_1, v_1) \in E$ *links the vertex* u_1 *with vertex* v_1 *[34]. For a visual example, refer to Figure 2.2, which illustrates a bipartite graph example of a football match frame.*

Weighted Graph

Definition 3 *Let* $G = (V, E)$ *be a weighted graph, where* V *is the set of vertices and* E *is the set of edges. The weight function* $w : E \to \mathbb{R}$ *assigns a real number to each edge in* E*.*

2.1.2 Complex Network Measurements

Complex network measurements aim at characterizing intricate topology, enabling the examination, representation, classification, and modeling of complex systems [15]. In this

[2]Several studies in the literature refer to two-mode graphs. The term 'modes' pertains to the diversity of referenced entity types.

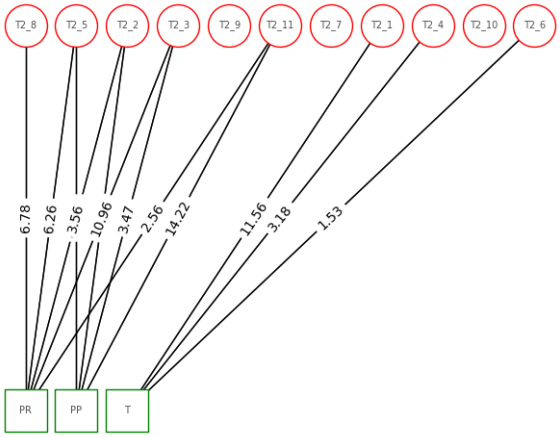


Figure 2.2: This bipartite graph example depicts a pass frame from a football match. In this representation, green square vertices represent the players executing passes and their intended targets, while the red circle vertices represent the defensive team. The edges between them are weighted based on the distances between the respective players, reflecting the spatial relationships in the context of the match.

section, we introduce measurements employed to quantify attributes of complex networks. Table 2.1 presents a comprehensive overview of measurements along with references to the defining equations.

Degree

In an undirected graph, the degree (D) of a vertex quantifies the number of connections that vertex has with other vertices in the graph. Specifically, we use k_i to denote the degree of a vertex v_i, which corresponds to the count of edges connected to v_i [15], which is defined in Equation 2.1:

$$k_i = \sum_{j \in N} (v_i, v_j) = \sum_{j} (v_j, v_i) \tag{2.1}$$

Betweenness Centrality

Betweenness centrality (B) is a metric in graph analysis that quantifies the importance of a vertex within a graph by counting the number of shortest paths between pairs of vertices that traverse through that particular vertex. This measurement offers insights into the vertex's role in maintaining the overall connectivity of the network. The betweenness centrality of a vertex v_i is computed as the sum of the fractions of all-pairs shortest paths that pass through v_i. A $(v_j, v_k) - path$ refers to a sequence of adjacent vertices that starts with v_j and ends with v_k. The length of such a path is defined by the number of edges it contains. In this context, the betweenness centrality of v_i is mathematically defined as the sum of the fractions.

$$c_B(v_i) = \sum_{v_j, v_k \in V} \frac{\sigma(v_j, v_k | v_i)}{\sigma(v_j, v_k)} \tag{2.2}$$

where V represents the set of vertices, $\sigma(v_j, v_k)$ denotes the number of shortest paths between v_j and v_k, and $\sigma(v_j, v_k|v_i)$ represents the count of those paths that traverse a specific vertex v_i. In cases where v_j equals v_k, $\sigma(v_j, v_k)$ is set to 1, and when v_i is a element of v_j, v_k,$\sigma(v_j, v_k|v_i)$ is set to 0 [7].

Closeness

Closeness centrality (C) measures how closely connected a vertex is to all other vertices in the graph. It is defined as the reciprocal of the sum of the shortest path distances from the vertex to all other vertices. A higher closeness score indicates a vertex's proximity to the rest of the vertices within the graph. In the case of disconnected graphs, this measurement is computed considering all vertices in the connected component containing the vertex of interest. The shortest path distance between vertices i and j is represented as d_{ij}, and stands for the length of the shortest path (v_i, v_j)-path [6].

Mathematically, the closeness centrality of a vertex v_i is expressed as the reciprocal of the average shortest path distance to v_i over all $n-1$ reachable vertices, as shown in Equation 2.3.

$$C(v_i) = \frac{n-1}{\sum_{v_j=1}^{n-1} d(v_j, v_i)}, \tag{2.3}$$

In this equation, $d(v_i, v_j)$ represents the shortest path distance between vertices v_i and v_j, and n denotes the total number of vertices that can access vertex v_i. It is important to note that the higher the closeness value, the greater the centrality.

Cluster Coefficient

The Cluster Coefficient (CC) is a measurement of the tendency for vertices in a graph to form clusters. It quantifies the proportion of triangles in which a specific vertex participates, relative to the maximum number of triangles it could be part of, this measurement is expressed at Equation 2.4. However, in the context of bipartite graphs, where vertices are divided into two separate, non-connected sets, and triangles cannot be formed, the computation of the cluster coefficient focuses on connection density [34]. Equations 2.5 and 2.6 provide mathematical expressions for the standard and bipartite-specific (CC_b) calculations of the cluster coefficient, respectively. A higher CC score suggests that a vertex is more likely to be associated with a closely-knit group of vertices within the graph.

$$c_{v_i} = \frac{2T(v_i)}{k_i(k_i-1)}, \tag{2.4}$$

Here, $T(v_i)$ represents the count of triangles that involve vertex v_i, and k_i denotes the degree of vertex v_i.

$$c_{b_{v_i}} = \sum_{v_j \in N(N(v_i))} \frac{c_{v_i v_j}}{|N(N(v_i))|}, \tag{2.5}$$

Table 2.1: Complex network measurements used in our study.

Measurement	Abbreviation	Symbol	Equation
Degree	D	k_i	2.1
Betweenness	B	$c_B(v_i)$	2.2
Closeness	C	$C(v_i)$	2.3
Cluster Coefficient	CC	c_{v_i}	2.4
Cluster Coefficient Bipartite	CC_b	$c_{b_{v_i}}$	2.5
Robins Alexander Clustering	RAC	CC_4	2.7

where $N(N(u))$ refers to the second-order neighbors of vertex u within the graph G, excluding vertex u itself. The term c_{uv} represents the pairwise clustering coefficient between vertices u and v. This coefficient can take on various forms, including:

$$dot : c_{v_i v_j} = \frac{|(N(v_i) \cap N(v_j)|}{|(N(v_i) \cup N(v_j)|} \tag{2.6}$$

Robins Alexander Clustering

Robins Alexander Clustering (RAC) is a bipartite measurement akin to the Cluster Coefficient (CC), but it places a distinct emphasis on detecting the presence of four-vertex cycles within bipartite graphs. Similar to CC, RAC quantifies the tendency of vertices in the graph to cluster together, but it specifically highlights a particular clustering pattern. The bipartite clustering coefficient, as defined by Robins and Alexander, is computed as four times the number of four-cycles (C4) divided by the number of three-paths (L3) in a bipartite graph [45]:

$$CC_4 = \frac{4 \times C_4}{L_3} \tag{2.7}$$

2.1.3 Classifiers

In this section, we outline the classifiers employed in our analysis, selected for benchmarking against the work by Merlin [36]. A classifier is a machine learning algorithm that is trained on a set of labeled data and used to make predictions on new unlabeled data. In our study, we compare the performance of common classifiers. Each classifier has its own strengths and weaknesses, and by evaluating their performance in our problem, we can gain insight into which classifiers are most effective for this task. We provide a description of each classifier. Table 2.2 shows classifiers' characteristics.

K-Neighbors Nearest - KNN

The KNN algorithm is a machine-learning technique used for classification and regression tasks. In KNN classification, the algorithm assigns a label to a new unlabeled data

point based on the labels of its K nearest neighbors in the training dataset. While KNN is easy to implement and often effective, it can be impacted by noisy and irrelevant input features [14]. Despite these limitations, KNN remains a popular choice for many classification problems due to its simplicity and nonlinear decision space [27].

Linear Discriminant Analysis - LDA

The LDA algorithm is designed to identify a linear combination of input features that best separates different classes. LDA can also be used for dimensionality reduction. However, its performance may be affected when the normality and equal covariance assumptions are not met. Additionally, noisy or irrelevant input features can increase the within-class variance, leading to classification errors. Despite these limitations, LDA produces simple decision boundaries that can be easily described and implemented. To classify a new observation, LDA calculates the Mahalanobis distance to the class centroids and assigns the observation to the closest one using a pooled covariance estimate [29].

Logistic Regression - LR

LR is a method for modeling the probability of a categorical outcome as a function of input features. One of the strengths of LR is its robustness to noisy or irrelevant input features, making it particularly useful for high-dimensional datasets. However, it assumes a linear relationship between the input features and the log odds, which may not hold for complex datasets with non-linear relationships. The LR algorithm estimates the probabilities of a sample belonging to a class, which must be continuous and bounded between 0 and 1. The name "logistic" is derived from the sigmoid (or logistic) function, Equation 2.8, used to map the input features to the output probabilities [5].

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{2.8}$$

Gaussian Naïve Byes - NB

NB is a probabilistic classification algorithm on Bayes' theorem, which supposes the conditional independence between input features. NB assumes that the probability of a certain class given an input feature vector can be calculated by multiplying the probabilities of each feature given the class. The NB classifier can be relatively robust to noisy and irrelevant input features, but the assumption of conditional independence between features given the class is its main drawback since such an assumption is often not true in real-world scenarios. NB estimates the class label using only the marginal and conditional probabilities of each feature. However, NB may not perform well on datasets where the input features are continuous or where the relationship between the input features and the class labels is nonlinear [5].

Random Forest - RF

RF is an ensemble classifier that uses many decision tree models and combines their predictions using majority voting to produce a single output. The algorithm is based

Table 2.2: Characteristics of various learning methods, where ✓represents effectiveness, ◇ indicates moderate performance, and - signifies lower efficacy [29].

Characteristic	KNN	LDA	LR	NB	RF	SVM (L2)
Resilience to outliers in input space	✓	-	✓	✓	✓	✓
Dealing with irrelevant inputs	-	-	✓	✓	✓	✓
Linear feature extraction capability	◇	✓	✓	-	-	✓
Interpretability	-	✓	✓	✓	✓	◇

on the concept of decision trees, where a tree is built by recursively splitting the data into subsets based on the values of the input features. The random feature selection process means that RF can still perform well even when some input features are noisy or irrelevant, as the irrelevant features are likely to be excluded in at least some of the decision trees. Additionally, the aggregation of multiple decision trees can help to reduce the impact of noisy features by averaging their effect on the final prediction. This makes RF a highly effective algorithm for classification tasks [38].

Support Vector Machine - SVM

The SVM algorithm uses a kernel to transform the original data into a higher dimensional space, where a hyperplane that optimally separates the data into two categories is found. SVM is capable of handling both linearly and non-linearly separable datasets through the use of kernel functions. However, SVM can be sensitive to noisy and irrelevant input features, which can affect its performance. To overcome this challenge, regularization techniques, such as L1 or L2, can be applied to penalize the model for using noisy or irrelevant input features [38].

2.2 Related Work

In this section, we survey and analyze prior studies that are relevant to our investigation into pass difficulty analysis. The goal is to provide readers with an overview of the state of the field, identify gaps, and position our work within the broader academic discourse.

2.2.1 Graph-based Representations in Football Analysis

Raabe et al. [42] introduced a graph representation that encodes each player from both teams as a vertex, incorporating edges weighted by player distances. Their use of Graph Convolutional Networks (GCN) facilitated binary classification for defensive success, distinguishing between possession wins and losses. Rodrigues et al. [47,49] proposed a unique approach known as Graph Visual Rhythms for temporal network analyses. Their method extracts vertex measurements over time and visually represents player rhythms through color-mapped plots. In a different context, Caetano et al. [10] applied graph analysis

to characterize interpersonal coordination among opponent players during offensive sequences in official matches. Their work aimed to discern coordination patterns in sequences leading to shots on goal compared to those ending in defensive tackles. Xenopoulos et al. [51] leveraged Graph Neural Networks (GNN) to predict sports outcomes. They constructed a fully connected graph in which vertices represent players and edges denote player distances. Their model demonstrated the potential of GNNs in the sports prediction domain. Brandt and Brefeld [8] presented a graph-based approach to football, treating passes as edges between player nodes. Through empirical comparisons, they evaluated various variants, including the PageRank algorithm, on football data from Bundesliga seasons. Transforming the approach into a predictor for winning teams, their results showed remarkable accuracies of over 57% using only three features. Vago et al. [50] introduced a novel tool for analyzing change patterns based on graphs in simulated football matches, providing a valuable resource for understanding dynamic aspects of gameplay. None of those studies focused on pass analysis.

2.2.2 Pass Analysis

Passes represent one of the most frequent events in football. For instance, during the 2014 World Cup, most teams executed over 500 passes in a single game, and in the 2010 World Cup, successful teams significantly engaged in more passes [33]. Statistical data related to passes are regarded as crucial predictors of success. These findings emphasize the importance of evaluating and enhancing passing strategies [24].

Despite this, traditional statistics, such as the number of assists, key passes, or successful pass completion percentage fail to consider the context in which passes are executed. Sophisticated analytical tools capable of dissecting vast amounts of data allow us to study the intricate dynamics of space creation and construction in football [9], a fundamental element related to passing strategies. Disrupting the opponent's defense stands as a crucial strategy in elite football [19].

Mendes et al. [16] conducted a thorough evaluation of the association between goal kick strategies and offensive outcomes. Their findings revealed a compelling statistic, indicating that long goal kicks presented a 64% chance of resulting in a successful offensive sequence.

In a comprehensive study, Kapsalis et al. [32] delved into the influence of various performance indicators on teams' goal-scoring abilities. The research underscored the significance of possession-oriented gameplay and well-executed progressive passes in achieving success in football. Notably, teams relying on long passes against well-organized defenses tended to be less effective in scoring goals. Conversely, teams with higher possession rates and a tendency for penetrating passes beyond the first 40 meters from their own goal exhibited a greater ability to score goals.

Goh et al. [26] investigated the most frequent methods of ball repossession, distribution, and movement patterns leading to goals scored in open play in football. Their findings highlighted the significance of shorter passing sequences, with 83.8% of goals resulting from such strategic plays.

In a unique approach, Jong et al. [17] explored teamwork using network analysis, com-

paring match outcomes, types, ladder halves, and tournament phases. Their observation indicated that successful league teams tend to have more players with connecting roles compared to tournament teams, emphasizing the need for adaptive match tactics.

Praça et al. [39], in turn, delved into the influence of match-related variables on teams' winning probabilities during the 2021/2022 Bundesliga. The results indicated that winning teams exhibited higher rates of accelerations, ball possessions, passes, successful passes, and shots compared to losing teams.

Merlin [36, 37] presented an approach for the automatic classification of football pass difficulty, discussing key spatiotemporal variables characterizing pass complexity. His research involved data from 465 passes, from which 32 independent variables were extracted. Pass difficulty was classified as low (56.5%), medium (22.6%), and high (20.9%). A significant conclusion drawn from the study was that high-difficulty passes are characterized by greater pressure on the receiving player. Passes were defined as vectors AB originating from the passing player at the moment and position the player passes the ball PP(t0) (A) and ending at the moment the receiving player touches the ball RP(t1) (B)[3][4], where the receiving player can be an opponent or not. Figure 3.1 illustrates the two moments of the passes. Merlin's research evaluated variables, such as the distance between the passing player and the nearest opponent, the number of opponents at distances of 1m, 2m, 5m, and 10m from the passing player, the passing player's speed, the distance between the receiving player and the nearest opponent, the number of opponents at distances of 1m, 2m, 5m, and 10m, pass distance, ball speed, and more. The study identified key variables for pass difficulty classification, including pressure on the receiver (5m, 10m) during the pass, distance to the nearest opponent, and pressure on the receiver (2m, 5m, and 10m) upon the ball's arrival.

To the best of our knowledge, no study in the literature has investigated the use of graph representations and machine learning methods for pass difficulty classification, a literature gap that this thesis bridges.

[3]In this study, we adopt the notation t_i and t_f, in place of t0 and t1 as utilized by Merlin, to denote the initial and final moments of passing.

[4]In this study, we employ the abbreviation PR, as opposed to Merlin's usage of RP, to represent the Passing Receiver Player.

Chapter 3

A Framework for Graph-based Pass Difficult Analysis

This chapter introduces the proposed framework for graph-based pass difficulty analysis. The components of the framework are presented and the implementation choices considered in this work are discussed in this chapter.

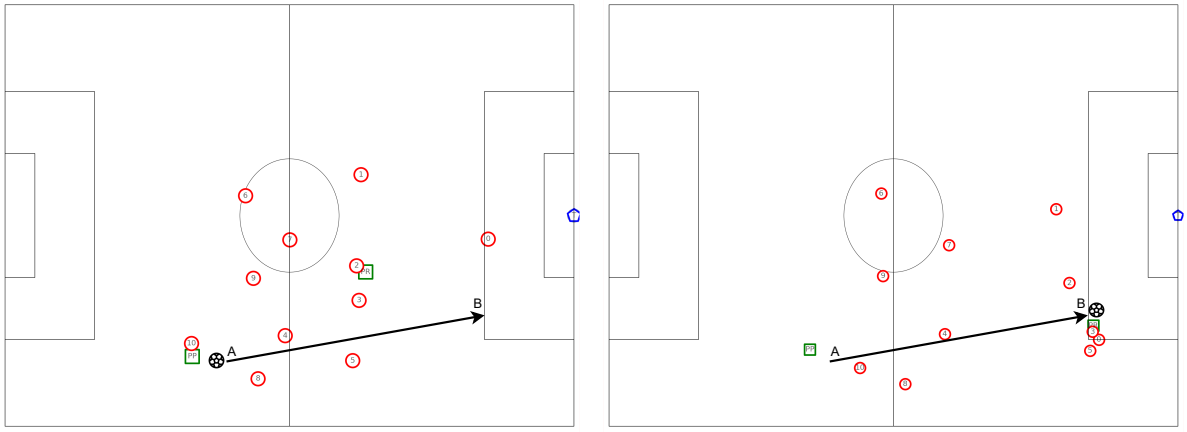
3.1 Introduction

Football is a collective sport characterized by spatiotemporal events at various scales, where player interactions can determine the success or failure of actions. Modeling and representing these interactions for performance analysis is a complex task [2,41]. Complex Network measurements are a valuable tool for such analysis, as they have been successfully applied to model match dynamics in various applications. Examples include analyzing goal-scoring passing networks [35], characterizing player contributions to the team [1], and identifying complex tactical patterns [49]. Passes are a prominent event in football [24], with successful teams executing significantly more passes [33]. Key pass statistics, such as the ratio of short/long passes [40] and passing accuracy [13], have been identified as predictors of success [24], underscoring the importance of evaluating and improving passing strategies for players and teams [24].

