open access

Published online on: https://journal.uii.ac.id/jurnalsnati/

Check for updates

DOI: 10.20885/snati.v4.i1.3

JURNAL SNATI
Sains, Nalar, dan Aplikasi Teknologi Informasi

Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi

Vol. 4 No. 1 (2025) 18 - 25 ISSN Media Electronic: 2807-5935

Forsyth-Edwards Notation in Chess Game Clustering: A Depth-Based Evaluation

Feri Wijayanto
Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia
feri.wijayanto@uii.ac.id

Abstract

Chess games clustering poses the challenge of accurately grouping games with similar strategies and positions, especially when the openings are similar. Previous research has used Portable Game Notation (PGN) as a feature for clustering, but its emphasis on move order can limit position transposition. This research addresses this limitation by evaluating Forsyth-Edwards Notation (FEN), which focuses on board position, as an alternative. Hierarchical clustering with complete linkage and K-means clustering were used to analyze 100 chess games at move depths of 20, 30, 40, and 60. Both methods effectively cluster games involving the English Opening and the Queen's Gambit Declined, with FEN providing slightly better differentiation than PGN. However, challenges remain in grouping French Defence variations, especially the Poulsen Attack and variations with 6.a3, due to positional similarities, as both variations share the same eight first moves. This study underlines the robustness of FEN for clustering tasks and its compatibility with hierarchical clustering, highlighting the important role of move depth. The results provide a basis for refining clustering methods and using larger data sets to deepen insights into chess strategies.

Keywords: *clustering analysis; text analysis; hierarchical clustering; k-means*

1. Introduction

Chess is a strategic game deeply rooted in intellect and complex problem-solving. Over centuries, it has evolved not only as a competitive sport but also as a subject of study to unravel the patterns and strategies behind successful gameplay. To effectively study these patterns, a structured system of documentation is essential. This compilation of game records serves as a database that can be utilized for analysis, enabling deeper insights into the strategies and tactics of chess. Forsyth-Edwards Notation (FEN) and Portable Game Notation (PGN) are two notations that are commonly used to document chess games. While PGN records the sequence of moves in a game, providing a dynamic view of the gameplay, FEN captures the static arrangement of pieces on the board at a specific point in time. This fundamental difference highlights their complementary nature: PGN is ideal for studying the progression of a game, whereas FEN is better suited for analyzing positional setups and snapshots of game states [1].

The central challenge in clustering chess games lies in accurately grouping games with similar strategies and positions, particularly when openings have inherent similarities. Previous research has primarily used PGN data to cluster chess games, emphasizing move sequences to categorize games based on opening strategies and move patterns. Techniques such as hierarchical clustering and K-means clustering have shown effectiveness in analyzing openings like the French Defence, Queen's Gambit Declined, and English Opening [2,3,4]. However, the reliance on move sequences in PGN can limit positional differentiation, especially in scenarios involving position transpositions or midgame variations. This limitation highlights the need for exploring alternative representations like FEN, which emphasize static board positions [5, 14].

Clustering, as a machine learning approach, has been widely applied in other fields where text analysis is essential [6]. For example, in natural language processing (NLP), clustering techniques are used to group similar documents, identify sentiment, or classify topics based on text content [7]. Notably, in the analysis of chess games, clustering has been applied within NLP contexts to uncover strategic patterns. The study in [8] employed DBSCAN clustering to classify chess openings based on positional features extracted from move sequences. These features included player ratings, opening codes, and game outcomes, enabling the identification and prediction of strategies tied to specific chess openings. Similarly, in social media analysis, clustering helps to categorize user behaviors or identify emerging trends [9]. In biological research, clustering methods are employed to analyze genetic

Received: 22-12-2024 | Revised: 03-01-2025 | Accepted: 07-01-2025 | Published: 15-01-2025

data or to group similar proteins [10]. These applications illustrate the versatility and effectiveness of clustering in uncovering patterns and relationships within textual and structural data [11]. In this context, applying clustering techniques to FEN data leverages similar principles, aiming to reveal the underlying strategic patterns in chess games as in [2, 12].

In the practice of studying chess openings, players often memorize move sequences or repertoires documented in PGN format. While this approach is useful for learning and replicating specific strategies, the evaluation of optimal moves frequently requires analyzing the board position, which is represented in FEN. Given this, FEN provides a powerful alternative for understanding the positional context of a game. It is hypothesized that FEN, with its emphasis on the static arrangement of pieces, can also serve as an effective feature for clustering chess games [13]. Although chess openings are theoretically categorized based on initial move sequences, subsequent variations can diverge significantly, making each game unique. This variability highlights the potential of clustering analysis to identify games based on their similarities, offering an alternative approach to understanding strategic patterns, even though it has been rarely utilized thus far. This opens up new possibilities for identifying positional patterns and strategic insights across different stages of play.

