




ARTIGO DE PESQUISA/RESEARCH PAPER


Modelagem de Passes no Futebol via Componentes do Expected Possession Value com Redes Totalmente Convolucionais


Football Pass Modeling via Expected Possession Value Components with Fully Convolutional Networks


Rafael Gomes   [Universidade Federal de Minas Gerais | rafaelgomes@dcc.ufmg.br]


Diogo Chaves  [Universidade Federal de Minas Gerais | diogochaves@dcc.ufmg.br]

João Marcos Campos  [Universidade Federal de Minas Gerais | joaomarcoscampos@dcc.ufmg.br]

Hugo Rios-Neto  [Universidade Federal de Minas Gerais | hugoriosneto@dcc.ufmg.br]

Adriano Pereira  [Universidade Federal de Minas Gerais | adrianoc@dcc.ufmg.br]

Wagner Meira Jr.  [Universidade Federal de Minas Gerais | meira@dcc.ufmg.br]

 Departamento de Ciência da Computação - Universidade Federal de Minas Gerais, Rua Reitor Pires Albuquerque, ICEx, Pampulha, Belo Horizonte, MG, 31270-901, Brasil.

Resumo. A avaliação de ações individuais no futebol é dificultada pela escassez de gols frente ao volume de eventos observados. O *framework* de *Expected Possession Value* (EPV) aborda esse problema decompondo o valor da posse em subcomponentes de passes, conduções e chutes, mas sua adaptação a novas bases de dados está longe de ser trivial. Neste trabalho, adaptamos os componentes de passe do EPV a 52 partidas da Premier League (2022–2023), construindo um *pipeline* de representação que codifica cada estado de jogo em 17 canais espaciais e os alimenta à arquitetura SoccerMap. Mesmo com cerca de $12\times$ menos partidas que o estudo original, o modelo de sucesso produz superfícies com boa calibração e que superam um *baseline* baseado em distância, enquanto o modelo de seleção posiciona o destino real entre os 10 pixels mais prováveis em 67,6% dos casos dentre 7.072 candidatos (*rank* mediano 3). Exploramos ainda essas superfícies por meio de uma análise *what-if* que perturba a posição do jogador-alvo em passes mal-sucedidos, revelando ajustes de posicionamento possíveis e demonstrando o potencial prático do modelo para revisão tática pós-jogo.

Abstract. Evaluating individual actions in soccer is hindered by the low frequency of goals relative to the volume of observed events. The Expected Possession Value (EPV) framework addresses this by decomposing possession value into subcomponents associated with passes, ball drives, and shots, but adapting it to new datasets is far from straightforward. We adapt the EPV pass components to 52 Premier League matches (2022–2023), building a representation pipeline that encodes each game state into 17 spatial channels and feeds them into the SoccerMap architecture. Even with roughly $12\times$ fewer matches than the original study, the success model yields well-calibrated surfaces that outperform a distance-only baseline, while the selection model ranks the true pass destination among the top 10 pixels in 67.6% of cases out of 7.072 candidates (median rank 3). We further exploit these surfaces through a what-if analysis that perturbs the target player's position on failed passes, revealing actionable positioning adjustments and demonstrating the model's practical value for post-match tactical review.

Palavras-chave: Aprendizado Profundo, Análise Esportiva, Futebol, Passes, Redes Neurais Convolucionais, EPV, Análise de Sensibilidade Posicional.

Keywords: Deep Learning, Sports Analytics, Football, Soccer, Passes, Convolutional Neural Networks, EPV, Positional Sensitivity Analysis

Recebido/Received: 16 June 2026 • Aceito/Accepted: 17 June 2026 • Publicado/Published: 10 July 2026

1 Introdução

Avaliar o impacto de ações individuais no futebol permanece um problema desafiador, sobretudo porque os eventos que definem o placar são raros em comparação ao grande volume de ações intermediárias que compõem uma posse. Nesse contexto, os passes ocupam papel central: constituem o principal mecanismo de progressão ofensiva e influenciam a criação de oportunidades muito antes da finalização. Modelar passes de forma contextualizada é, portanto, uma etapa fundamental para estimar o valor de uma posse de bola.