A pass can be represented as a vector $\overrightarrow{AB}$ connecting the position A of the passing player (PP) at timestamp t_i – PP_{t_i}(A) – to the position B of the passing receiver (PR) at timestamp t_f – PR_{t_f}(B). Within this context, t_i refers to the moment when the passing player touches the ball, while t_f is the moment when the receiver player contacts the ball [36]. This definition does not consider the pass as an instantaneous event, restricted to a single frame, but as a sequence of frames. Figure 3.1 illustrates examples of players performing passes, in which green squares represent the PP and PR, red circles represent players of the opposing team, and a blue pentagon represents the target.

According to Merlin [36], the main tactical variables that explain pass classification are, in order of relevance:

- the number of opponents between PRt_f and the target;



(a) t_i (b) t_f

Figure 3.1: Illustration of a football pass, the vector $\overrightarrow{AB}$ has the origin A at the Passing Player (PP) position at t_i and destination B at expected Passing Receiver (PR) position at timestamp t_f. (a) t_i is the initial instant when the PP touches the ball, while (b) t_f is the final instant when the Real passing Receiver (RR) has contact with the ball, who may be a PP teammate (successful pass), or a player of the opposing team (unsuccessful pass).

- player density on a ring of radius 5m around PRt_i;
- number of outplayed opponents;
- player density on a ring of radius 5m around PRt_f;
- nearest opponent to PRt_f;
- nearest opponent to PRt_i;
- ball progress;
- player density on a ring of 2m around PRt_f;
- player density on a ring of radius 10m around PRt_f;
- the velocity of PRt_f;
- player density on a ring of radius 10m around PRt_i;
- displacement of PR; and
- distance PRt_f from the target.

In his study, Merlin [36] showed the importance of the proximity between opponent players and their evolution over time for the pass difficulty classification. His findings motivated us to model such interactions as a complex network that could be explored toward characterizing the pass difficulty. In this context, the proximity between players can be modeled in the definition of edges, which leads to several formulations. In this

work, we use a distance threshold between players by taking into account their (x, y) coordinates on the football field, and we investigate methods based on the Delaunay triangulation. Besides, we analyze different graph topologies based on bipartite and one-mode graphs. Graph modeling allows complex network measurement calculations and quantitative analyses of players' interactions during passing events.

This chapter is organized as follows. Section 3.2 introduces the graph-based pass difficult analysis framework. Next, Section 3.3 describes the different graph models analyzed. Next, Section 3.4 discusses the complex network measurements extracted. Then, Section 3.5 details the feature vectors extracted. Finally, Section 3.6 presents the classification step.

3.2 Overview

With the aim of supporting pass difficulty analysis, we propose the analysis framework presented in Figure 3.2. The framework comprises three main steps: (a) Graph Extraction, (b) Complex Network Measurement Computation, and (c) Classification. In the following, each component is described.

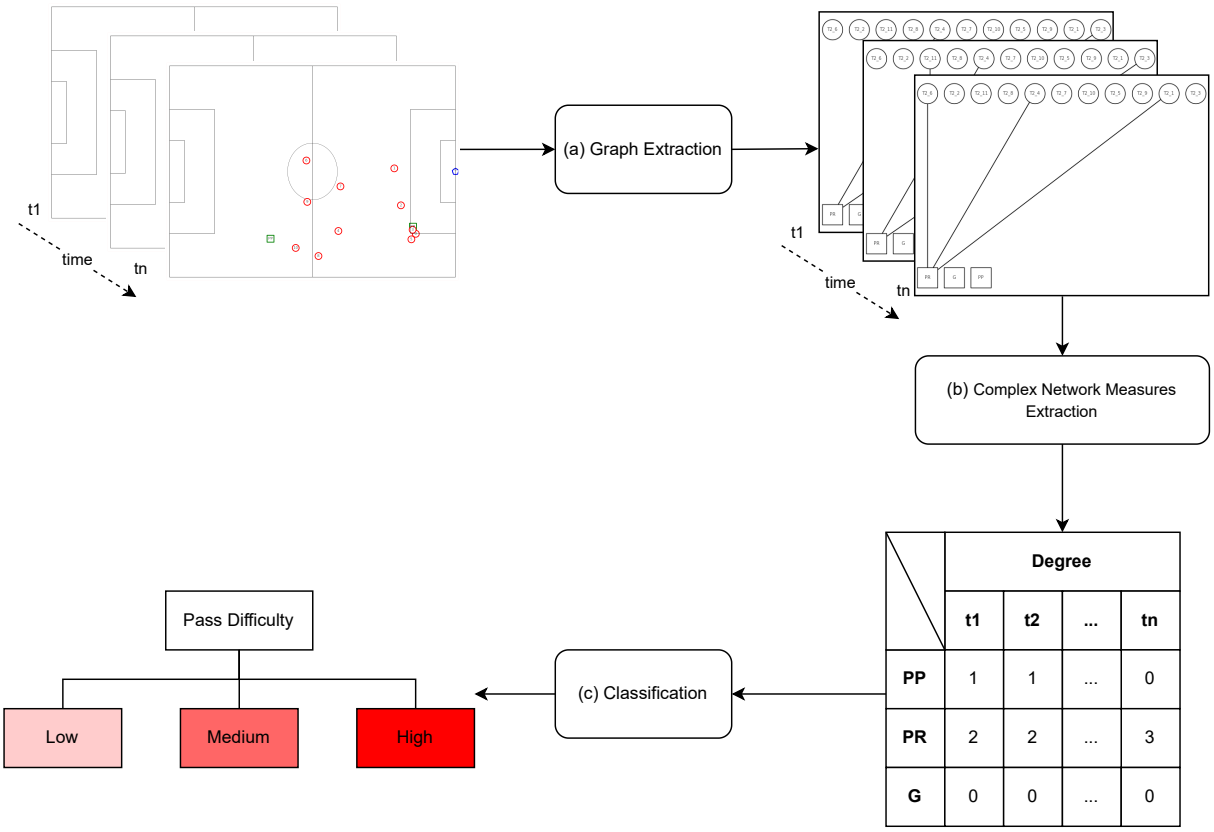


Figure 3.2: Framework for Pass Classification composed of three steps: (a) Graph construction based on players' positions on the pitch for frames in the interval (t_1, t_n). (b) Computation of complex network measurements and feature organization into a matrix, where each row represents a player and each column represents a moment in time. (c) Classification process that receives features passing as input and determines the difficulty category (low, medium, and high).

(a) Graph Extraction: This step concerns building graphs that have vertices representing players and edges based on the players' proximity (defined in terms of their locations and distance). The input is (x, y) coordinates of players on the football field. The different formulations explored for the definition of graph vertices and edges are described in Section 3.3. Furthermore, bipartite and one-mode graph types were analyzed.

(b) Complex Network Measurement Computation: This step refers to the computation of complex network measurements from graphs obtained. (b). The measurements are extracted from the most relevant vertices, i.e., the ones representing players involved in the passing event, and the target. Later, the measurements are used to construct feature vectors as detailed in Section 3.5. Our study uses complex network measurements introduced in Section 2.1.2 and characterizes the vertex centrality.

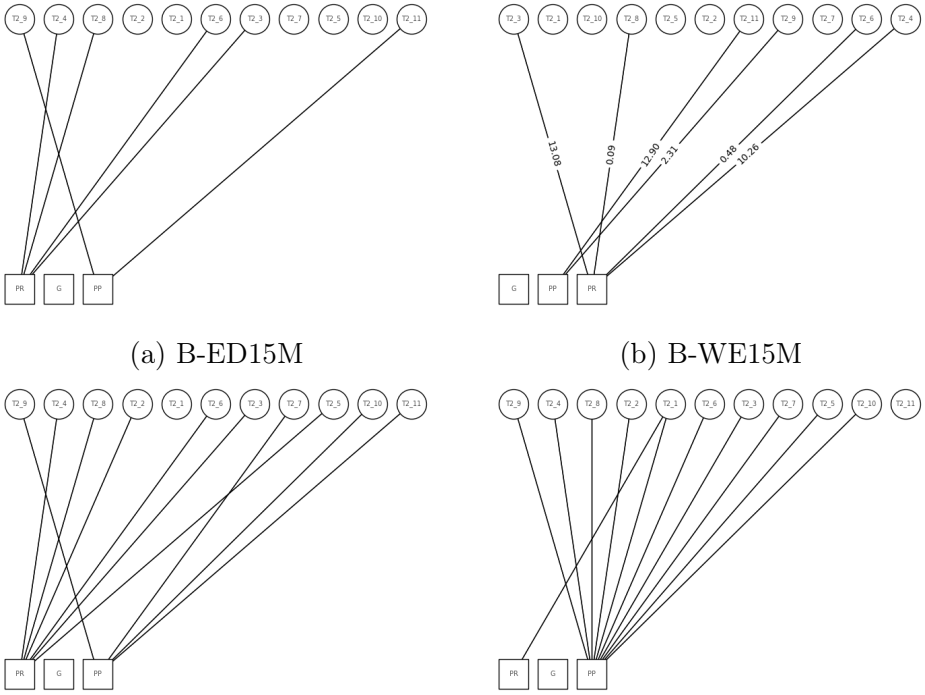
(c) Classification: In this step, the feature vectors computed in step (b) are combined with different classifiers aiming at the creation of a classification system for the pass difficulty problem. Section 2.1.3 describes the classifiers considered in our study.

3.3 Graph Representation

The definition of vertices and edges may lead to different graph configurations, which encode different players' spatial distribution properties. This work investigates which graph representation would be suitable to encode the football pass difficulty. In the following, we present the different alternatives for constructing graphs considered in our study.

Table 3.1: Overview of vertex representations.

Vertex Type	**Name**	**Definition**
A	Attackers	Vertices refer to players of a single team.
O	Opponent-Aware	Vertices refer to the two players involved in the passing event, all opponents and target (goal).
B	Attacker-Opponent Bipartite	Bipartite graphs are used to model disjoint sets of vertices U and V, where U has vertices that represent the players that perform the pass or the target and V represents the opponents.



(a) B-ED15M (b) B-WE15M

(c) B-xD15M (d) B-OI

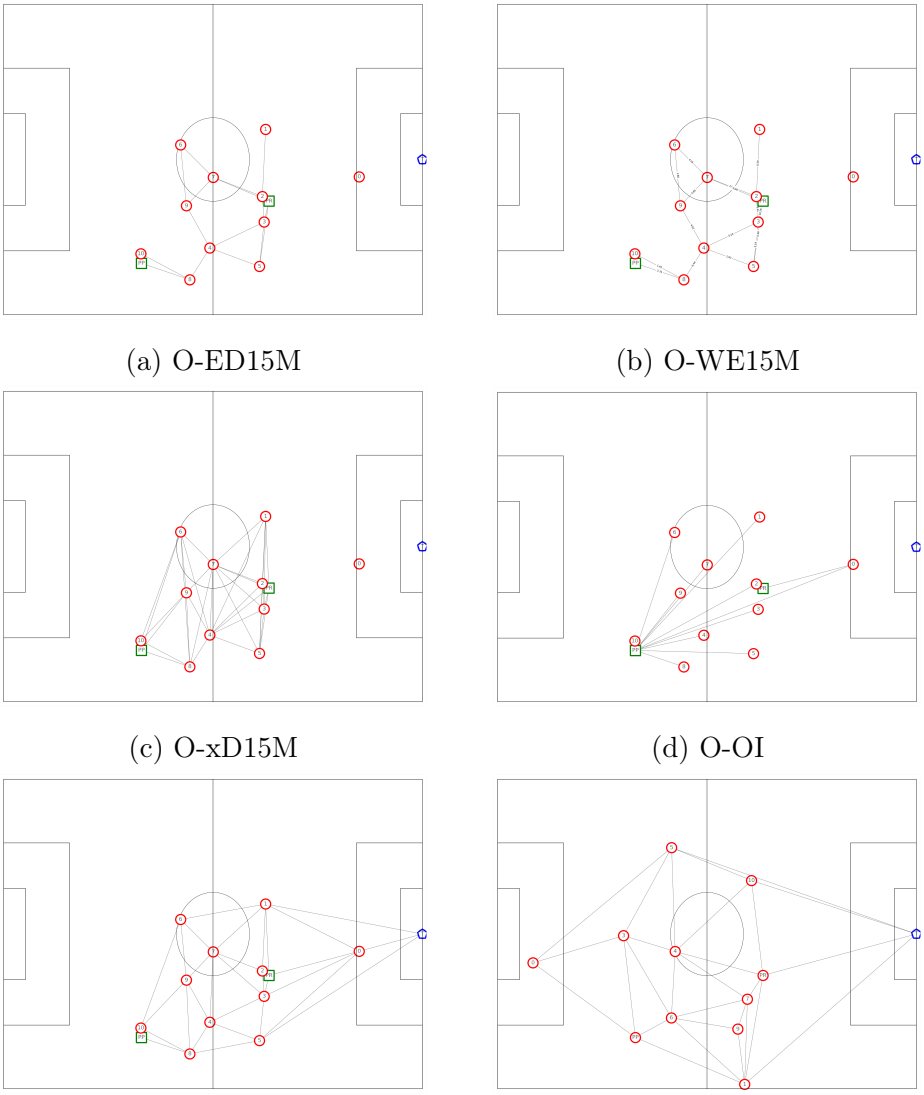
Figure 3.3: Graph representations constructed from the same frame using bipartite vertices and different edge definitions. Square vertices represent passing players and their target, while circle vertices represent opponent players.

Vertex

In graph theory, vertices represent entities or objects that may interact with each other. In our models, we represent football players as vertices. We introduce the specific graph formulations that we consider in the following.

Attackers (A): Modeling graphs in which vertices refer to players of a single team allows encoding their technical-tactical interaction. In particular, we investigate patterns encoded in graphs defined in terms of how attacking players interact with each other. We believe that encoding how players are involved with offensive actions may provide relevant information for characterizing pass exchanges that aim to create opportunities for scoring by encoding how players are involved in offensive actions.

Opponent-Aware (O): The presence of opponents around players that perform passes impacts the pass difficulty [36]. Therefore, modeling graphs that represent the interaction of opponent players can reveal relevant tactical features. Since this study focuses on pass difficulty analysis, we explore the interaction between players involved with a passing action and their opponents. The graph $G = (V, E)$ is defined in such a way that the vertices belonging to set V refer to the two players involved in the passing event and all opponents. Set V also contains a special vertex associated with the target (goal), the vertex highlighted as a blue pentagon in Figure 3.4.



(a) O-ED15M (b) O-WE15M

(c) O-xD15M (d) O-OI

(e) O-Dl (f) A-Dl

Figure 3.4: Graph representations constructed from the same frame using different edge and vertex definitions. In the Opponent-Aware (O) graphs, green square vertices represent passing players, red circle vertices represent opponent players, and blue pentagon vertices represents the target of passing players. In the Attackers (A) graphs, red vertices represent players on passing team, and blue vertex represent their target.

Attacker-Opponent Bipartite (B): Bipartite graphs are used to model disjoint sets of vertices [3]. In this formulation, we consider two independent sets of vertices to represent each team, and a set of edges between sets of vertices to model inter-team interactions. More precisely, the formulation refers to a graph bipartite $G = (U, V, E)$ where U has vertices that represent the players that perform the pass or the target and V represents the opponents. The set of edges E is defined as $E \subseteq U \times V$. In the context of team sports, bipartite graphs are used to represent each team as a set of vertices and inter-team interactions as a set of edges.

Edges

Edges represent relationships among vertices (e.g., interaction among players). Many definitions can be used in football analysis, such as if players interact in an event and if they are in the same football field sector. In the passing difficulty assessment, we consider the following formulations for edges:

Table 3.2: Overview of edge representations.

Edge Type	**Name**	**Definition**
ED	Euclidean Distance	A threshold based on the Euclidean distance among players is used to define if two vertices are connected or not.
WE	Weighted Edge	The edge weight is defined in terms of the Euclidean distance among players. A threshold is used to define if two vertices are connected or not.
xD	Axis-x Distance	A threshold based on the distance in the Axis-x is used to define if two vertices are connected or not.
OI	Opponent Interference	Edges are connected if they represent opponents who, along the x axis, are in front of the opponent target (goal).
Dl	Delaunay	Edges form triangles that do not contain vertex in their interior.

Euclidean Distance - ED: Number of opponents within a radius of 2m, 5m, 10m, and 15m to players that perform the pass are some more important variables to determine the passing difficulty [36]. One way to model players' proximity relies on defining a distance threshold, used to determine if two vertices are connected or not. More formally, a graph $G = (V, E)$ has edges $(v_i, v_j) \in E$ linking vertices v_i and v_j

if $d_{ij} < T$, where d_{ij} is the Euclidean distance (Equation 3.1) of players i and j in the field and T is a threshold. This representation provides insight into how players tend to cluster together in the field, and how their movements may be influenced by nearby opponents or teammates.

$$d_{ij} = \sqrt{(v_{i_x} - v_{j_x})^2 + (v_{i_y} - v_{j_y})^2} \tag{3.1}$$

Weighted Edge - WE: In a weighted graph formulation, weights are assigned to edges to represent costs, length, or any other magnitude and units of measurement that can be used for modeling the problem. In our study, we model a graph in which the distance between players is assigned as an attribute of the edges linking them. A graph $G = (V, E)$ contains edges $(v_i, v_j) \in E$ connecting vertices v_i and v_j, if $d_{ij} < T$, where d_{ij} is the Euclidean distance between players i and j in the field, T is a threshold defined. The weight function is defined as $w_{ij} = T - d_{ij}$. This allows us to capture the relative strength of connections between players in our analysis. By considering the distance between players as edge weights, we can better understand the importance of certain player interactions in the game.