O *framework* de *Expected Possession Value* (EPV) Fernández *et al.* [2021] constitui uma das formulações mais relevantes para esse problema ao decompor o valor da posse

em componentes associados a passes, conduções e chutes, combinando dados de evento, dados de rastreamento e modelagem probabilística. Em particular, os componentes de passe destacam-se pelo uso da arquitetura SoccerMap Fernández and Bornn [2020], capaz de produzir superfícies completas de predição sobre o campo e, assim, oferecer uma representação tática mais rica das opções disponíveis em cada jogada.

Apesar de sua relevância, a aplicação do EPV em novas bases de dados não é imediata. Sua adaptação exige sincronização entre eventos e *frames*, padronização da orientação do jogo, reorganização dos jogadores em papéis ofensivos e defensivos e construção de uma representação matricial compatível com a arquitetura. O principal desafio não está apenas

na reutilização da rede, mas na transformação de dados brutos heterogêneos em uma entrada capaz de preservar a estrutura tática da jogada.

Este trabalho adapta os componentes de passe do EPV para uma nova base da Premier League (temporada 2022–2023), fornecida pela Gradient Sports e composta por dados de evento e rastreamento de 52 partidas — cerca de $12\times$ menos do que as 633 utilizadas no estudo original. A principal contribuição está na construção de um *pipeline* de representação que converte o estado de jogo em 17 canais matriciais, combinando ocupação espacial de atacantes e defensores com informações geométricas e cinemáticas. Desse modo, o estudo vai além da reaplicação de uma arquitetura conhecida e mostra como torná-la operacional em um ambiente de dados distinto e mais restrito.

Os resultados indicam que a adaptação é promissora. O modelo de sucesso produz superfícies probabilísticas coerentes, cuja média predita coincide com a taxa observada no conjunto de teste e coerentes com a lógica tática esperada, superando um *baseline* baseado em distância euclidiana em todas as métricas avaliadas. O modelo de seleção, apesar de gerar distribuições visualmente concentradas, posiciona o destino real entre os 10 pixels de maior probabilidade em 67,6% dos casos, com *rank* mediano de 3 dentre 7.072 posições candidatas. Além dos resultados preditivos, exploramos as superfícies de sucesso por meio de uma análise de sensibilidade posicional do tipo *what-if*: ao perturbar a posição do jogador-alvo em passes mal-sucedidos, o modelo sugere que, sob uma perturbação local e mantendo as demais condições fixas, determinadas posições alternativas seriam associadas a maior probabilidade predita de sucesso, demonstrando seu potencial como ferramenta de análise tática pós-jogo.

Especificamente, as contribuições deste trabalho são:

1. a adaptação de um *pipeline* de pré-processamento para sincronizar dados de evento e rastreamento e convertê-los em uma entrada compatível com a SoccerMap;
2. a proposição de uma representação espaço-temporal em 17 canais matriciais, enriquecida com informações posicionais, geométricas e cinemáticas;
3. a implementação e avaliação quantitativa dos modelos de probabilidade de sucesso e de seleção de passes em 52 partidas da Premier League;
4. uma análise de sensibilidade posicional (*what-if*) que explora as superfícies de sucesso de forma contrafactual, perturbando a posição do jogador-alvo em passes mal-sucedidos e quantificando o ganho de probabilidade associado.

2 Trabalhos Relacionados

2.1 Valoração de ações com dados de evento

Abordagens iniciais de valoração apoiaram-se em dados de evento — registros discretos contendo tipo, localização e resultado de cada ação. O valor esperado de posses foi formulado por meio de cadeias de Markov sobre estados discretos Rudd [2011], ideia estendida pela métrica *Expected Threat* (xT), que deriva o valor de transições entre localizações de uma grade discretizada do campo Singh [2019]. A probabilidade de finalização nos 10 segundos seguintes

a um passe foi utilizada como estimativa de seu risco e recompensa Power *et al.* [2017], enquanto o *framework* VAEP generalizou a valoração ao estimar o impacto de qualquer tipo de ação sobre a probabilidade de gol a partir de sequências de eventos Decroos *et al.* [2019]. Embora relevantes, essas abordagens compartilham representações simplificadas do estado de jogo, sem incorporar a dinâmica simultânea dos 22 jogadores e da bola.

2.2 Modelos com dados de rastreamento e deep learning

A disponibilidade de dados de rastreamento óptico — posições de todos os jogadores e da bola a 10–25 Hz — viabilizou modelos de maior resolução espaço-temporal. No contexto específico de passes, um modelo que combina a trajetória inicial da bola com os vetores de movimento de todos os jogadores atingiu AUC de 93,4% na estimativa de probabilidade de conclusão Anzer and Bauer [2022]. No basquete, o EPV de posses foi estimado por um modelo estocástico multirresolução que combina macro e micro-transições Cervone *et al.* [2016], servindo de inspiração para a formulação no futebol.

O avanço mais abrangente nesse campo é o *framework* de *Expected Possession Value* (EPV) Fernández *et al.* [2021], que formula o valor da posse como um processo de decisão markoviano decomposto em passes, conduções e chutes. Para os componentes de passe, o *framework* emprega a arquitetura SoccerMap — uma rede totalmente convolucional que processa camadas matriciais em três escalas de resolução ($1\times$, $1/2\times$, $1/4\times$) e aprende superfícies completas a partir de uma única correspondência *ground-truth* por exemplo Fernández and Bornn [2020]. Treinado com 633 partidas da Premier League (2013–2015), o modelo demonstrou estimativas calibradas para todos os componentes e possibilitou aplicações como análise de estratégias de pressão e mapeamento de relações de valor entre jogadores.

O presente trabalho replica os componentes de passe desse *framework* com um subconjunto significativamente menor de partidas da mesma liga, porém de uma temporada mais recente (2022–2023), investigando a robustez da abordagem sob restrições de volume de dados.

3 Metodologia

A principal dificuldade deste trabalho não está apenas em treinar modelos sobre uma nova base, mas em tornar operacional, em um ambiente de dados distinto, um *framework* originalmente concebido para outra escala e outra organização de informações. Para isso, foi necessário construir um *pipeline* completo de adaptação, capaz de sincronizar dados de evento e rastreamento, padronizar a orientação das jogadas, reorganizar os jogadores segundo o contexto ofensivo e defensivo da posse e, por fim, converter cada situação de passe em uma representação matricial compatível com a SoccerMap. Essa etapa é central para o artigo, pois é ela que viabiliza a transferência do EPV para uma base menor e estruturalmente diferente, preservando o máximo possível da informação tática contida no estado de jogo.

3.1 Dados

Os dados foram fornecidos pela Gradient Sports e compreendem 52 partidas da Premier League inglesa, temporada

2022–2023. Para cada partida estão disponíveis *dados de evento* (tipo, jogador, localização, resultado) e *dados de rastreamento* (posições bidimensionais de todos os jogadores e da bola, com vetores de velocidade e aceleração). A sincronização entre ambos permite associar a cada passe um *snapshot* completo do estado espaço-temporal do jogo.

O conjunto foi dividido em 45 jogos para treino e validação (80%/20%) e 7 jogos para teste.

3.2 Pré-processamento

O pré-processamento compreendeu três etapas. Primeiro, os dados de evento de cada partida foram filtrados para reter apenas passes, com seus atributos essenciais. Em seguida, para cada passe extraiu-se o *frame* de rastreamento correspondente ao instante de início do evento, contendo identificadores, coordenadas, velocidades e acelerações de todos os jogadores e da bola. Os identificadores (tanto do jogador em si quanto do seu time) foram normalizados sequencialmente e, ao comparar o time do jogador com o time do portador da bola no momento do *frame*, os jogadores foram rotulados como time atacante ou defensor. Por fim, todas as coordenadas foram espelhadas para que o time com posse ataque sempre da esquerda para a direita, garantindo comparabilidade entre exemplos.