Axis-x Distance - xD: The football field length is usually represented in axis x and denotes the attacking direction. This axis has information about the team's progress toward the opponent's goal. In this formulation, edges are defined in terms of the x coordinate associated with players. Formally, the graph $G = (V, E)$ contains edges if $(v_i, v_j) \in E$ links vertices v_i and v_j, and $dx_{ij} < T$, where dx_{ij} is the distance in axis x of players i and j in the field and T is a threshold. By focusing on the directional relationships, we can better understand how players work together to move the ball up the field and create scoring opportunities.

Opponent Interference - OI: The number of opponents between the target and the passing receiver player at t_f in relation to the x axis has been pointed out as one the most relevant variables to characterize the pass difficulty [36]. In our formulation, the graph $G = (V, E)$ contains an edge if $(v_i, v_j) \in E$ if $p_i < p_j < p_t$ or $p_t < p_j < p_i$ where p_i is the position of the attacking player i along the x axis, p_j is the position of the opponent player j position along the x axis, and p_t is the position of the target in the x axis. By analyzing these interference patterns, we can insights into how the opposing team may attempt to disrupt the flow of the game and limit the effectiveness of key players.

Delaunay - Dl: Delaunay triangulation defines a planar graph that has vertices with specific positions. Graph edges form triangles that do not contain vertex in their interior. In the football context, a Delaunay graph encodes information about the proximity of players. A graph $G = (V, E)$ contains edges if $(v_i, v_j) \in E$ are defined by computing the Delaunay triangulation. By using the Delaunay triangulation, we can better understand the patterns of player movement and positioning during the game.

We standardize the nomenclature to refer to different configuration setups considered in our experiments. Thus, a graph configuration is defined as **V-EEDDD-M**, where V indicates the vertex type, E refers to the criterion used to define edges, DDD indicates the threshold distance (but not necessarily present), and M indicates the complex network measurements. Examples of the use of those patterns are presented in Figure 3.4 and Figure 3.3.

3.4 Complex Network Measurement Extraction

We compare six complex network measurements for vertices that represent the players that perform the pass and the target for each instant of time. The measurements are described in Section 2.1.2. We aim to characterize players' connectivity and influence at passing events, as well as the dynamics of the network. Table 3.3 presents examples of complex network measurements from the graph presented in Figure 3.5.

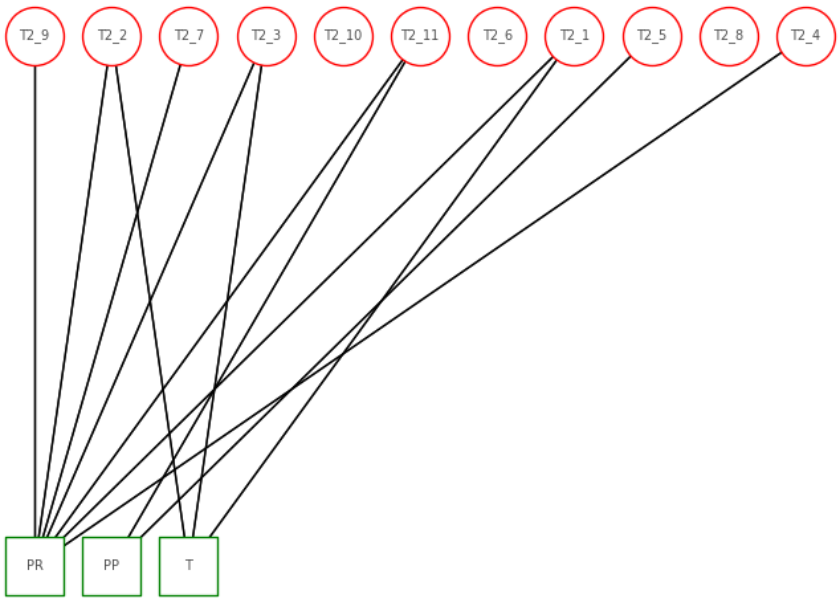


Figure 3.5: Bipartite graph, B-ED15M, extracted from t_i from pass event with difficulty high. The green square on the bottom represents the players that perform the pass (PR and PP) and the target (T). The red circles on the top represent players of the opponent team (T2_i, with $i \in [2, 11]$).

3.5 Feature Vector Extraction

In our formulation, passing events will be classified according to feature vectors that contain complex network measurements extracted from passing frames over time. The feature vector will then be used as the input of classifiers. In our study, the vertices associated with the Passing Player (PP), Passing Receiver (PR), and Target (T) are of interest. Feature vectors are constructed by employing different formulations depending on the

Table 3.3: Extracted Complex Network Measurements from the Graph in Figure 3.5. Degree (D), Betweenness (B), Closeness (C), Cluster Coefficient Bipartite (CC_b), and Robins Alexander Clustering (RAC).

Vertex	**D**	**B**	**C**	**CC_b**	**RAC**
PP	2	0.1184	0.4438	0.1250	0.3871
PR	7	0.4934	0.8242	0.2768	0.3871
Target	3	0.0197	0.4438	0.4286	0.3871

number of timestamps considered and the use of fusion schemes. Those formulations are presented next. Table 3.4 presents Tactical Statistic Variable with graph configurations similarity.

Binary Temporal Scale The feature vector in a binary temporal scale considers complex network measurements related to passing player (PP), passing receiver (PR), and target (T) at timestamps t_i and t_f. A single complex network measurement is computed for each vertex. The final vector will, therefore, comprise six features, two for each vertex.

Multiple Temporal Scales In this formulation, we encode into the feature vector properties of graphs related to *all* timestamps between t_i and t_f. The goal is to represent the temporal evolution of the graphs during a passing event. Figure 3.6 illustrates the feature extraction process.

Feature Fusion Our study also assesses the effectiveness of classifiers when multiple feature vectors are combined. We selected the features of the best multiple temporal scales. Those features are concatenated into a new vector.

3.6 Classification

In this study, we evaluate six different supervised classification algorithms, which are described in detail in Section 2.1.3. The objective of supervised classification is to estimate the function f using a set of observed points consisting of a sample vector $X = (X_1, X_2, \ldots, X_n)$ and a corresponding variable Y [31]. The classification process involves the following steps:

Pre-processing: To evaluate the performance of our machine learning model, we divided the labeled dataset into two subsets: a training set comprising (75%) of the data, and a test set comprising the remaining (25%). To ensure that the test set is representative of the entire dataset, we used the shuffle split technique to randomly shuffle the data before splitting it into training and test sets. We also utilized a seed value to ensure that the shuffling process generates the same set of splits every

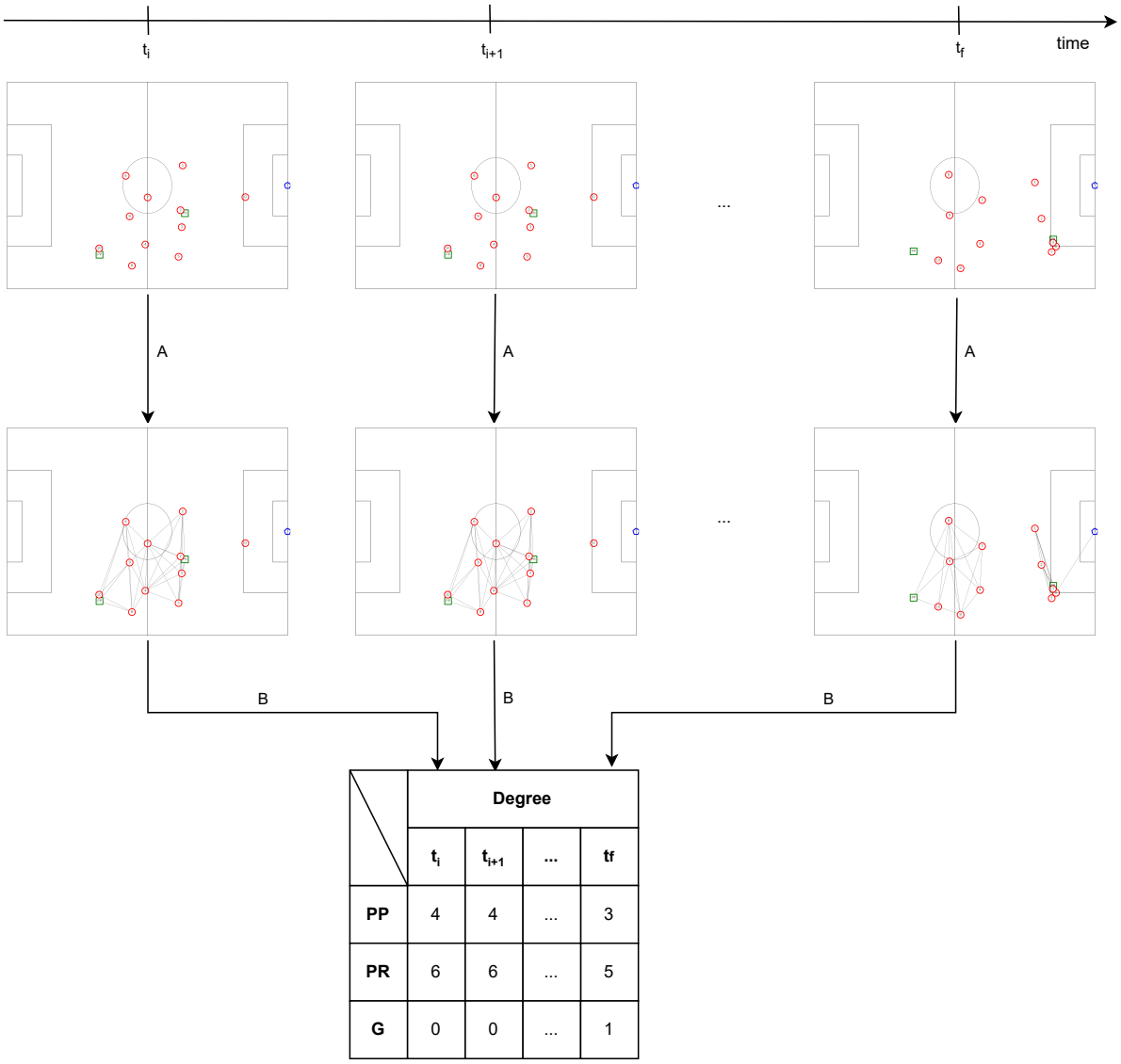


Figure 3.6: Flowchart illustrating how complex network measurements are computed over time and encoded into feature vectors. This flow comprises two steps: (A) Graph construction based on players' positions on the pitch for each frame on time interval (t_i, t_f). (B) Computation of complex network measurements and feature vector encoding. Each row represents a player and each column, a moment in time.

time, which is important for reproducibility and for comparing the performance of different models using the same test set. This approach helps ensure that our results are not dependent on any specific subset of the training data and that our model has generalized well to the entire training set.

Standardization is a widely used preprocessing technique in machine learning, especially for algorithms that are sensitive to the scale of the input features. Standardizing the features ensures that they are on a similar scale and that no single feature dominates the learning algorithm. To achieve this, we applied the standard scaler preprocessing technique, which transforms the features of a dataset to have a mean of zero and a standard deviation of one, as shown in Equation 3.2.

Table 3.4: Tactical Statistic Variable with graph configurations similarity. We use degree measurement because is the measurement that better and directly references the variables

Tactial Statistic Variable	**Graph Configuration**
Opponents between PR t_f and target	OI-D
Density (5m) PR t_i	ED5M-D
Outplayed opponents	-
Density (5m) PR t_f	ED5M-D
Nearest opponent PR t_f	WE15M-D
Nearest opponent PR t_i	WE15M-D
Ball progress	-
Density (2m) PR t_f	ED2M-D
Density (10m) PR t_f	ED10M-D
Velocity PR t_f	-
Density (10m) PR t_i	ED10M-D
Displacement PR	-
Distance PR t1 to target	-

$$x_{scaled} = \frac{x - \mu}{\sigma}, \tag{3.2}$$

where x is a feature, μ is the mean of the feature values, σ is the standard deviation of the feature values, and x_{scaled} is the standardized feature.

Training: We fitted the function f using the training subset, where the input vector X consists of complex network measurements extracted from pass events graphs, while the categorical output Y is defined by labels: low, medium, and high difficulty. The objective of the training classification process is to reduce the error ϵ as expressed in Equation 3.3, by employing statistical learning techniques that take into account the relationship between the input features and the output labels. The objective is to obtain a model that generalizes well to unseen data, i.e., one that performs accurately on the validation and test subsets.

$$Y = f(X) + \epsilon \tag{3.3}$$

where ϵ is the error resulting of f inaccuracy.

To tune the hyperparameters of our machine learning models, we used the Grid Search technique, which exhaustively searches a specified parameter grid to find the best combination of hyperparameters for a given estimator. To ensure the validity of our results and mitigate overfitting, we used the cross-validation scheme, which assesses the performance of a model by training and validating it on multiple subsets of the data. Specifically, we employed 5-fold cross-validation, where each fold consists of a validation set of 10% of the data and the remaining four equally

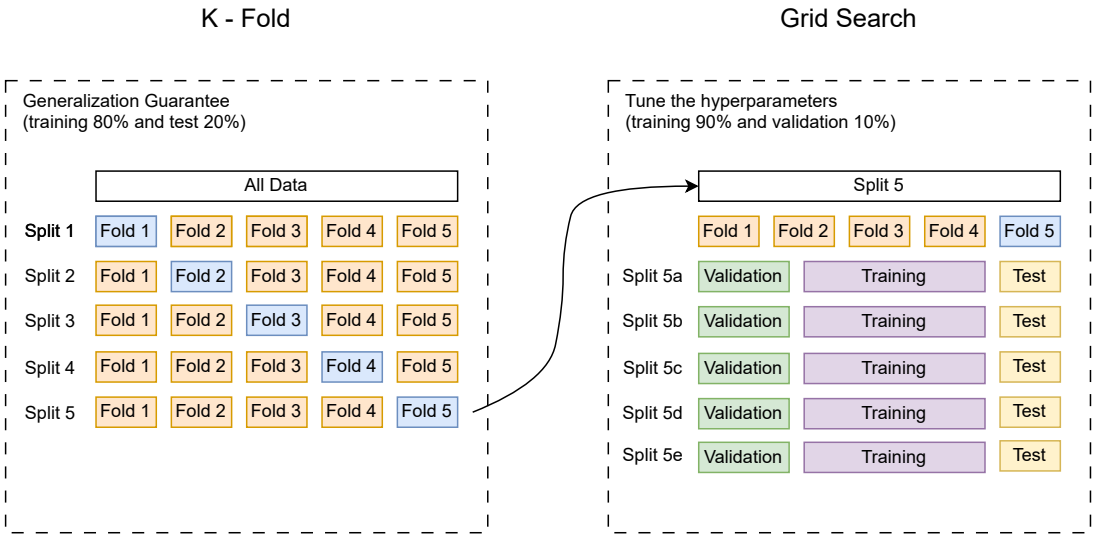


Figure 3.7: The entire dataset was partitioned into 5 folds and repeated 5 times. Each iteration designated one fold as the test dataset. The partitioned folds underwent a grid search process, dividing the training dataset into 10% and 90% subsets. The larger subset was utilized for training and hyperparameter validation of the classifiers.

sized folds for training. The shuffle split technique was used to generate these folds, randomly shuffling the data and splitting it into 5-folds. We also ensured the reproducibility and comparability of results by using a seed value that guarantees the same set of splits is generated every time the algorithm is run. Table 3.5 shows hyperparameters variation for each classifier and Figure 3.7 illustrates the K-fold cross-validation and grid search process.

Prediction: The training process provides an estimate $\hat{f}$ for f, allowing us to predict $\hat{Y}$ using Equation 3.4.

$$\hat{Y} = \hat{f}(X) \tag{3.4}$$

Table 3.5: Hyperparameter variations employed in grid search process for each classifier. K-Neighbors Nearest (KNN), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Gaussian Naïve Byes (NB). Inverse of regularization strength (C). Weight is the function used for prediction consists of two approaches: uniform weighting, where all points within each neighborhood are assigned equal weights, and distance weighting, which assigns weights to points based on the inverse of their distances.

Classifiers	**Hyperparameters**
KNN	Number of Neighbors: 3, 5, 7, 9, 11, 13, 15 Weight: uniform, distance Metric: Euclidean, manhattan
LR	C: $10^{-4} \leq x \leq 10^{4}$, with 15 steps Penalty: L1, L2 multi-class: one-vs-rest, multinominal
RF	Number estimators: 3, 5, 10, 20, 50 Max depth: 3, 5, 10, 20, 50
SVM (L2)	gamma: $2^{-15} \leq x \leq 2^{5}$, with 10 steps C: $2^{-5} \leq x \leq 2^{10}$, with 10 steps decision function: one-vs-one, one-vs-rest
LDA	-
NB	-

Chapter 4

Materials and Methods

This chapter presents a detailed description of the materials used and the methodology employed in the conducted research. The purpose of this chapter is to provide readers with an overview of the approach taken to address the research objectives and to ensure the transparency and reproducibility of the study. By detailing the materials, data collection procedures, and analytical methods, this chapter guides future researchers interested in replicating or extending our study.

4.1 Dataset

We utilized the dataset employed in a previous study conducted by Merlin [36]. That allows us to compare results obtained using the proposed methodology with his, which can be considered the state-of-the-art baseline for the pass difficult classification problem. This shared dataset consists of 465 passes extracted from four matches of the 2016 Brazilian Football Championship's first division. By using the same dataset, our work allows for a direct comparison between the findings of both studies.

Corner kicks and free kicks were excluded from the analysis. Two Sony Handycam HDR-CX405 digital cameras were used to record the matches in HD resolution (720p image resolution, 1280×720 pixels) at an acquisition frequency of 30 Hz (sub-sampling of 15 Hz, or 15 frames per second). The semi-automatic tracking system of the DVideo software was used to determine the 2D positions of the players [21, 22]. Each player was identified by a number ($p = 1, 2, \ldots, 14$) corresponding to their position on the team, including the main players and three substitutes who participated in the match. The 2D coordinates of the players were defined as $X_p(t)$ and $Y_p(t)$, where t represents each instant of time and X and Y represent height and length on the field, respectively.