3.3 Formalização matemática

Seguimos a formulação original do EPV Fernández *et al.* [2021]; seja $G \in \{1, -1, 0\}$ o resultado de uma posse (gol marcado, gol sofrido, fim do tempo) e T_t o *snapshot* espaço-temporal no instante t . O valor esperado da posse é:

$$EPV_t = \mathbb{E}[G | T_t] \quad (1)$$

O EPV é decomposto segundo as ações possíveis $A = \{\rho, \delta, \varsigma\}$ (passe, condução, chute):

$$EPV_t = \sum_{a \in A} \mathbb{P}(A=a | T_t) \mathbb{E}[G | A=a, T_t] \quad (2)$$

Focamos no componente de passe ($A=\rho$). Seja $D_t \in L$ a localização de destino, com L uma grade 104×68 . O valor esperado do passe integra sobre todas as localizações possíveis:

$$\mathbb{E}[G | A=\rho, T_t] = \quad (3)$$

$$\int_L \mathbb{P}(D_t | A=\rho, T_t) \mathbb{E}[G | A=\rho, D_t, T_t] dD_t \quad (4)$$

O termo $\mathbb{E}[G | A=\rho, D_t, T_t]$ é decomposto pelo resultado do passe $O_\rho \in \{0, 1\}$:

$$\mathbb{E}[G | A=\rho, D_t, T_t] = \quad (5)$$

$$\sum_{o \in \{0,1\}} \mathbb{P}(O_\rho=o | D_t, T_t) \mathbb{E}[G | O_\rho=o, D_t, T_t] \quad (6)$$

Essa decomposição origina os componentes implementados: probabilidade de sucesso $\mathbb{P}(O_\rho=1 | D_t, T_t)$ e probabilidade de seleção $\mathbb{P}(D_t | A=\rho, T_t)$.

3.4 Representação do estado de jogo

Cada situação de passe é codificada em um tensor de $17 \times 104 \times 68$ canais matriciais (Tabela 1). A representação combina ocupação espacial de atacantes e defensores, distâncias à bola e ao gol, decomposição angular da relação bola–gol, velocidades por eixo de jogadores, bola e portador, contagens de jogadores entre a bola e cada célula e distância à origem do evento.

Apesar de o trabalho original de Fernández *et al.* [2021] conter 13 canais, decidimos abrir mão de alguns e adicionar outros, considerando que, aqui, o objetivo final era um modelo voltado apenas para passes. Nesse sentido, o conjunto de 17 canais mantém os sinais mais informativos e estáveis extraídos diretamente do tracking: ocupação dos jogadores, distâncias à bola e ao gol, geometria básica, movimento e medidas gerais de pressão/espaco.

Assim, o modelo recebe informação suficiente para avaliar qualidade e disponibilidade de passes, sem depender de muitas *features* direcionais derivadas, que seriam mais complexas e sensíveis a ruído.

Tendo isso em vista, esses canais codificam não apenas a disposição dos jogadores, mas como o espaço está sendo ocupado e bloqueado no instante da decisão, estabelecendo a ponte entre os dados brutos e uma modelagem campo-inteiro compatível com a SoccerMap.

Tabela 1. Canais matriciais da representação do estado de jogo.

#	Categoria	Canal
1	Localização	Posição do time atacante
2	Localização	Posição do time defensor
3	Distância	Distância à bola
4	Distância	Distância ao gol
5	Ângulo	$\cos \theta$ (bola–gol)
6	Ângulo	$\sin \theta$ (bola–gol)
7	Ângulo	Ângulo ao gol
8	Velocidade	Velocidade da bola — eixo x
9	Velocidade	Velocidade da bola — eixo y
10	Contagem	Atacantes entre bola e célula
11	Contagem	Defensores entre bola e célula
12	Velocidade	Velocidade do portador da bola
13	Velocidade	Velocidade dos atacantes — eixo x
14	Velocidade	Velocidade dos atacantes — eixo y
15	Velocidade	Velocidade dos defensores — eixo x
16	Velocidade	Velocidade dos defensores — eixo y
17	Distância	Distância ao início do evento