The technical and tactical aspects of passes were evaluated by two coaches and researchers who assessed the difficulty level of 465 passes. The determination of difficulty was based solely on observational assessments by the coaches, without relying on statistical or mathematical criteria. Each pass was assigned a class of either 1 (low difficulty), 2 (medium difficulty), or 3 (high difficulty). For validation purposes, only the passes on which both coaches agreed were considered. If there was a disagreement, a third coach was consulted for the final decision. The distribution of pass difficulty levels is presented

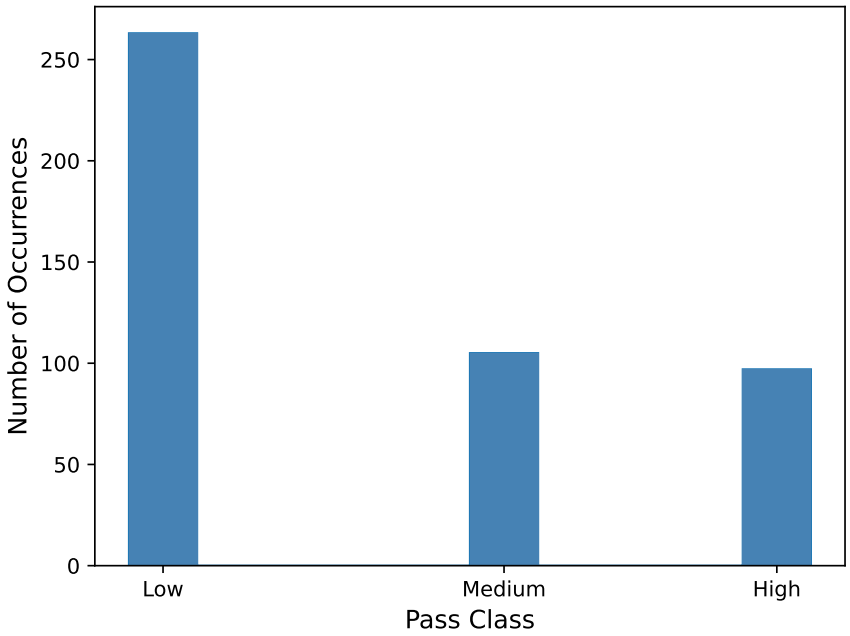


Figure 4.1: The distribution of 465 passes is shown across three different classes defined according to the pass difficulty: low, medium, and high.

in Figure 4.1. The entire dataset was split into three subsets, with 116 (25%) samples for the test dataset, 35 (7.5%) samples for the validation dataset, and 314 (67.5%) samples for the training dataset.

4.2 Evaluation Metrics

To account for the imbalanced nature of the dataset, with the prevalence of low-difficulty passes being higher than the others, we evaluated the models by computing the balanced accuracy (BACC) for each of the 5 test folds. BACC is a widely used metric for imbalanced datasets as it gives equal weight to each class, regardless of the number of samples in each class. We selected the model with the best BACC as a result of the experiment. The BACC is defined in Equation 4.1.

$$BACC = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} + \frac{TN_i}{TN_i + FP_i} \tag{4.1}$$

where N is the total number of classes, TP_i is the number of true positives for class i, FN_i is the number of false negatives for class i, TN_i is the number of true negative for class i, and FP_i is the number of false positive for class i.

4.3 Statistical Analysis

We employed Tukey's Honestly Significant Difference Test (Tukey's HSD Test) to perform posthoc analysis of multiple comparisons, which compares the mean of each group with the mean of every other group while controlling the family-wise error rate for a given confidence level (e.g., 95%) across any pair-wise comparisons [4]. In the context of analyzing which graph elements are better to represent pass difficulty in football, Turkey's HSD Test becomes a valuable tool. This test allows for a comprehensive examination of various types of graph elements, such as nodes and edges, to assess if there are statistically significant variations in pass difficulty scores between them. By conducting this test, researchers can identify which specific graph elements lead to significant differences in representing pass difficulty. This statistical approach adds rigor to the analysis, providing a reliable means to distinguish between graph elements and aiding in the selection of the most effective representations for conveying pass difficulty in football. The primary objective of this test is to identify which pairs of groups have significant differences in means. Tukey's Test is a variation of the t-statistic designed for multiple comparisons, and the statistic with the best performance is denoted by q_s, as defined in Equation 4.2, and the HSD is defined in Equation 4.3 [4, 48].

$$q_s = \frac{Y_{max} - Y_{min}}{SE}, \tag{4.2}$$

where Y_{max} and Y_{min} are the larger and smaller means of the two groups being compared. SE is defined as the standard error of the entire design. If the confidence intervals of two pairs contain 0, then there is no significant difference.

$$HSD = q_s \sqrt{\frac{MSE}{n}} \tag{4.3}$$

The HSD is a statistic that can be used to determine significant differences between two groups. If the absolute value of the difference between the two groups' means is greater than or equal to the HSD, the difference is significant, the interval is expressed in Equation 4.4.

$$|Y1 - Y2| \geq HSD \tag{4.4}$$

where $|Y1 - Y2|$ denotes the absolute difference between the means of two groups.

To analyze the BACC, we take into consideration confidence intervals, as expressed in Equation 4.6

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (ac_i - \overline{ac})^2} \tag{4.5}$$

where $\overline{ac}$ is the accuracy average and N is the number of accuracy.

$$ci = \left\{ \overline{ac} - 1.96\frac{\sigma}{N}, \overline{ac} + 1.96\frac{\sigma}{N} \right\} \tag{4.6}$$

where the factor 1.96 indicates the confidence interval of 95%.

4.4 Experimental Protocol

A data-driven approach was employed in this study to evaluate graph representations for characterizing pass exchange difficulty. Figure 4.2 illustrates our study protocol and the steps taken. We considered three lines of investigation:

1. Classifiers and Graph Elements: we evaluated various graph models and measurements. For each pass, two graphs were extracted: one from the initial and another from the final frame.

2. Temporal Analysis: we conducted a comprehensive examination of graph models and measurements to uncover significant findings from a multilevel temporal perspective. Unlike the previous analysis, which focused only on the initial and final frames of each pass, we evaluated multiple frames that comprise the passing event. This approach allowed us to capture the dynamic evolution of passing events more effectively. We considered different intervals of frames, starting from the initial moment when the passing player first touches the ball up to the moment when the pass the completed.

3. Feature Combination: we analyzed the combination of feature vectors that encode different graph models and frame intervals.

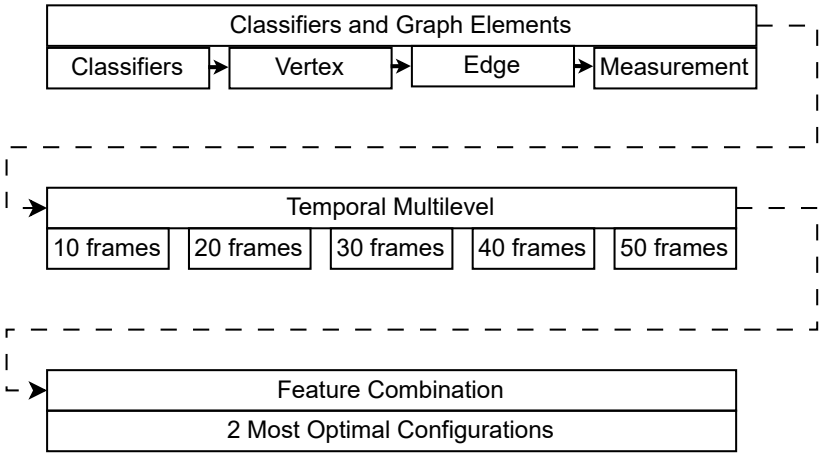


Figure 4.2: Experimental Protocol: The first step was to define the classifier and graph elements that best represent the difficulty of passes. The sub-steps involved determining the best classifier, the best vertex representation, the best edge representation, and the best graph measurements in that order. The second step was concerned with determining the optimal frame interval that captures a passing event. Sub-steps involved include the assessment of different time intervals, ranging from 10 to 50 frames. In the final step, we investigated the complementary aspect of the features, aiming to enhance the overall classification accuracy through their combination.

Table 4.1: The 161 combinations involving vertices, edges, and measurements. Among the edges, namely ED (Euclidean Distance), xD (Axis-x Distance), and WE (Weighted Edge), we considered variations in the distance: 2m, 5m, 10m, and 15m. Combinations tested are marked with ✓, while infeasible ones are marked with '-'. By systematically exploring these combinations, we gained insights into diverse configurations and identified the most effective ones. The measurements evaluated include Betweenness (B), Closeness (C), Cluster Coefficient (CC), Cluster Coefficient Bipartite (CC_b), Degree (D), and Robins Alexander Clustering (RAC), along with edge types Euclidean Distance (ED), Delaunay(Dl), Opponent Interference (OI), Axis-x Distance (xD), and Weighted Edge (WE).

Edges	**Vertex and Measurement**												
	Attackers				**Opponent-Aware**				**Attacker-Opponent Bipartite**				
	B	**C**	**CC**	**D**	**B**	**C**	**CC**	**D**	**B**	**C**	**CC_b**	**D**	**RAC**
Dl	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-
ED	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OI	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
xD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
WE	-	✓	✓	✓	-	✓	✓	✓	-	✓	✓	✓	✓

4.4.1 Classifiers and Graph Elements

This section provides an overview of the evaluation protocol adopted to address the raised research questions. The evaluation protocol aims to disentangle the relationship between pass difficulty classification and various graph elements, such as vertex, edges, and measurements. We examine how these elements impact pass difficulty classification and how they can be optimized to improve performance in a binary temporal resolution. We evaluate the elements, using the initial frame t_i and final frame t_f of each pass and complex network measurements from vertices PP (passing player), PR (passing receiver), and T (target). Through a comprehensive analysis of vertex, edge, and measurement factors, we demonstrate that certain graph configurations and metrics have a significant impact on pass difficulty classification. Performed experiments aim to characterize the importance of different graph elements in pass difficulty classification, we aimed to identify the key elements that can provide valuable insights for coaches, enabling them to make informed decisions to optimize pass efficiency and enhance overall team performance. Figure 4.3 presents the graph components considered in our study.

As shown in Figure 4.2, initially, the classifiers were evaluated, and those that exhibited better and significantly comparable results were selected for the subsequent evaluation of vertices. Similarly, as in the previous step, vertex representations were assessed, and only those that yielded better and significantly comparable results were chosen for the subsequent evaluation. Analogously, the evaluation of edges and graph measures followed the same protocol. In this way, the best factors for characterizing the difficulty of passing through graph models are taken to the next evaluation step. Table 4.1 summarizes the combinations of graph elements.

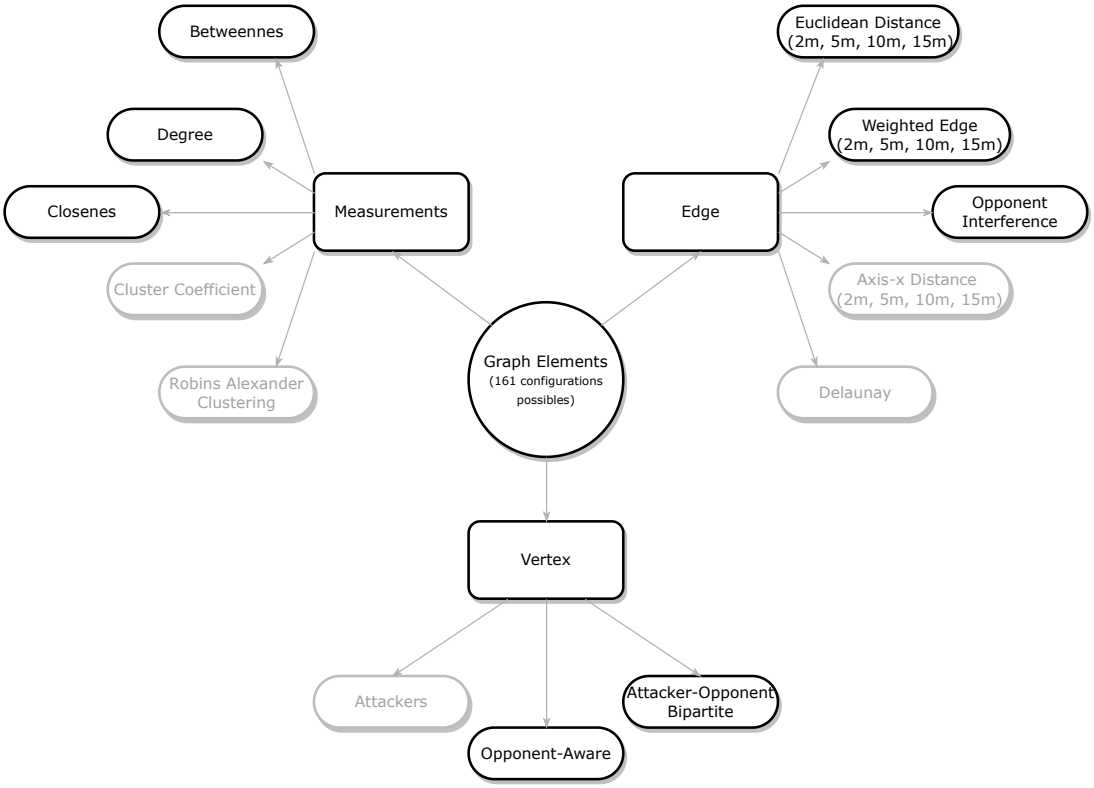


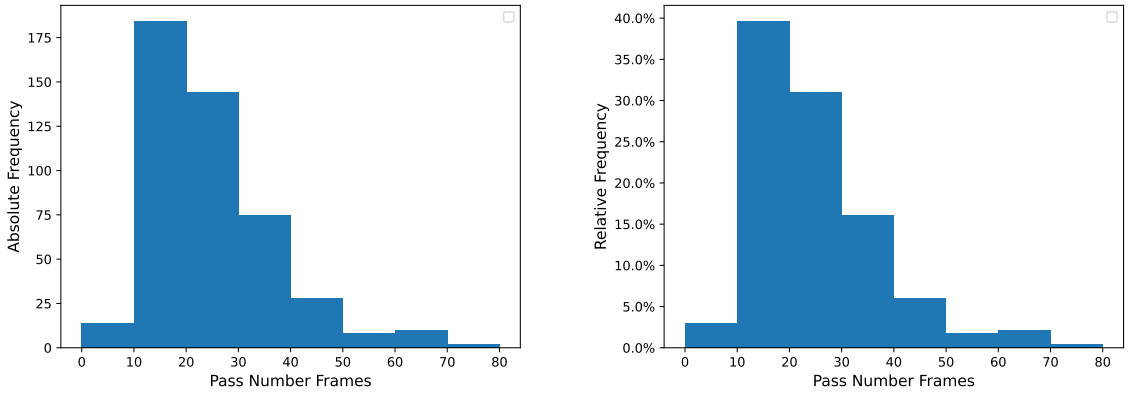
Figure 4.3: This diagram shows the graph elements (vertex, edge, and graph measurements) considered in our study. The elements highlighted in black are used in the multiscale temporal analysis.

4.4.2 Temporal Multilevel Scale

We begin our analysis of temporal networks by examining the pass at a lower temporal resolution, specifically focusing on the initial frame t_i and final frame t_f. In order to gain a more detailed understanding of networking evolution, we increase the temporal resolution. Figure 4.4 illustrates the distribution of pass duration, revealing that over 85% of passing events occur within a range of 10 to 50 frames. As described in Section 4.1, considering a sampling of 15 Hz, this corresponds to a pass duration between 667 ms and 3,333 ms. To determine the optimal frame interval for capturing passing events, we analyze passing intervals from the initial frame at timestamp t_i to the moment defined by thresholds of 10, 20, 30, 40, and 50 frames.

4.4.3 Feature Fusion

Our research also delves into the assessment of classifier effectiveness through the fusion of multiple feature vectors. To achieve this, we amalgamated the two most optimal configurations for each classifier, both on multilevel and binary scales. A breakdown of the number of features incorporated into the input of the multilevel frames interval can be found in Table 4.2. Notably, the size of the final vector is contingent upon the specific temporal level that has been merged in this fusion process.



(a) Absolute Frequency

(b) Relative Frequency

Figure 4.4: Distribution of passes based on the number of frames. (a) Histogram showing the distribution of passes according to the number of frames. The x-axis represents the number of frames, and the y-axis represents the frequency or count of passes falling within each frame range. (b) Percentage of passes based on the number of frames. The x-axis represents the number of frames, and the y-axis represents the percentage of passes relative to the total number of passes.

Table 4.2: Size of feature vectors by edge and temporal level. Euclidean Distance (ED), Opponent Interference (OI), and Weighted Edge (WE).

Edge	**Temporal Level**	**Size of the Feature Vector**
ED5M, ED10M, ED15M, WE5M, WE10M, WE15M	Binary (t_i and t_f)	6
	10 frames	30
	20 frames	60
	30 frames	90
	40 frames	120
	50 frames	150
OI	Binary (t_i and t_f)	4
	10 frames	20
	20 frames	40
	30 frames	60
	40 frames	80
	50 frames	100

Chapter 5

Results and Discussion

This chapter presents the results of the study and provides a comprehensive discussion of the findings in light of the research objective. The purpose of this chapter is to present the empirical outcomes and to engage in a meaningful interpretation of the data, thereby addressing the research questions and contributing to the broader body of knowledge in the field.

In the following sections, we present the key findings, organized according to the themes that emerged from the analysis. We also provide relevant statistical analyses and visual representations to support our interpretations. Through an examination of the data, we aim to uncover meaningful patterns and relationships that light on the research questions and contribute to a deeper understanding of the pass difficulty representation by graphs. Subsequently, we engage in a comprehensive discussion of the results. This discussion explores the implications of the findings, identifies potential limitations, and offers some insights into the theoretical, practical, and methodological implications of the study.

5.1 Classifiers and Graph Elements

In this section, we present the results of the analysis aimed at identifying the classifiers and the graph elements that best represent the pass difficulty classification problem. The objective of this analysis was to determine which specific elements, such as vertices, edges, or measurements, have a significant impact on accurately classifying pass difficulty.

First, we aimed to identify the most effective classifier for accurately classifying pass difficulty. We evaluated several machine learning algorithms, including LR, SVM, and RF, among others. Each classifier was trained and tested on the dataset, and its performance was assessed using the BACC metric.

Second, we examined the role of vertices in the classification task. We analyzed different vertex representations, Attackers, Opponent-Aware, and Attacker-Opponent Bipartite, and assessed their impact on the overall classification accuracy. We employed statistical measures to determine the significance of each vertex representation in predicting pass difficulty

Next, we investigated the influence of different edge types on classification perfor-

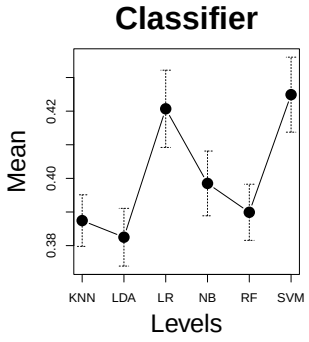
mance. We explored edge representations such as OI, ED5M, and ED10M. By analyzing the mean balanced accuracy and conducting statistical tests, we determined the relative importance of each edge representation in capturing pass difficulty.

Furthermore, we examined the role of measurements derived from the graph structure. These measurements included Degree, Betweenness Centrality, and Closeness Centrality, among others. We calculated these measurements for PP (passing player), PR (passing receive), and T (target) vertices and explored their relationship with pass difficulty classification.

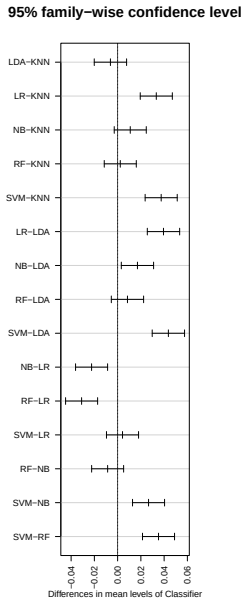
Through comprehensive statistical analyses, we identified the graph elements that significantly contribute to pass difficulty classification. The following subsections provide a detailed overview of the results, discussing the findings for each element individually and highlighting their relative importance.

5.1.1 Classifiers Results

We assessed the effectiveness of the classifiers by evaluating the features extracted from the combinations summarised in Table 4.1 from PP (passing player), PR (passing receiver), and T (target) vertices. Using the initial frame ti and final frame tf of each pass. Each combination serves as an input to the classifiers detailed in Section 2.1.3.



(a) Mean Accuracy



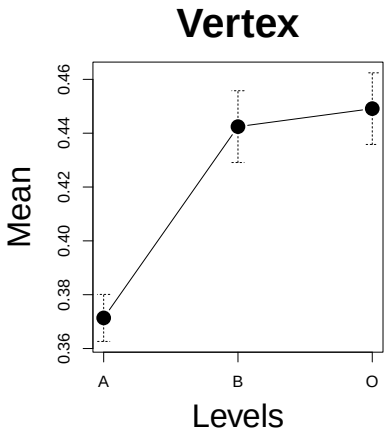
(b) Significant Difference Interval

Figure 5.1: Statistical test to compare the performance of multiple classifiers in binary temporal experiments. Figure 5.1a displays the mean accuracy and confidence intervals of each classifier, with larger error bars indicating greater variability. Figure 5.1b presents the results of the TukeyHSD test, which compares the means of different groups in the dataset. The figure displays confidence intervals for all pairwise comparisons of group means. If the confidence interval for a comparison includes zero, this indicates that the difference between the means is not statistically significant.

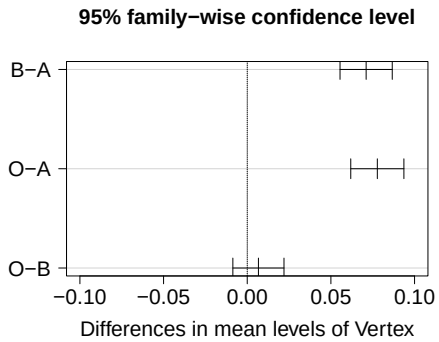
Figure 5.1a displays the mean accuracy for each classifier, with the confidence interval expressed by Equation 4.6. Figure 5.1b indicates whether there are significant differences among the classifiers. The LR and SVM classifiers had the highest average accuracy, with no significant difference observed among the other classifiers. In conclusion, the LR and SVM (with the added benefit of L2 regularization) classifiers achieve higher BACC. These classifiers excel in identifying and differentiating between relevant and irrelevant data, leading to precise analysis and informed decision-making. The regularization techniques also allow them to handle noise and irrelevant features more effectively, resulting in better accuracy and more robust models. Consequently, we utilized LR and SVM classifiers in subsequent evaluations.

5.1.2 Vertex Representations Results

We analyzed the performance of vertex types using Tukey's HSD post hoc test. Figure 5.2 presents the results. Our analysis showed that the accuracy of the *Attackers* vertices was only around 37%, indicating random performance in a three-class universe. On the other hand, the *Opponent-Aware* and *Attacker-Opponent Bipartite* vertices showed similar accuracy levels of around 44% and were statistically comparable. Therefore, we conclude that the B and O vertices exhibit superior performance as they highlight the interplay between passing players and opponents, providing a more effective representation of patterns related to the pass exchange difficulty. As a result, we consider just B and O vertices in the subsequent analysis.



(a) Mean Accuracy



(b) Significant Difference Interval

Figure 5.2: Figure 5.2a displays the mean accuracy and confidence intervals of each vertex representation. Moreover, the overlap between their error bars suggests that their means are not significantly different. Figure 5.2b shows the results of the TukeyHSD test, which compares the means of different groups in the dataset. The figure displays confidence intervals for all pairwise comparisons of group means. If the confidence interval for a comparison includes zero, it indicates that the difference between the means is not statistically significant.

Table 5.1: This summary considers all vertex, edge, and measurement combinations possible, except for the Attacker representation due to the results of Section 5.1.2. The table below indicates possible combinations with an ✓and impossible combinations with a '-'. Edges ED, xD, and WE have distance variations between 2m, 5m, 10m, and 15m, resulting in a total of 111 possible combinations. Betweenness (B), Closeness (C), Cluster Coefficient (CC), Cluster Coefficient Bipartite (CC_b), Degree (D), and Robins Alexander Clustering (RAC). Euclidean Distance (ED), Delaunay(Dl), Opponent Interference (OI), Axis-x Distance (xD), and Weighted Edge (WE).

Edges	**Vertex and Measurement**								
	Opponent-Aware				**Attacker-Opponent**				**Bipartite**
	B	**C**	**CC**	**D**	**B**	**C**	**CC_b**	**D**	**RAC**
Dl	✓	✓	✓	✓	-	-	-	-	-
ED	✓	✓	✓	✓	✓	✓	✓	✓	✓
OI	✓	✓	✓	✓	✓	✓	✓	✓	✓
xD	✓	✓	✓	✓	✓	✓	✓	✓	✓
WE	-	✓	✓	✓	-	✓	✓	✓	✓

5.1.3 Edge Results

To answer the question 'Identification of edge representations for football matches that are suitable for evaluating pass efficiency', we evaluated five different edge representations, as detailed in Section 3.3.

Figure 5.3 compares the mean balanced accuracy for all edge representations. The OI (opponent interference) representation achieved the highest balanced accuracy, approximately 48%, while the Dl representation had the lowest mean BACC, around 36%. This low accuracy can be considered a random performance in a three-class universe. Additionally, both OI and Dl were found to be statistically significantly different. To provide a more detailed comparison, we also present further analysis of the OI representation compared to other types of edge representations.

Figures 5.4a and 5.4b show the mean balanced accuracy and the results of Tukey's Test for OI and ED edge types, respectively. The OI edge type achieved the highest balanced accuracy and was found to be statistically significantly different from ED2M, while the difference in the mean balanced accuracy of the other edge types was not statistically significant. Figures 5.5a and 5.5b show the mean balanced accuracy and the results of Tukey's Test for OI and WE edge types, respectively. The OI edge type achieved the highest balanced accuracy and was found to be statistically significantly different from WE2M, while the difference in the mean balanced accuracy of the other edge types was not statistically significant

Figures 5.6a and 5.6b show the mean balanced accuracy and Tukey's Test results for OI and Dx edge types, respectively. The OI edge type achieved the highest balanced accuracy and was found to be statistically significantly different from all the Dx representations. Our analysis shows that edges that consider opponent interference or proximity are the

most significant for pass difficulty classification. However, very close proximity, such as 2 meters, does not show good performance for this problem, likely because it is a situation that does not occur frequently. Therefore, we selected the edges with the best performance with no statistically significant difference for further analysis, they are: OI, ED5M, ED10M, ED15M, WE5M, WE10M, and WE15M.

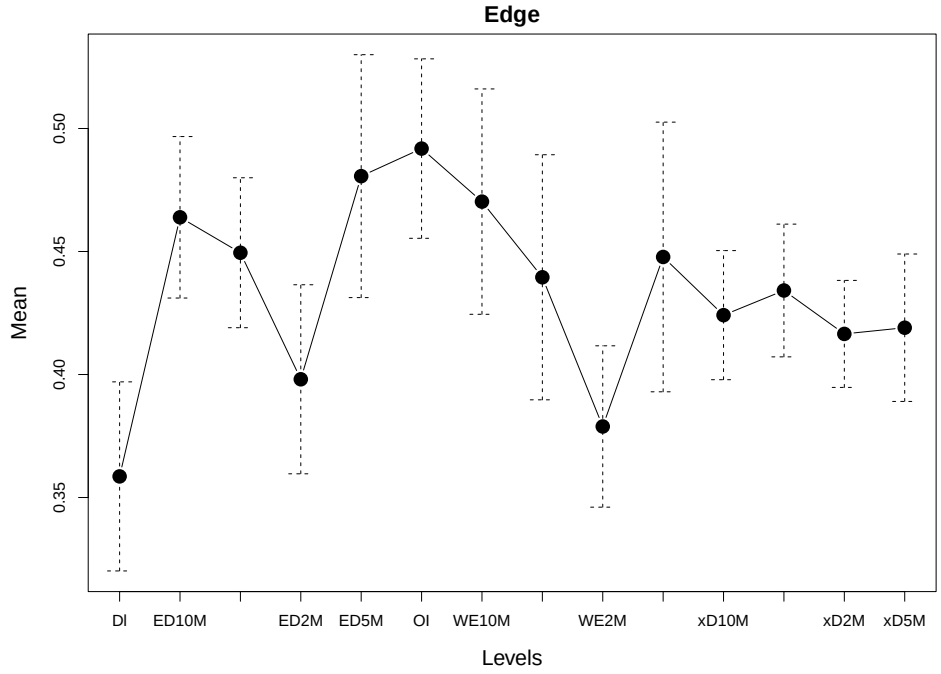
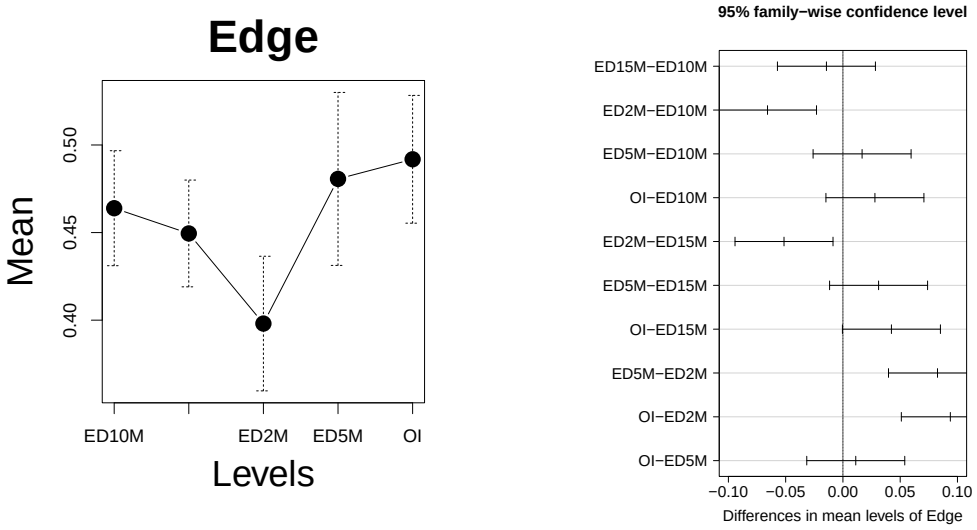


Figure 5.3: Mean BACC for all edges representations with respectively confidence intervals.

5.1.4 Graph Measurement Results

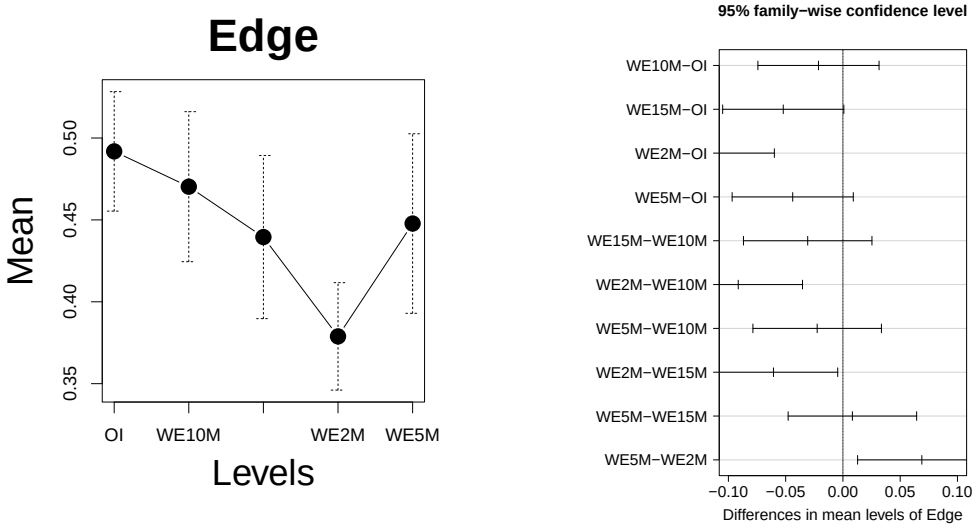
We addressed the question 'Identification of graph measurements for pass efficiency assessment', by evaluating five graph measurements, as described in detail in Section 2.1.2. We evaluated various measurements extracted from combinations, as summarized in Table 5.2, to classify pass difficulty in a binary temporal resolution. Each combination constituted a feature vector input to the LR and SVM classifiers, which we determined to be the best classifiers for our purpose, as discussed in Section 5.1.1. The results of Tukey's test analysis, with measurements as the factor, are presented in Figure 5.7. The analysis revealed that the Degree measurement achieved an accuracy rate of approximately 55%, which was statistically significantly different from the other measurements, and showed the highest mean BACC. These findings suggest that a measurement expressing the nominal value of connections between players is suitable for the pass difficulty classification problem.

Summary of Findings: Our analysis indicates that graph elements, specifically vertex, edge, and measurements, have a significant impact on pass difficulty classification. The most effective representations are those that consider opponent interactions such as



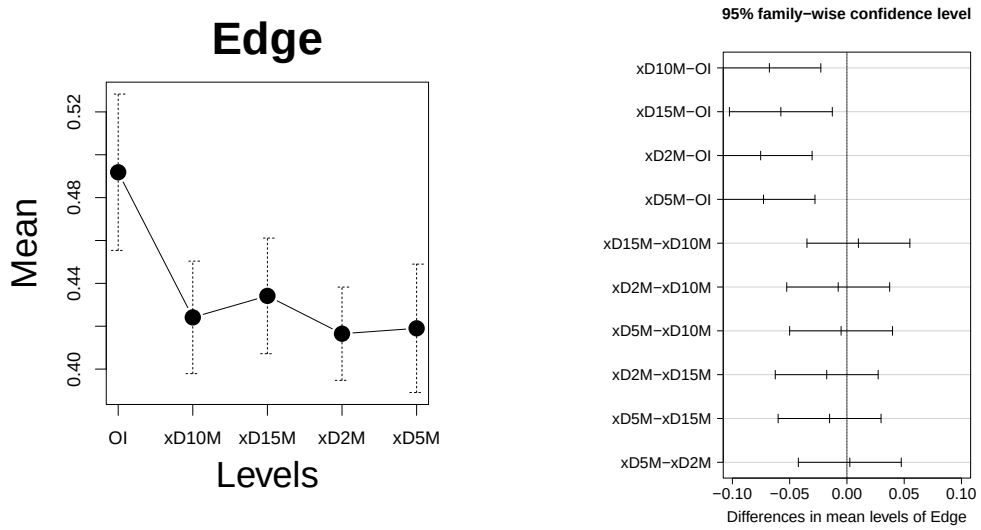
(a) Mean BACC (b) Confidence Interval

Figure 5.4: Statistical test that compared the performance of OI and ED edge representations experiments. Sub-figure 5.4a shows the Mean BACC for OI and ED edges representations with respectively confidence intervals. Sub-figure 5.4b shows the Confidence Interval from TukeyHSD for OI and ED edges representations.



(a) Mean BACC (b) Confidence Interval

Figure 5.5: Statistical test that compared the performance of OI and WE edge representations experiments. Sub-figure 5.5a shows the Mean BACC for OI and WE edges representations with respectively confidence intervals. Sub-figure 5.5b shows the Confidence Interval from TukeyHSD for OI and WE edges representations.

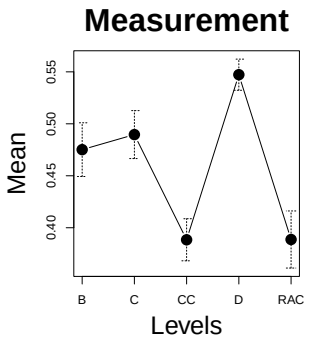


(a) Mean BACC (b) Confidence Interval

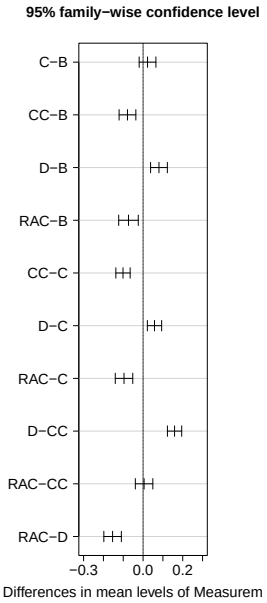
Figure 5.6: Statistical test that compared the performance of OI and xD edge representations experiments. Sub-figure 5.6a shows the Mean BACC for OI and xD edges representations with respectively confidence intervals. Sub-figure 5.6b shows the Confidence Interval from TukeyHSD for OI and xD edges representations.

Table 5.2: This summary considers all vertex, edge and measurement combinations possible, except for the Attacker representation due to the results of Section 5.1.2 and Edge xD and Dl representation follow the discuss in Section 5.1.3. The table below indicates possible combinations with an ✓and impossible combinations with a '-'. Edges ED, and WE have distance variations between 5m, 10m, and 15m, resulting in a total of 57 possible combinations. Betweenness (B), Closeness (C), Cluster Coefficient (CC), Cluster Coefficient Bipartite (CC_b), Degree (D), and Robins Alexander Clustering (RAC). Euclidean Distance (ED), Opponent Interference (OI), and Weighted Edge (WE).