3.5 Arquitetura dos modelos

Os três modelos compartilham o bloco de extração de *features* da arquitetura SoccerMap Fernández and Bornn [2020], uma rede totalmente convolucional. O tensor $17 \times 104 \times 68$ passa por camadas convolucionais com 32 e 64 filtros (5×5 , *stride* 1) na escala original ($1 \times$) e é subamostrado por *max-pooling* para as escalas $1/2 \times$ e $1/4 \times$. Predições intermediárias em cada escala são fundidas por *upsampling* e camadas de fusão linear, produzindo uma superfície $104 \times 68 \times 1$. Os modelos diferem apenas no bloco de predição e na função de perda.

3.5.1 Probabilidade de sucesso do passe

Estima $\mathbb{P}(O_\rho=1 \mid D_t, T_t)$. Uma sigmoide é aplicada elemento a elemento à saída do SoccerMap:

$$\hat{P}_{ij} = \sigma(z_{ij}) = \frac{e^{z_{ij}}}{e^{z_{ij}} + 1} \quad (7)$$

A perda é a entropia cruzada binária calculada no pixel de destino d do passe observado:

$$\mathcal{L}_{\text{suc}} = -[y \log \hat{P}_d + (1 - y) \log(1 - \hat{P}_d)] \quad (8)$$

onde $y \in \{0, 1\}$ é o resultado observado. O gradiente dessa perda escalar é retropropagado por toda a rede, aprendendo a superfície completa a partir de uma única correspondência *ground-truth*.

3.5.2 Probabilidade de seleção do passe

Estima $\mathbb{P}(D_t \mid A=\rho, T_t)$. A matriz 104×68 é vetorizada e a função *softmax* é aplicada:

$$\hat{S}_k = \frac{e^{v_k}}{\sum_{j=1}^{W \cdot H} e^{v_j}}, \quad k = 1, \dots, W \cdot H \quad (9)$$

garantindo $\sum_k \hat{S}_k = 1$. A perda é $\mathcal{L}_{\text{sel}} = -\log \hat{S}_d$.

3.5.3 Valor esperado da posse a partir de passes

Estima $E[G \mid A = \rho, O_\rho, D_t, T_t]$. Reutiliza pesos pré-treinados do modelo de sucesso. A saída é mapeada ao intervalo $[-1, 1]$ por $\hat{V}_{ij} = 2\sigma(z_{ij}) - 1$, e a perda é o erro quadrático médio:

$$\mathcal{L}_{\text{epv}} = (r - \hat{V}_d)^2 \quad (10)$$

onde r é a recompensa de longo prazo associada ao resultado da posse Fernández *et al.* [2021]. A integração deste componente ao *pipeline* constitui extensão direta do presente trabalho e será abordada em continuidade futura.

3.6 Análise de sensibilidade posicional

Para avaliar a coerência tática das superfícies aprendidas e demonstrar uma aplicação prática do modelo de sucesso, propomos uma análise do tipo *what-if* sobre passes mal-sucedidos. O procedimento consiste em três etapas.

Primeiro, para cada passe falho do conjunto de análise, identifica-se o jogador-alvo e sua posição no instante do passe. Em seguida, essa posição é perturbada em uma vizinhança de aproximadamente 3 a 6 metros ao redor da posição projetada pelo vetor de velocidade corrente do jogador, gerando um conjunto de posições alternativas candidatas. Para cada posição alternativa, o tensor de estado de jogo completo ($17 \times 104 \times 68$) é reconstruído — atualizando todos os canais derivados, como distâncias, contagens de jogadores ultrapassados e ocupação espacial — e o modelo de sucesso é reavaliado. A probabilidade de sucesso na nova posição do alvo é então comparada à probabilidade original.