Edges	**Vertex and Measurement**								
	Opponent-Aware				**Attacker-Opponent**				**Bipartite**
	B	**C**	**CC**	**D**	**B**	**C**	**CC_b**	**D**	**RAC**
ED	✓	✓	✓	✓	✓	✓	✓	✓	✓
OI	✓	✓	✓	✓	✓	✓	✓	✓	✓
WE	-	✓	✓	✓	-	✓	✓	✓	✓



(a) Mean BACC

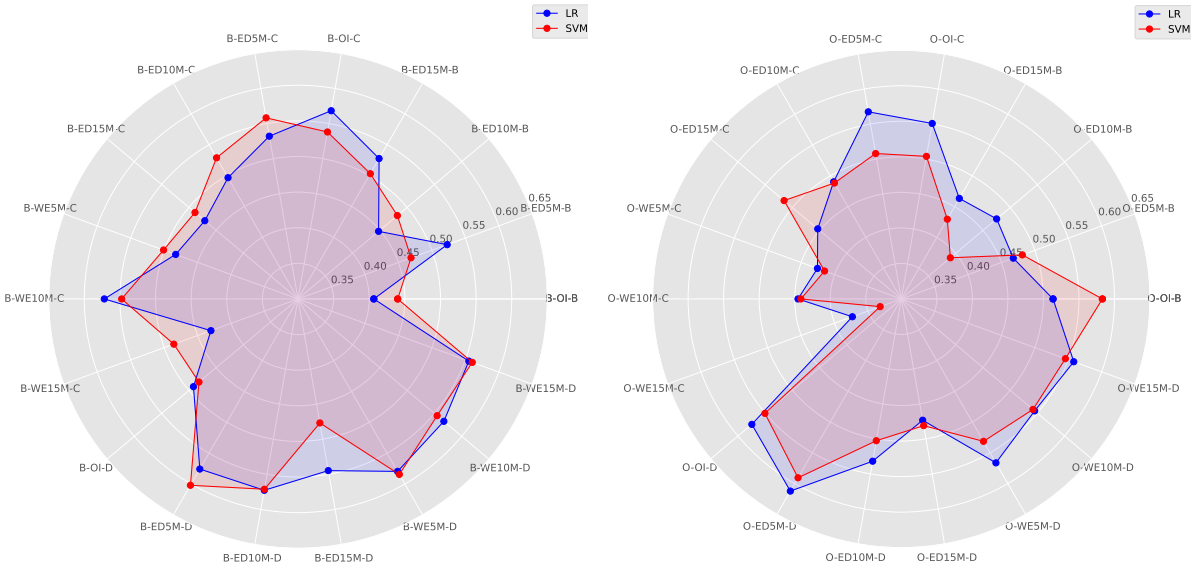


(b) Confidence Interval from TukeyHSD

Figure 5.7: Figure 5.2a compares the performance of various complex network measurements. The plot displays the mean BACC and confidence intervals for each measuremen. Figure 5.2b presents the outcomes of the TukeyHSD test, which evaluates the means of various groups in the dataset. The figure exhibits confidence intervals for all pairwise comparisons of group means. If the confidence interval for a comparison encompasses zero, it implies that the difference between the means is not statistically significant.

proximity and interference, as well as simple measurements that count the number of relations between vertices. Table 5.3 summarizes the most effective elements and their brief explanations. In Figure 5.10, we demonstrate how different factors impact each other, which confirms that the Attackers vertex representation and Delaunay edge representation are not suitable for pass difficulty classification. Furthermore, we observed a direct impact between edge representation and measurements. Among the measurements, the C measurement has the highest mean BACC, and we decided to include it along with B, which does not have a statistically significant difference between them, as shown in Figure 5.7. Finally, we present a mean BACC for the relevant elements in Table 5.4 and visualization is present in Figure 5.8.

The confusion matrices depicted in Figure 5.9 reveal the highest accuracy matrix for the O-ED5M-D configuration, with Logistic Regression as the runner-up. Interestingly, while the classifier effectively identified passes of low and high difficulty, it struggled to classify passes of medium difficulty. This observation suggests either the classifiers' superior performance in filtering out irrelevant data and noise or the potential dismissal of medium-difficulty samples during model training.



(a) Attacker-Opponent Bipartite Models (b) Opponent-Aware Models

Figure 5.8: Comparing the mean BACC of the best graph configurations in binary temporal resolution.

Table 5.3: Summary of the most effective graph element. Attacker-Opponent Bipartite (B), and Opponent-Aware (O). Euclidean Distance (ED), Opponent Interference (OI), and Weighted Edge (WE). Degree (D).

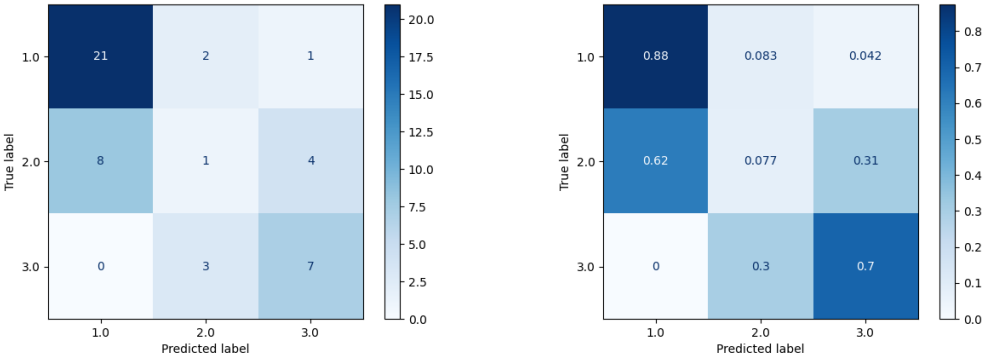
Factor	**Best Representations**	**Findings**
Vertex	B and O	Highlight the interplay between passing players and opponents.
Edge	OI, ED5M, ED10M, ED15M, WE5M, WE10M And WE15M	Consider opponent interference and/or proximity.
Measurements	D	Encode number of connections between players.

5.2 Temporal Multilevel Scale

We start our analyses of temporal networks from a lower temporal resolution considering the initial frame t_i and final frame t_f. To explore the process of network evolution in detail we increment the temporal resolution. Figure 4.4 presents the pass duration distribution. We observe that more than 85% of the passing events occur between 10 and 50 frames, as described in Section 4.1 the sampling is 15 Hz then usually the pass duration is between 667 ms and 3,333 ms. Aiming to identify the frame interval that better presents the passing we analysed passing intervals from the frame t_i to the thresholds 10, 20, 30, 40,

Table 5.4: Summary of the mean BACC performance for 33 relevant combinations of graph elements in binary temporal resolution for pass difficulty classification.

Combination	SVM	LR
B-ED5M-B	0.523	0.469
B-ED5M-C	0.532	0.558
B-ED5M-D	0.576	0.602
B-ED10M-B	0.448	0.482
B-ED10M-C	0.497	0.529
B-ED10M-D	0.573	0.571
B-ED15M-B	0.528	0.503
B-ED15M-C	0.471	0.489
B-ED15M-D	0.545	0.477
B-OI-B	0.406	0.440
B-OI-C	0.568	0.538
B-OI-D	0.525	0.550
B-WE5M-C	0.483	0.501
B-WE5M-D	0.580	0.584
B-WE10M-C	0.572	0.548
B-WE10M-D	0.567	0.555
B-WE15M-C	0.430	0.486
B-WE15M-D	0.555	0.560
O-ED5M-B	0.468	0.481
O-ED5M-C	0.567	0.508
O-ED5M-D	**0.612**	0.591
O-ED10M-B	0.475	0.390
O-ED10M-C	0.490	0.488
O-ED10M-D	0.532	0.503
O-OI-B	0.514	0.583
O-OI-C	0.551	0.504
O-OI-D	0.575	0.550
O-WE5M-C	0.426	0.415
O-WE5M-D	0.566	0.531
O-WE10M-C	0.445	0.441
O-WE10M-D	0.545	0.542
O-WE15M-C	0.373	**0.332**
O-WE15M-D	0.558	0.546



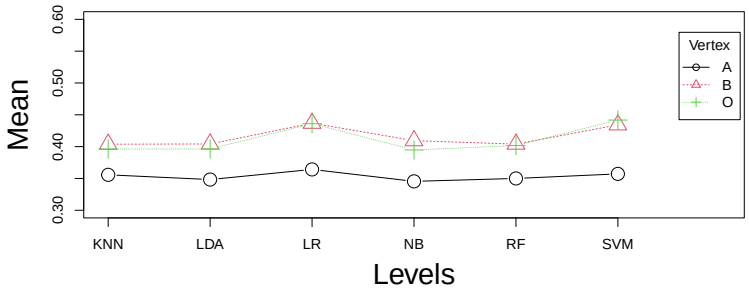
(a) Confusion Matrix (b) Normalized Confusion Matrix

Figure 5.9: Confusion matrices, visualized on a blue scale, pertain to the graph configuration O-ED5M-D. The x-axis denotes the predicted labels by the LR classifier, while the y-axis corresponds to the true labels of samples. Labels are categorized as follows: 1 = low difficulty, 2 = medium difficulty, and 3 = high difficulty. Figure 5.9a presents the nominal values, while Figure 5.9b illustrates the normalized values.

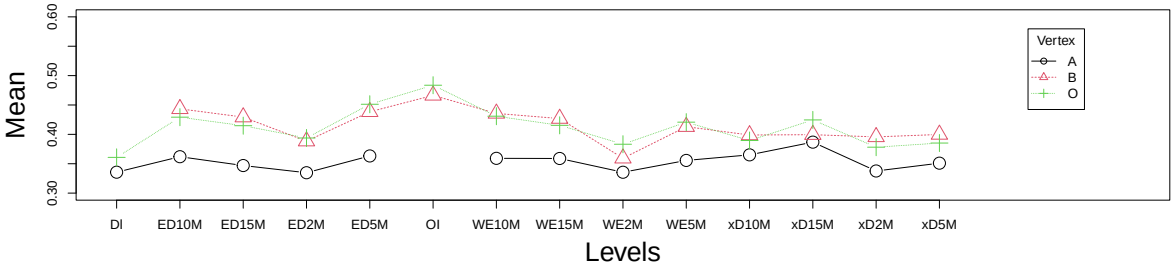
and 50 frames.

Figures 5.11 and 5.12 present a radar plot that provides a comprehensive visual comparison of frame intervals and different graph configurations. General patterns can be observed. First, the classifiers demonstrate similar trends across frame intervals, with each classifier having distinct highest or lowest points but overall displaying comparable behavior. Secondly, the Attacker-Opponent Bipartite vertex exhibits a lower concentration of points in the central area and showcases more consistent radar curves, indicating a more reliable configuration. Conversely, the Opponent-Aware vertex displays a higher concentration of points in the central area and presents a less uniform appearance, suggesting a relatively less consistent configuration.

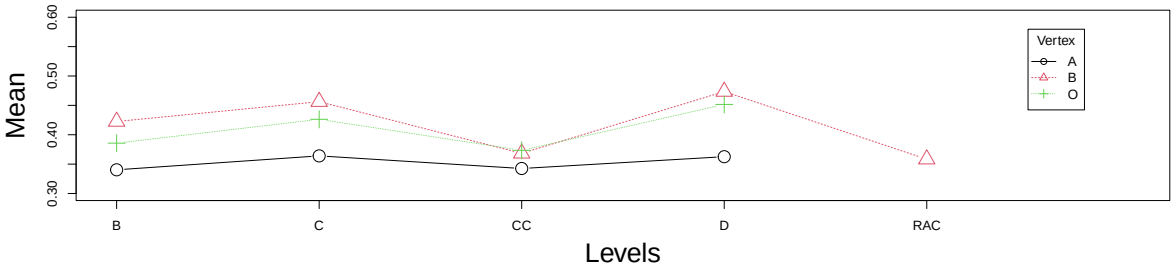
In Figure 5.11a, focusing on the Attacker-Opponent Bipartite vertex and LR classifier, the B-OI-C 50 frames achieved the highest Balanced Accuracy (BACC) with a value of 0.605. On the other hand, the B-OI-B Binary configuration obtained the lowest BACC of 0.406, resulting in a notable difference of 20% between the two. Another aspect observed is that BACC of edge ED and Degree e Closeness measurement present a crescent behavior for distance threshold BACC 5M $>$ BACC 10 M $>$ BACC 15M for the majority frame intervals. It indicates that the 5M threshold distance has more information about the pass difficulty, which was evidenced by Merlin's study [36]. Another aspect analyzed was the consistency of different frame intervals. We observed that the 50-frame interval exhibited the lowest number of axes with poor Balanced Accuracy (BACC) values and the second-highest number of axes with the best BACC values. Moreover, the 50-frame interval did not have any BACC values close to the radar center. This indicates that the 50-frame interval is a favorable choice, further supporting the hypothesis that more detailed data leads to more consistent results. Notably, it also yielded the highest BACC value in the plot. In contrast, the binary interval displayed a higher number of BACC values both at the higher and lower ends of the axes, indicating inconsistent behavior. Additionally, it had the lowest BACC value in the plot. Another notable observation is the configuration



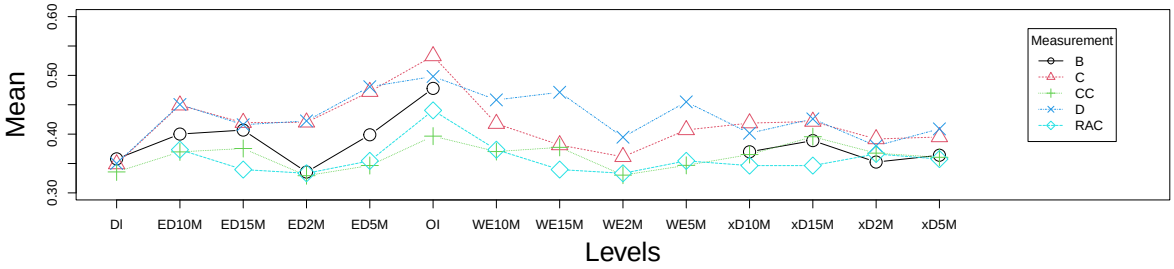
(a) Classifier vs Vertex



(b) Edge vs Vertex



(c) Measurement vs Vertex

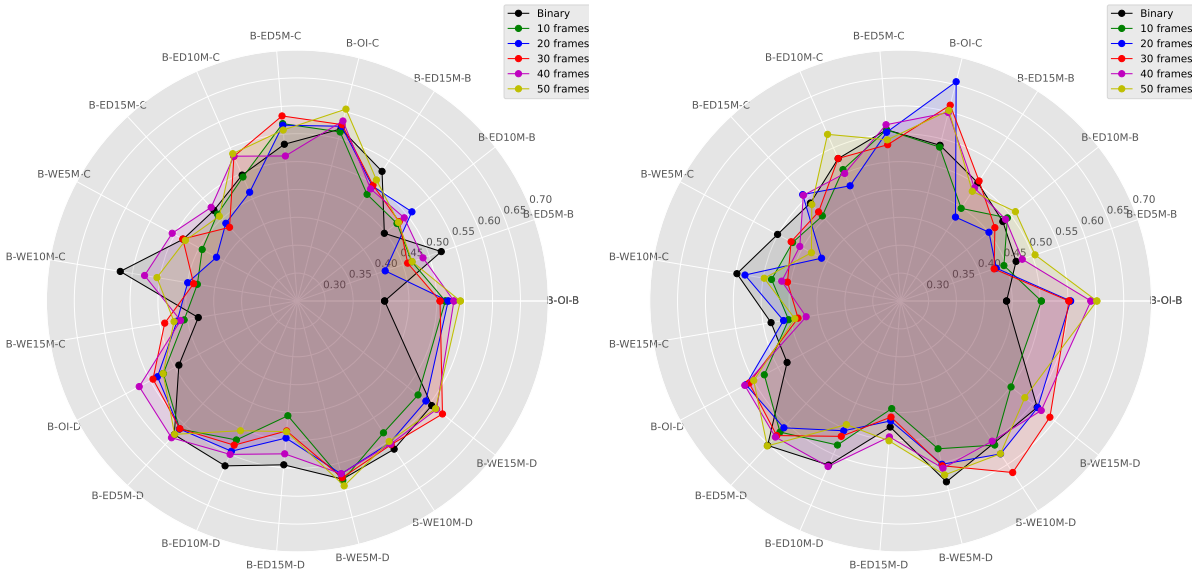


(d) Edge vs Measurement

Figure 5.10: Cross-comparison of binary temporal factors for pass difficulty classification. Figures 5.10a, 5.10a, and 5.10c show the comparative analysis of classifiers, edge representations, and measurements, respectively, for each vertex. In Figure 5.10d, we observe a direct impact between edge and measurement factors.

with the largest spacing between points, which is WE10M-C. In this configuration, the binary, 40, and 50 frame intervals demonstrated the highest Balanced Accuracy (BACC)

values, while the 10, 30, and 20 frame intervals showed the lowest BACC values. This suggests that for this specific configuration, a greater level of temporal detail positively influences its performance. On the other hand, the configuration with more condensed points is WE5M-D. In this case, all intervals exhibited BACC values ranging between 0.55 and 0.60, which are considered good values. This indicates that this configuration performs well, regardless of the level of temporal detail. Overall, these findings highlight the importance of considering both temporal detail and the specific configuration when assessing the performance of the classifiers.