Casos em que o reposicionamento produz aumento superior a 10 pontos percentuais na probabilidade predita são selecionados como exemplos de posicionamento subótimo, permitindo identificar ajustes locais que teriam favorecido a conclusão do passe. A perturbação mantém o vetor de velocidade original do jogador e não altera as posições dos demais agentes, constituindo uma aproximação de primeira ordem que isola o efeito do posicionamento do receptor.

4 Resultados e Discussão

4.1 Avaliação quantitativa

Para o modelo de sucesso, reportamos métricas probabilísticas avaliadas no pixel de destino de cada passe e comparamos com um *baseline* de regressão logística treinado unicamente sobre a distância euclidiana do passe (Tabela 2). Para o modelo de seleção, adotamos acurácia top- k e *rank* do destino real, métricas mais adequadas quando o espaço de saída compreende 7.072 posições (Tabela 3). Ambas as avaliações utilizam o conjunto de teste (7 partidas, 6 973 passes).

Tabela 2. Avaliação probabilística do modelo de sucesso. O *baseline* utiliza regressão logística sobre a distância euclidiana do passe.

Métrica	SoccerMap	Baseline
AUC-ROC (\uparrow)	0,622	0,588
Brier Score (\downarrow)	0,138	0,141
Log-Loss (\downarrow)	0,444	0,456
Average Precision (\uparrow)	0,876	0,850

O modelo supera o *baseline* em todas as métricas. O AUC-ROC de 0,622 reflete a dificuldade intrínseca de discriminação em um cenário com taxa de sucesso de 82,9%, onde o espaço para separação entre classes é comprimido. A calibração em agregado é precisa: a probabilidade média predita (0,830) coincide com a taxa de sucesso observada (0,829). Ressalta-se que o valor principal do modelo reside nas superfícies campo-inteiro, cuja coerência tática é avaliada nas seções seguintes, e não apenas na predição pontual.

4.2 Superfícies de probabilidade de sucesso

A Figura 1 apresenta a superfície gerada para uma situação representativa do conjunto de teste. Nela, passador é o jogador do time de ataque marcado em preto; a superfície representa a distribuição de probabilidade de sucesso do passe mostrado ao redor do campo, sendo o valor de probabilidade representado por cores. Cores mais quentes, como vermelho, representam probabilidades mais altas, enquanto cores frias, como azul, representam probabilidades mais baixas, conforme mostrado pela barra lateral. As setas representam o movimento dos jogadores, sendo a direção correspondente à direção do jogador e o comprimento referente à velocidade.

A superfície preserva padrões taticamente esperados: corredores quentes emergem no entorno de companheiros em melhores condições de apoio, enquanto regiões ocupadas ou protegidas pela linha defensiva permanecem frias. O mapa também diferencia com clareza opções curtas de passes longos que atravessam zonas congestionadas, indicando que a representação proposta transferiu para a nova base relações espaço-temporais fundamentais do jogo.

4.3 Análise de sensibilidade posicional

A Figura 2 ilustra um exemplo da análise *what-if* descrita na Seção 3.6, aplicada a um passe malsucedido do conjunto de teste. No cenário original (esquerda), o jogador-alvo encontra-se em zona de cobertura defensiva densa e a probabilidade de sucesso predita é baixa. Ao deslocar o alvo poucos metros em direção a um espaço entre linhas (direita), o tensor é reconstruído e o modelo reavaliado: a probabilidade de sucesso

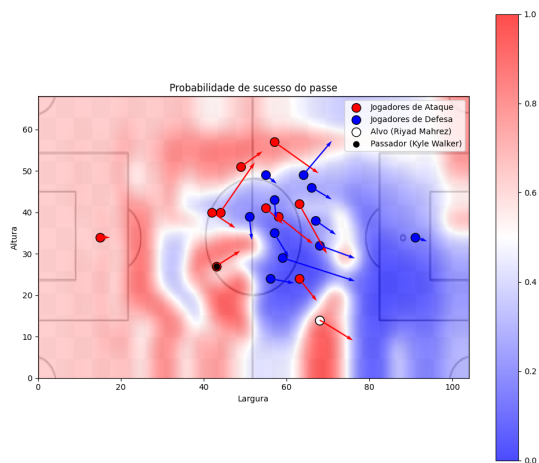


Figura 1. Superfície de probabilidade de sucesso do passe para uma jogada do conjunto de teste. A malha prevista acompanha a ocupação ofensiva e defensiva do lance e evidencia como a chance de conclusão varia ao longo do campo.

sobe de forma expressiva, consistente com a saída da zona de marcação.

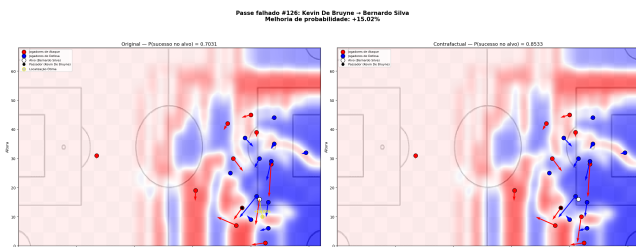


Figura 2. Análise contrafactual para um passe falhado de Kevin De Bruyne em direção a Bernardo Silva. À esquerda, o cenário original; à direita, o reposicionamento local sugerido pela superfície. Mantidas as demais condições do lance, a probabilidade estimada de sucesso no alvo aumenta de 0,7031 para 0,8533, evidenciando o potencial da abordagem para diagnósticos táticos mais finos.

Esse comportamento, observado sistematicamente nos casos analisados, evidencia que as superfícies aprendidas são sensíveis ao posicionamento local dos jogadores e produzem variações de probabilidade coerentes com a lógica tática, reforçando sua utilidade como ferramenta de análise pós-jogo.

4.4 Probabilidade de seleção do passe

Tabela 3. Acurácia top-*k* e rank do modelo de seleção (7.072 pixels por superfície).

Métrica	Valor
Top-1	37,6%
Top-5	58,1%
Top-10	67,6%
Top-50	84,3%
Top-100	89,9%
Rank médio	56 / 7.072
Rank mediano	3 / 7.072

O modelo produz distribuições concentradas sobre o campo, comportamento esperado dado que a *softmax* opera sobre 7.072 elementos. A avaliação por top-*k*, mostrada na Tabela 3, exhibe as frequências nas quais o pixel correto aparece nas *k* maiores probabilidades retornadas pelo modelo. Ela mostra que essa concentração é altamente informativa: o

destino real situa-se entre os 10 pixels mais prováveis em 67,6% dos passes, com posição mediana no ranqueamento de apenas 3.

Esses resultados demonstram a robustez do modelo simplificado: mesmo treinado com volume de dados significativamente menor e sem as *features* contextuais definidas no trabalho original, a grande maioria das previsões situa-se no top-1% das probabilidades geradas. A aplicação de *temperature scaling* pós-treinamento resultou em temperatura ótima $T = 1,09$, valor próximo de 1, confirmando que os logits já se encontram adequadamente calibrados — diferentemente do fator $T = 0,5$ reportado no estudo original Fernández *et al.* [2021], que dispunha de maior volume de dados e, portanto, superfícies de logits com maior amplitude a ser corrigida.

4.5 Discussão

O modelo de sucesso mostrou-se resiliente à redução de dados: com cerca de 1/12 do volume de partidas e em uma temporada quase uma década mais recente, a SoccerMap manteve superfícies taticamente coerentes e calibração precisa. A mesma liga (Premier League) permite isolar o volume como principal variável, sugerindo que as relações espaço-temporais fundamentais do jogo permanecem estáveis. O modelo de seleção revelou desempenho superior ao sugerido pela inspeção visual das superfícies: o rank mediano de 3 indica que a rede aprendeu a localizar a região de destino com alta precisão, mesmo com volume reduzido. Ainda assim, a representação proposta não incorpora *features* contextuais presentes no *framework* original, como linhas de pressão dinâmicas e *pitch control* Fernández and Bornn [2018], cuja inclusão constitui direção natural de extensão.