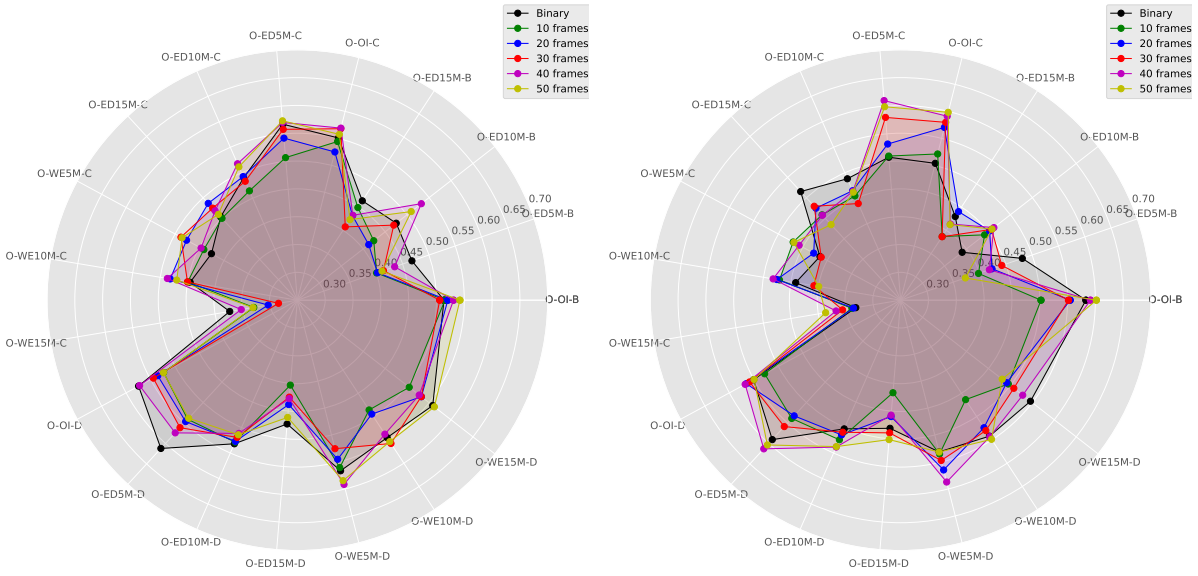
(a) LR classifier

(b) SVM classifier

Figure 5.11: Attacker-Opponent Bipartite graph showcases the performance ratings of eighteen graph configurations and measurements across six frame intervals: Binary, 10, 20, 30, 40, and 50 frames. Each line in the plot represents a specific frame interval, while the data points along the axes indicate the rating for each graph configuration.

Moving to 5.11b, which examines the same vertex type but with an SVM classifier, the B-OI-C with 20 frames yielded the highest BACC at 0.656, while the B-WE5M-C with 20 frames recorded the lowest BACC of 0.411, representing an approximate 25% difference. Another aspect observed is that BACC of edge ED and Degree measurement present a crescent behavior for distance threshold BACC 5M $>$ BACC 10 M $>$ BACC 15M for majority frame intervals. It indicates that the 5M threshold distance has more information about the pass difficulty, which was evidenced by Merlin's study [36]. Another aspect analyzed was the consistency of different frame intervals. We observed that the 50-frame interval exhibited the lowest number of axes with poor Balanced Accuracy (BACC) values and the highest number of axes with the best BACC values. Moreover, the 50-frame interval did not have any BACC values close to the radar center. This indicates that the 50-frame interval is a favorable choice, further supporting the hypothesis that more detailed data leads to more consistent results. The highest BACC value in the plot belongs to the 20-frame interval. In contrast, the binary interval displayed a near number of BACC values both at the higher and lower ends of the axes, indicating inconsistent

behavior. Additionally, the 20-frame interval had the lowest BACC value in the plot. Another notable observation is the configuration with the largest spacing between points, which is OI-B. In this configuration, 50, and 40-frame intervals demonstrated the highest Balanced Accuracy (BACC) values, while the binary and 10-frame intervals showed the lowest BACC values. This suggests that for this specific configuration, a greater level of temporal detail positively influences its performance. On the other hand, the configuration with more condensed points is ED5M-C. In this case, all intervals exhibited BACC values ranging between 0.50 and 0.60, which are considered good values. This indicates that this configuration performs well, regardless of the level of temporal detail. Overall, these findings highlight the importance of considering both temporal detail and the specific configuration when assessing the performance of the classifiers.



(a) LR classifier

(b) SVM classifier

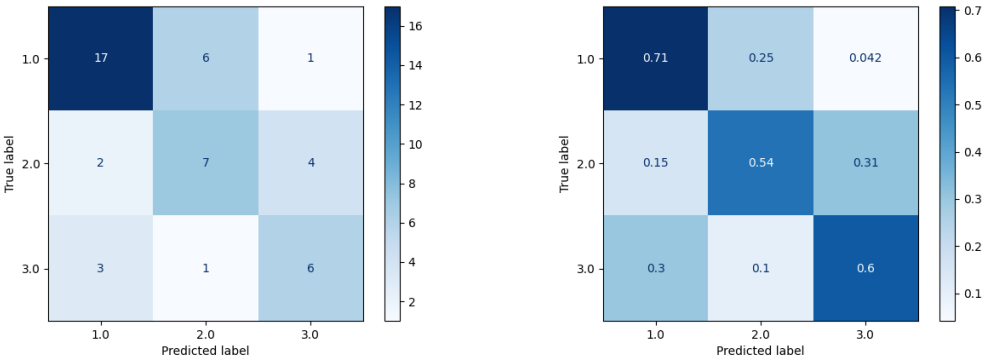
Figure 5.12: Opponent-Aware graph and showcases the performance ratings of eighteen graph configurations and measurements across six frame intervals: Binary, 10, 20, 30, 40, and 50 frames. Each line in the plot represents a specific frame interval, while the data points along the axes indicate the rating for each graph configuration.

Shifting the focus to Figure 5.12a, analyzing the Opponent-Aware vertex and LR classifier, the O-ED5M-D Binary configuration exhibited the highest BACC of 0.612, whereas the O-WE15M-C with 30 frames showed the lowest BACC at 0.284, indicating a considerable difference of around 30%. Another aspect observed is that BACC of edge ED and Degree e Closeness measurement present a crescent behavior for distance threshold BACC 5M $>$ BACC 10 M $>$ BACC 15M for all frame intervals. It indicates that the 5M threshold distance has more information about the pass difficulty, which was evidenced by Merlin's study [36]. Another aspect analyzed was the consistency of different frame intervals. We observed that the 40-frame interval exhibited the lowest number of axes with poor Balanced Accuracy (BACC) values and the second-highest number of axes with the best BACC values. Moreover, the 40-frame interval did not have any BACC values close to the radar center. This indicates that the 40-frame interval is a favorable choice,

further supporting the hypothesis that more detailed data leads to more consistent results. The highest BACC value in the plot belongs to the binary frame interval. In contrast, the binary interval displayed a near number of BACC values both at the higher and lower ends of the axes, indicating inconsistent behavior. Additionally, the 30-frame interval had the lowest BACC value in the plot. Another notable observation is the configuration with the largest spacing between points, which is ED10M-B. In this configuration, 40-, and 50-frame intervals yielded the highest Balanced Accuracy (BACC) values, while the 20, and 10-frame intervals showed the lowest BACC values. This suggests that for this specific configuration, a greater level of temporal detail positively influences its performance. On the other hand, the configuration with more condensed points is ED10M-D. In this case, all intervals exhibited BACC values ranging between 0.50 and 0.55, which are considered good values. This indicates that this configuration performs well, regardless of the level of temporal detail. Overall, these findings highlight the importance of considering both temporal detail and the specific configuration when assessing the performance of the classifiers.

Finally, Figure 5.12b explores the Opponent-Aware vertex and SVM classifier. We can observe that the O-ED5M-D with 40 frames achieved the highest BACC of 0.613, while the O-WE15M-C Binary configuration recorded the lowest BACC at 0.332, resulting in a similar 30 percentage point difference in performance. Another aspect observed is that BACC of edge ED and Degree measurement present a crescent behavior for distance threshold BACC 5M $>$ BACC 10 M $>$ BACC 15M for all frame intervals. It indicates that the 5M threshold distance has more information about the pass difficulty, which was evidenced by Merlin's works [36]. Another aspect analyzed was the consistency of different frame intervals. We observed that the 40-frame interval exhibited the lowest number of axes with poor Balanced Accuracy (BACC) values and the highest number of axes with the best BACC values. Moreover, the 40-frame interval did not have any BACC values close to the radar center. This indicates that the 40-frame interval is a favorable choice, further supporting the hypothesis that more detailed data leads to more consistent results. The highest BACC value in the plot belongs to the 40-frame interval. In contrast, the binary interval displayed a near number of BACC values both at the higher and lower ends of the axes, indicating inconsistent behavior. Additionally, binary frames interval had the lowest BACC value in the plot. Another notable observation is the configuration with the largest spacing between points, which is ED5M-C. In this configuration, 40-, and 50-frame intervals yielded the highest Balanced Accuracy (BACC) values, while the binary, and 10-frame intervals showed the lowest BACC values. This suggests that for this specific configuration, a greater level of temporal detail positively influences its performance. On the other hand, the configuration with more condensed points is ED10M-D. In this case, all intervals exhibited BACC values ranging between 0.50 and 0.55, which are considered good values. This indicates that this configuration performs well, regardless of the level of temporal detail. Overall, these findings highlight the importance of considering both temporal detail and the specific configuration when assessing the performance of the classifiers.

Figure 5.13 displays the confusion matrix with the highest accuracy for the B-OI-C configuration, utilizing 20 frames with the SVM classifier. Notably, the main diagonal



(a) Confusion Matrix (b) Normalized Confusion Matrix

Figure 5.13: Confusion matrices, visualized on a blue scale, pertain to the graph configuration B-OI-C for 20 frames. The x-axis denotes the predicted labels by the SVM classifier, while the y-axis corresponds to the true labels of samples. Labels are categorized as follows: 1 = low difficulty, 2 = medium difficulty, and 3 = high difficulty. Figure 5.13a presents the nominal values, while Figure 5.13b illustrates the normalized values.

exhibits darker shades compared to the previous confusion matrix (Figure 5.9), which attained the highest accuracy among binary temporal resolution configurations. This underscores the advantageous impact of temporal detail on classifying passes of medium difficulty. Specifically, the central square reflects a higher percentage of correctly classified instances.

Summary of Findings: Table 5.5 highlights optimal configurations and frame intervals identified in our experiments. Notably, for Vertex B, regardless of the classifier, the OI edge and closeness measurements emerge as the most effective, suggesting that a bipartite graph aptly captures how the number of opponents between the player and the target impacts pass difficulty. Closeness, indicating a vertex's proximity to the rest of the graph, provides insights into the presence or absence of opponents. Conversely, for Vertex O, irrespective of the classifier, the ED5M edge and degree measurements stand out. This representation encodes the pressure on the player within a 5 meters threshold.

Table 5.6 presents a complementary multilevel analysis. Notably, for degree measurements and the ED edge, the smaller threshold distances exhibit significantly higher Balanced Accuracy (BACC), emphasizing the importance of positional data close to the player. However, the 2-meter threshold seems to be limited to capturing crucial information adequately. Additionally, the analysis reveals that intervals of 40 and 50 frames stand out as the most consistent, displaying a higher number of data points distributed farther from the radar center.

Table 5.7 provides insights into the consistency of intervals. Remarkably, the binary frame consistently exceeds the lower end of the axes by more than one, while the 40 and 50 frames consistently maintain a maximum of one. Notably, the binary frame exhibits a considerable number of instances reaching the higher end of the axes, although generally comparable to the occurrences observed with the 40 and 50 frames.

Table 5.5: Optimal configurations and frame intervals, showcasing the best Balanced Accuracy (BACC) for each Vertex and Classifier pairs.

Vertex & Classifier	Configuration	Interval	BACC
B & LR	B-OI-C	50 frames	0.605
B & SVM	B-OI-C	20 frames	0.656
O & LR	O-ED5M-D	binary	0.612
O & SVM	O-ED5M-D	40 frames	0.613

Table 5.6: Complementary multilevel analysis. For each Vertex and Classifier pairs, we enumerate measurements where the Balanced Accuracy (BACC) for 5M ED threshold exceeds that for 10M, and BACC for 10M surpasses that for 15M. The third column indicates the intervals with consistent distance from the radar center across configurations.

Vertex & Classifier	Measurements where a smaller distance threshold has greater BACC	Optimal Frame Interval with Consistent BACC
B & LR	D and C	50 frames
B & SVM	D	50 frames
O & LR	D and C	40 frames
O & SVM	D	40 frames

Table 5.7: The consistency of frame intervals for each vertex and classifier pairs.

Vertex & Classifier	Interval (frames)	Number of BACC values at the higher end of the axes	Number of BACC values at the lower end of the axes
B & LR	50	4	1
B & LR	binary	6	4
B & SVM	50	6	1
B & SVM	binary	4	2
O & LR	40	5	1
O & LR	binary	7	2
O & SVM	40	7	0
O & SVM	binary	4	5

5.3 Feature Fusion

Two feature fusions were carried out, the first one considering the best configurations for the LR classifier, which are B-OI-C-50f with 100 features resulting in a BACC of 0.60, and the configuration O-ED5M-D-binary with 6 features and a BACC of 0.61. This first fusion resulted in a vector with 106 features and a BACC of 0.66. The second fusion, on the other hand, considered the configurations with the best BACC for the SVM classifier: O-ED5M-D-40f with 120 features, yielding a BACC of 0.61, and the configuration B-OI-C-20f with 40 features and a BACC of 0.65. This second fusion produced a vector with

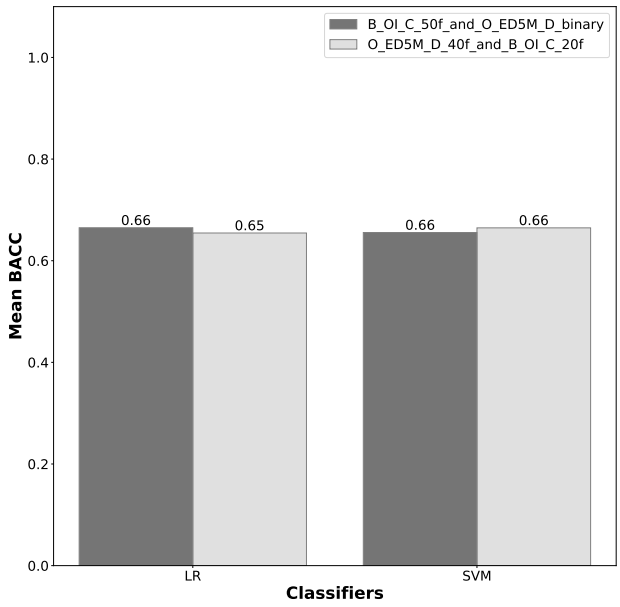
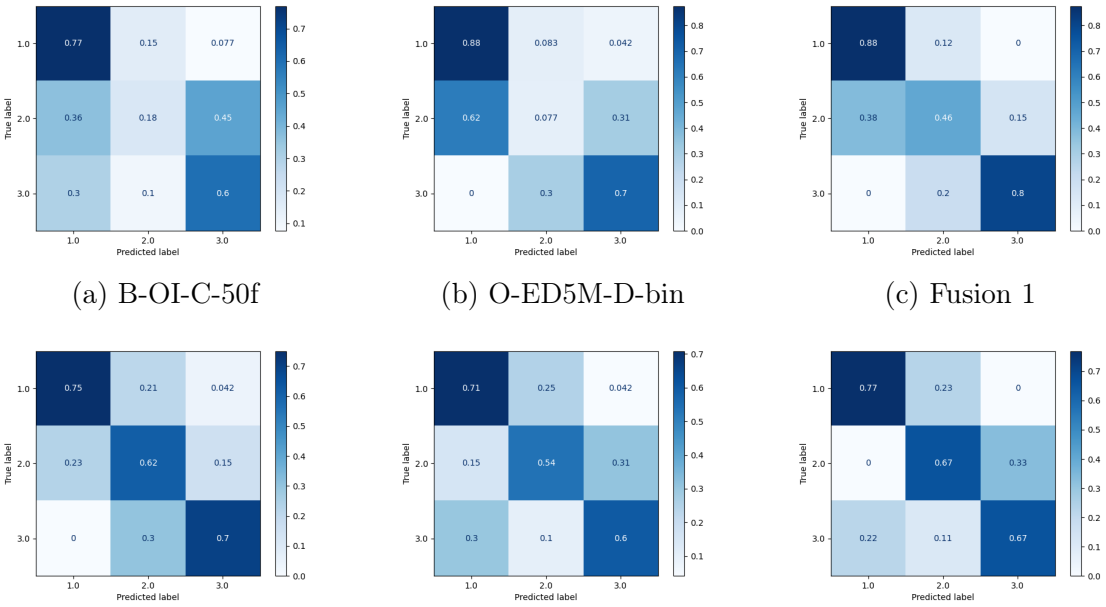


Figure 5.14: Mean Balanced Accuracy (BACC) of Feature Fusion for LR and SVM Classifiers.

160 features and a BACC of 0.66. The fusion results can be found in Figure 5.14.

Figure 5.15 showcases the confusion matrices representing the highest accuracy for each configuration used in fusion. The top row matrices correspond to results obtained with the Logistic Regression classifier, while the bottom row matrices represent outcomes from the Support Vector Machine classifier. Fusion occurs between the optimal configurations for each classifier. Interestingly, the Support Vector Machine classifier exhibits a darker main diagonal compared to Logistic Regression. However, Logistic Regression provides more distinct differentiation between low and high-difficulty classes. Notably, the combinations of features benefited the classification of medium-difficulty pass samples for both classifiers.



(a) B-OI-C-50f (b) O-ED5M-D-bin (c) Fusion 1

(d) O-ED5M-D-40f (e) B-OI-C-20f (f) Fusion 2

Figure 5.15: Confusion matrices are depicted in a blue scale for the fusion features. The x-axis represents the predicted labels for the classifiers, while the y-axis represents the true labels of samples. Labels are defined as follows: 1 = low difficulty, 2 = medium difficulty, and 3 = high difficulty. Figure 5.15a presents the normalized confusion matrix for the B-OI-C configuration with 50 frames and the LR classifier. Figure 5.15b presents the normalized confusion matrix for the O-ED5M-D configuration with binary temporal level and LR classifier. Figure 5.15c presents the normalized confusion matrix for the fusion of configurations B-OI-C with 50 frames and O-ED5M-D with binary classification for the LR classifier. Figure 5.15d presents the normalized confusion matrix for the O-ED5M-D configuration with 40 frames classified by SVM. Figure 5.15e presents the normalized confusion matrix for the B-OI-C configuration with 20 frames. Figure 5.15f presents the normalized confusion matrix for the fusion of configurations O-ED5M-D with 40 frames and B-OI-C with 20 frames.

Chapter 6

Conclusions

This chapter summarizes the main contributions of the work and outlines potential avenues for future research.

6.1 Contributions

Sport analysis is a burgeoning field that capitalizes on the wealth of big data made available through the implementation of automatic monitoring systems [25]. This vast data reservoir has been harnessed to facilitate the knowledge discovery process, aiding in the formulation of effective match strategies and training programs across various sports. Football, being a globally popular sport with a significant economic impact, has particularly benefited from such analyses.