5 Conclusão

Este trabalho adaptou os componentes de passe do *framework* de *Expected Possession Value* Fernández *et al.* [2021] a 52 partidas da Premier League (2022–2023), construindo um *pipeline* capaz de converter dados brutos de evento e rastreamento em uma representação espaço-temporal de 17 canais matriciais compatível com a arquitetura SoccerMap Fernández and Bornn [2020].

Os resultados demonstram que a adaptação é viável mesmo com uma redução de cerca de 12× no volume de dados. O modelo de sucesso produziu superfícies bem calibradas e taticamente coerentes, superando um *baseline* baseado em distância. O modelo de seleção, avaliado por acurácia top-*k*, posicionou o destino real entre os 10 pixels mais prováveis em 67,6% dos passes, com rank mediano de 3 dentre 7.072 candidatos. A análise de sensibilidade posicional mostrou que as superfícies aprendidas respondem de forma coerente a perturbações no posicionamento do jogador-alvo, evidenciando potencial concreto para análise tática pós-jogo.

Como trabalhos futuros, destacam-se a conclusão do modelo de valor esperado da posse a partir de passes, a incorporação de *features* contextuais como *pitch control* Fernández and Bornn [2018] e linhas de pressão dinâmicas, e a ampliação da base de dados para avaliar a curva de aprendizado dos modelos.

Declarações complementares

Agradecimentos

Agradecemos à Gradient Sports pela cessão dos dados e aos monitores e colaboradores **Thiago Costa Porto**, **Gabriel Reis** e **Gabriel Valadão** pelas ideias e discussões.

Financiamento

Este trabalho é parcialmente financiado por CNPq, CAPES, FAPESP e projeto IAIA-INCT em IA.

Contribuições dos autores

A concepção do estudo e a definição dos objetivos principais foram realizadas por Rafael Gomes e Diogo Chaves. A metodologia foi desenvolvida por Rafael Gomes e Diogo Chaves, sob a orientação de Hugo Rios-Neto, que também supervisionou tecnicamente o projeto. Os experimentos foram conduzidos por Rafael Gomes, Diogo Chaves e João Marcos Campos, e as análises de dados foram realizadas conjuntamente por Rafael Gomes e João Marcos Campos. A aquisição de financiamento foi responsabilidade de Adriano Pereira e Wagner Meira Jr., que também coordenaram o projeto. Todos os autores leram e aprovaram o manuscrito final.

Conflitos de interesse

Os autores declaram que não têm nenhum conflito de interesses.

Disponibilidade de dados e materiais

Os conjuntos de software gerados durante o estudo atual estão disponíveis em: https://github.com/Dtchaves/Simplified_EPV_in_PFF_Data.

Os conjuntos de dados analisados durante o estudo atual serão disponibilizados mediante solicitação.

Referências

- Anzer, G. and Bauer, P. (2022). Expected passes. *Data Mining and Knowledge Discovery*. DOI: 10.1007/s10618-021-00810-3.
- Cervone, D., D'Amour, A., Bornn, L., and Goldsberry, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*.
- Decroos, T., Bransen, L., Haaren, J. V., and Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. DOI: 10.1145/3292500.3330758.
- Fernández, J. and Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *MIT Sloan Sports Analytics Conference*.
- Fernández, J. and Bornn, L. (2020). Soccermap: A deep learning architecture for visually-interpretable analysis in soccer.
- Fernández, J., Bornn, L., and Cervone, D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*. DOI: 10.1007/s10994-021-05989-6.
- Power, P., Ruiz, H., Wei, X., and Lucey, P. (2017). Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. DOI: 10.1145/3097983.3098051.

Rudd, S. (2011). A framework for tactical analysis and individual offensive production assessment in soccer using markov chains. In *New England Symposium on Statistics in Sports*.

Singh, K. (2019). Introducing expected threat (xT). Blog post.