In this exploration, the integration of spatial and temporal data poses a significant challenge, compelling researchers to develop systems capable of modeling, visualizing, and analyzing the dynamic nature of football. As part of this effort, the present thesis delved into the application of complex network measures for the analysis of passing difficulty, with a focus on the players' spatial distribution during passes. This entails a consideration of the spatiotemporal dynamics inherent in football passes. The primary objective was to contribute tools for modeling and analyzing such information, thus supporting the knowledge discovery process.

The study was structured in alignment with the research questions outlined in Section 1.2. The identified contributions are as follows:

- The first question delved into the investigation of **optimal graph representations for capturing patterns associated with pass difficulty**. In Section 4.4.1, we explored various vertex and edge representations to address this inquiry. Our findings revealed that vertex representations exclusive to the attacking team proved ineffective. In contrast, bipartite and opponent-aware representations demonstrated comparable performance. Notably, edges reflecting players' Euclidean proximity and opponent interference along the x-axis emerged as effective representations. Conversely, representations, such as Delaunay triangulation and players' distance along the x-axis, were deemed less effective.

- The second question aimed to investigate **which graph measures are well-suited for capturing patterns associated with pass difficulty**. This investigation is detailed in Section 4.4.1, where we assessed centrality and cycle-related measurements. Our results indicate that the most effective measurement is the centrality measure Degree. These findings suggest that a measurement reflecting the sheer quantity of connections between players is optimal for the pass difficulty classification problem. Additionally, other centrality measures, such as betweenness and closeness, demonstrated commendable performance. Conversely, cycle-related measurements like cluster coefficient and Robins Alexander clustering exhibited less favorable performance.

- The third question was concerned with the investigation of **which classification system is suitable to determine the level of pass difficulty**. We addressed this question in Section 4.4.1. Among the investigated some classifiers, the LR and SVM (with the added benefit of L2 regularization) classifiers achieved the highest BACC scores. These classifiers excelled in identifying and differentiating between relevant and irrelevant data, and handle noise and irrelevant features more effectively, resulting in better accuracy and more robust models. In the configuration O-ED5M-D (Opponent-aware, Euclidean distance with a 5-meter threshold, and Degree measurement), employing the SVM classifier, an accuracy of 61% was achieved.

- The fourth question aimed to investigate **whether multilevel temporal representations contain crucial information that enhances the accuracy of pass difficulty analysis**. This inquiry is addressed in Section 4.4.2. We analyzed the impact of different intervals of frames on Balanced Accuracy (BACC) in pass difficulty classification.

  For bipartite configurations with both the LR and SVM classifiers, the 50-frame interval showed notable characteristics. It exhibited the lowest number of axes with poor BACC values (just one for both classifiers) and the second-highest number of axes for LR, and the highest number for SVM. Furthermore, the 50-frame interval did not have any BACC values close to the radar center. This suggests that the 50-frame interval has a consistent behavior and it is a favorable choice, providing additional evidence that more detailed data leads to more consistent results. In contrast, the binary interval demonstrated a higher number of BACC values, totaling 6 for LR and 4 for SVM, at the higher end, and 4 for LR and 2 for SVM at the lower end of the axes. These findings indicate inconsistent behavior in the binary interval.

  For opponent-aware configurations with both the LR and SVM classifiers, the 40-frame interval exhibited notable characteristics. It demonstrated the lowest number of axes with poor BACC values (just one for LR and zero for SVM) and the second-highest number of axes for LR and the highest number for SVM (with 5 for LR and 7 for SVM). Moreover, the 40-frame interval did not have any BACC values close to the radar center. This indicates that the 40-frame interval has a consistent behavior and is a favorable choice, providing additional evidence that more detailed

data leads to more consistent results. In contrast, the binary interval displayed a totaling 7 for LR and 4 for SVM, at the higher end, and 2 for LR and 5 for SVM at the lower end of the axes. These findings suggest inconsistent behavior in the binary interval.

For multilevel evaluation, the optimal temporal resolution was found to be 20 frames, utilizing the B-OI-C (Bipartite attackers-opponent, Opponent Interference, and Closeness measurement)configuration with an SVM classifier. This feature set achieved an accuracy of 65%, aligning with the agreement index of the two annotator coaches.

- The remaining question, which pertains to investigating **whether the combination of features can enhance pass difficulty classification accuracy and provide deeper insights into the underlying patterns**, was explored in Section 4.4.3. In this section, we analyzed the combination of the two best configurations and frame intervals for each classifier (LR and SVM). The results demonstrated a marginal increment in accuracy, just 1%. Consequently, we conclude that the best configurations in multilevel analysis are not incremental or complementary.

The study was guided by two primary hypotheses: *(1) Temporal graphs and their associated complex-network measurements effectively model pass difficulty classification; and (2) Detailed temporal representations yield improved classification accuracy.* We demonstrated that graph representations achieved an accuracy of 61%, indicating their efficacy in modeling the problem. Additionally, we observed that higher temporal resolutions improved accuracy from 61% with binary resolution to 65%, aligning closely with the agreement index of the two annotator coaches. Therefore, we conclude that the hypotheses put forth were confirmed.

6.2 Future Work

The work carried out opens up new research opportunities. Below, we enumerate some of these possibilities for investigation:

- Expanding the scope of our research, a promising avenue for future exploration involves extending the analysis of graphs to encompass a broader spectrum of game events. Beyond the specific focus on pass difficulty, examining graphs in the context of other game events presents an opportunity to gain a holistic understanding of the underlying dynamics of sports gameplay. This extension could involve investigating patterns, connectivity, and centrality measures in the representation of diverse in-game activities. By delving into the analysis of graphs in various game events, we aim to unravel unique insights into strategic interactions, player dynamics, and overall team performance, thereby contributing to a more comprehensive and nuanced comprehension of the intricacies within the sporting domain;

- Investigate the integration of advanced computational models, specifically Deep Learning architectures, to refine our classifier toolkit within the sports analytics

framework. This initiative would involve a meticulous process of selecting, designing, and training neural network models, such as Convolutional Neural Networks (CNNs) for analyzing spatial patterns in game events and Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks for capturing the temporal dependencies of sequences of events. The computational endeavor would extend to applying transfer learning techniques to leverage pre-trained models for enhancing learning efficiency and accuracy with relatively smaller sports datasets.

- Graph neural networks present a promising venue to learn representations without relying on handcraft feature representations designed to capture structures and proprieties of graphs useful for answering relevant research questions in sports analysis [43, 51];

- An exciting avenue for future research lies in the in-depth analysis of graphs specifically tailored to sequences of events where a team is in possession of the ball. By focusing on these sequences, we aim to unravel intricate patterns and interactions among players, shedding light on the team's strategic decision-making processes, ball movement dynamics, and overall gameplay effectiveness. This analysis could involve studying graph properties, such as node centrality and connectivity, within the context of continuous possession. Understanding the underlying graph structures during these sequences may provide valuable insights into team coordination, offensive strategies, and the factors influencing successful ball progression. Exploring the graphical representation of possession sequences holds the potential to enhance our comprehension of team dynamics and inform strategic insights for coaches and analysts in the realm of sports analytics;

- An intriguing avenue for future exploration involves the utilization of temporal visualization techniques, particularly through the implementation of Visual Graph Rhythm, as proposed by Rodrigues [49]. This approach holds the promise of offering enhanced insights into the temporal dynamics of information;

- Exploring the dynamic aspects of player interactions within a sports context through continuous graph analysis could be a valuable avenue. Investigate how graph structures evolve during gameplay, focusing on player movements, collaborations, and strategic formations. This approach could provide deeper insights into the fluid nature of team dynamics during different phases of a match;

- Another promising research venue relies on investigating fusion strategies to combine different classification models, trained on different graph measurements, toward achieving a more effective classification system [18]; and

- Develop and refine graph-based metrics tailored specifically for evaluating team performance in sports. Consider creating novel measures that capture the effectiveness of team coordination, communication, and adaptability. By establishing comprehensive metrics, researchers can offer sports professionals a more nuanced and data-driven approach to assessing team dynamics and making informed decisions for improvement.

Bibliography

[1] E. Arriaza-Ardiles, J.M. Martín-González, M.D. Zuniga, J. Sánchez-Flores, Y. de Saa, and J.M. García-Manso. Applying graphs and complex networks to football metric interpretation. *Human Movement Science*, 57:236–243, 2018.

[2] Natàlia Balague, Carlota Torrents, Robert Hristovski, Keith Davids, and Duarte Araújo. Overview of complex systems in sport. *Journal of Systems Science and Complexity*, 26(1):4–13, Feb 2013.

[3] Albert-László Barabási. *Network Science*. Cambridge University Press, 2018.

[4] Yoav Benjamini and Henry Braun. John w. tukey's contributions to multiple comparisons. *The Annals of Statistics*, 30(6):1576–1594, 2002.

[5] Giuseppe Bonaccorso. *Machine Learning Algorithms: A Reference Guide to Popular Algorithms for Data Science and Machine Learning*. Packt Publishing, 2017.

[6] Stephen P. Borgatti and Daniel S. Halgin. Analyzing affiliation networks. *The SAGE Handbook of Social Network Analysis*, pages 417–433, 2014.

[7] Ulrik Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.

[8] Markus Brandt and Ulf Brefeld. Graph-based approaches for analyzing team interaction on the example of soccer. In *CEUR Workshop Proceedings*, volume 1970, page 10 – 17, 2015.

[9] Lotte Bransen, Jan Van Haaren, and Michel van de Velden. Measuring soccer players' contributions to chance creation by valuing their passes. *Journal of Quantitative Analysis in Sports*, 15(2):97–116, 2019.

[10] Fabio Caetano, Paulo Santiago, Ricardo Torres, Sergio Cunha, and Felipe Moura. Interpersonal coordination of opposing player dyads during attacks performed in official football matches. *Sports biomechanics*, pages 1–16, 05 2023.

[11] Christopher Carling, Tom Reilly, and A Mark Williams. *Performance assessment for field sports*. Routledge, 2008.

[12] Philippe Chassy. Team play in football: How science supports f. c. barcelona's training strategy. *Psychology*, 04:7–12, 01 2013.

[13] Christian Collet. The possession game? a comparative analysis of ball retention and team success in european and international football, 20072010. *Journal of sports sciences*, 31, 10 2012.

[14] Padraig Cunningham and Sarah Delany. k-nearest neighbour classifiers. *Mult Classif Syst*, 54, 04 2007.

[15] Luciano da F. Costa, Francisco A. Rodrigues, Gonzalo Travieso, and Paulino R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, jan 2007.

[16] Matheus Henrique de Amorim Mendes, Vitor Hugo Santos Rezende, and Gibson Moreira Praça. Association between goal kick strategies and the offensive outcome in the uefa euro 2020. *Human Movement*, 24(3):64 – 70, 2023.

[17] Laura M.S. de Jong, Paul B. Gastin, Lyndell Bruce, and Dan B. Dwyer. Teamwork and performance in professional women's football: A network-based analysis. *International Journal of Sports Science and Coaching*, 18(3):848 – 857, 2023.

[18] Fabio A. Faria, Jefersson A. dos Santos, Anderson Rocha, and Ricardo da S. Torres. A framework for selection and fusion of pattern classifiers in multimedia recognition. *Pattern Recognition Letters*, 39(0):52 – 64, April 2014. Advances in Pattern Recognition and Computer Vision.

[19] Javier Fernández and Luke Bornn. Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *MIT Sloan Sports Analytics Conference 2018*, 03 2018.

[20] Pascual Figueroa, Neucimar Leite, and Ricardo Barros. Tracking soccer players aiming their kinematical motion analysis. *Computer Vision and Image Understanding*, 101:122–135, 02 2006.

[21] Pascual J. Figueroa, Neucimar J. Leite, and Ricardo M.L. Barros. A flexible software for tracking of markers used in human motion analysis. *Computer Methods and Programs in Biomedicine*, 72(2):155–165, 2003.

[22] Pascual J. Figueroa, Neucimar J. Leite, and Ricardo M.L. Barros. Background recovering in outdoor image sequences: An example of soccer players segmentation. *Image and Vision Computing*, 24(4):363–374, 2006.

[23] Iztok Fister, Karin Ljubic Fister, Ponnuthurai Nagaratnam Suganthan, Matjaz Perc, and Iztok Fister. Computational intelligence in sports: Challenges and opportunities within a new research domain. *Appl. Math. Comput.*, 262:178–186, 2015.

[24] Floris Goes, Matthias Kempe, Rens Meerhoff, and Koen A.P.M. Lemmink. Not every pass can be an assist: A data-driven model to measure pass effectiveness in professional soccer matches. *Big Data*, 7, 09 2018.

[25] F.R. Goes, L.A. Meerhoff, M.J.O. Bueno, D.M. Rodrigues, F.A. Moura, M.S. Brink, M.T. Elferink-Gemser, A.J. Knobbe, S.A. Cunha, R.S. Torres, and K.A.P.M. Lemmink. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, 21(4):481–496, 2021. PMID: 32297547.

[26] Alvin M Goh, Eric J Drinkwater, Craig A Harms, Mark Scanlan, Robert U Newton, and Fadi Ma'ayah. Characteristics of goals scored in open play at the 2017 and 2018 australian national cerebral palsy football championship. *International Journal of Sports Science and Coaching*, 18(3):858 – 866, 2023.

[27] Jacob Goldberger, Sam Roweis, Geoff Hinton, and Ruslan Salakhutdinov. Neighbourhood components analysis. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, NIPS'04, page 513–520, Cambridge, MA, USA, 2004. MIT Press.

[28] Laszlo Gyarmati, Haewoon Kwak, and Pablo Rodriguez. Searching for a unique style in soccer. *CoRR*, abs/1409.0308, 2014.

[29] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009.

[30] Michael Horton, Joachim Gudmundsson, Sanjay Chawla, and Joël Estephan. Classification of passes in football matches using spatiotemporal data. *ACM Transactions on Spatial Algorithms and Systems*, 3, 07 2014.

[31] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013.

[32] Marios Kapsalis, Spyridon Plakias, Angelos Kyranoudis, Artemis Zarkadoula, Aikaterini Lathoura, and Themistoklis Tsatalas. Exploring the impact of possession-based performance indicators on goal scoring in elite football leagues. *Journal of Physical Education and Sport*, 23(8):2004 – 2015, 2023.

[33] Matthias Kempe, Martin Vogelbein, and Stephan Nopp. The cream of the crop: Analysing fifa world cup 2014 and germany's title run. *Journal of Human Sport and Exercise*, 11(1):42–52, 2016.

[34] Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, 2008.

[35] Scott Mclean, Paul M. Salmon, Adam D. Gorman, Nicholas J. Stevens, and Colin Solomon. A social network analysis of the goal scoring passing networks of the 2016 european football championships. *Human Movement Science*, 57:400–408, 2018.

[36] Murilo Merlin. *New approach to analyse the passing in soccer matches using multivariate and machine learning techniques*. PhD thesis, Universidade Estadual de Campinas, Faculdade de Educação Física, Campinas, SP, 2020.

[37] Murilo Merlin, Allan Pinto, Alexandre Gomes de Almeida, Felipe Arruda Moura, Ricardo da Silva Torres, and Sergio Augusto Cunha. Classification and determinants of passing difficulty in soccer: a multivariate approach. *Science and Medicine in Football*, 6(4):483–493, 2022. PMID: 36412184.

[38] Raúl Montoliu, Raúl Martín-Félez, Joaquín Torres-Sospedra, and Adolfo Martínez-Usó. Team activity recognition in association football using a bag-of-words-based method. *Human Movement Science*, 41:165–178, 2015.

[39] Gibson Moreira Praça, André Luiz Braga Jacinto, Guilherme de Sousa Pinheiro, Cristóvão de Oliveira Abreu, and Varley Teoldo da Costa. What are the key performance indicators related to winning matches in the german bundesliga? *International Journal of Performance Analysis in Sport*, 23(4):284 – 295, 2023.

[40] Joel Oberstone. Differentiating the top english premier league football clubs from the rest of the pack: Identifying the keys to success. *Journal of Quantitative Analysis in Sports*, 5:10–10, 01 2009.

[41] Pedro Passos, Keith Davids, Duarte Araújo, Natasha Sophia C Paz, J. Minguéns, and Joana Mendes. Networks as a novel tool for studying team ball sports as complex social systems. *Journal of science and medicine in sport*, 14(2):170–6, 2011.

[42] Dominik Raabe, Reinhard Nabben, and Daniel Memmert. Graph representations for the analysis of multi-agent spatiotemporal sports data. *Applied Intelligence*, 53(4):3783–3803, 2023.

[43] Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93, 2022.

[44] Markel Rico-González, Jose Pino Ortega, Amaia Mendez, Filipe Clemente, and Arnold Baca. Machine learning application in soccer: A systematic review. *Biology of Sport*, 40:249–263, 01 2023.

[45] Garry Robins and Malcolm Alexander. Small worlds among interlocking directors: Network structure and distance in bipartite graphs. *Computational & Mathematical Organization Theory*, 10(1):69–94, 2004.

[46] Daniele C. Uchoa Maia Rodrigues, Felipe A. Moura, Sergio Augusto Cunha, and Ricardo da S. Torres. Graph visual rhythms in temporal network analyses. *Graphical Models*, 103:101021, 2019.

[47] Daniele Cristina Uchoa Maia Rodrigues. *Complex network measurements in graph-based spatio-temporal soccer match analysis*. PhD thesis, Universidade Estadual de Campinas, Instituto de Computação, Campinas, SP, 2017.

[48] John W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5(2):99–114, 1949.

[49] Daniele C. Uchoa Maia Rodrigues, Felipe A. Moura, Sergio Augusto Cunha, and Ricardo da S. Torres. Graph visual rhythms in temporal network analyses. *Graphical Models*, 103:101021, 2019.

[50] Nicolò Oreste Pinciroli Vago, Yuri Lavinas, Daniele Rodrigues, Felipe Moura, Sergio Cunha, Claus Aranha, and Ricardo da Silva Torres. Integra: An open tool to support graph-based change pattern analyses in simulated football matches. In *Proceedings - European Council for Modelling and Simulation, ECMS*, page 228 – 234, 2020.

[51] Peter Xenopoulos and Claudio Silva. Graph neural networks to predict sports outcomes. In *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021*, page 1757 – 1763, 2021.