This study aims to explore the application of clustering techniques to FEN data to address the limitations of PGN in representing positional nuances. By adapting methods such as hierarchical clustering and K-means clustering, we seek to identify positional groupings that reflect tactical and strategic elements of the game [2]. Additionally, this research aims to test the hypothesis that FEN, with its emphasis on static positional representation, can effectively cluster chess games. A comparative analysis will be conducted to evaluate the clustering results using FEN against those derived from PGN, highlighting the strengths and applicability of each format [14]. The use of FEN as a data source introduces unique challenges and opportunities, particularly in representing and analyzing positional information rather than move sequences. Through this research, we aim to expand the scope of chess game analysis, demonstrating the utility of FEN clustering in enhancing our understanding of chess strategies.

2. Research Methods

This section describes the research methods employed to test the hypothesis that Forsyth-Edwards Notation (FEN) can serve as an effective basis for clustering chess games. While the general approach follows methodologies from prior research on PGN data, key adaptations have been made to suit the unique characteristics of FEN. The research methodology used in this study is summarized in Figure 1.

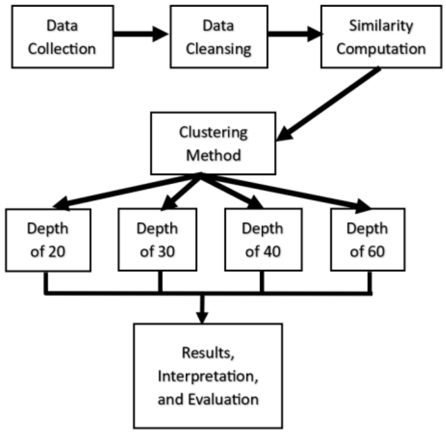

Figure 1. Research Methodology

2.1. Data Collection

The dataset for this study consists of chess games sourced from https://www.chessgames.com/ in PGN format. We use the same data as previous research, which includes 20 chess games for each of the following openings: French Defence (Advance Variation, Steinitz Variation), French Defence (Advance Variation, Paulsen Attack), French Defence (Advance Variation, with 6. a3), Queen’s Gambit Declined, and English Opening (symmetrical variation, main line), for a total of 100 games [5].

Forsyth-Edwards Notation (FEN) is a standard notation used to describe the current state of a chessboard [14]. It encodes the positions of all pieces, the active player, castling rights, en passant target squares, and the halfmove and fullmove counters. The core part of the FEN string, which represents the arrangement of pieces, uses a combination of characters where uppercase letters (e.g., 'R' for rook, 'N' for knight) denote white pieces and lowercase letters (e.g., 'r', 'n') denote black pieces.

To better understand FEN and its relation to PGN, consider the following example:

From a PGN record, a specific move sequence might appear as follows:

1. e4 e6 2. d4 d5 3. e5 c5 4. c3 Nc6 5. Nf3 Qb6 6. a3

This sequence translates into a board state, which in FEN format is represented as:

r1b1kbnr/pp3ppp/1qn1p3/2ppP3/3P4/P1P2N2/1P3 PPP/RNBQKB1R b KQkq - 0 6

Breaking this FEN string down:

a) **r1b1kbnr/pp3ppp/1qn1p3/2ppP3/3P4/P1P2N2 /1P3PPP/RNBQKB1R**: This section encodes the position of pieces on the board.
b) **w**: Indicates it is White's turn.
c) **KQkq**: Castling rights for both sides.
d) **-**: No en passant target square.

e) **0**: Halfmove counter.
f) **6**: Fullmove counter.

The chessboard has 64 squares, 8 rows and 8 columns. The columns are coded from ‘a’ to ‘h’ from left to right, while the rows are coded from ‘1’ to ‘8’ starting from the white pieces. The first part, (a), is used to map the positions of the pieces starting from the eighth row separated by ‘/’ on each row in sequence. The number of empty squares between the pieces is represented by numbers.

Each move in a chess game generates a unique FEN string, as the position of pieces on the board changes with every move [15]. For comprehensive analysis, it is crucial to generate and record the FEN notation for every single move across all games. This detailed recording captures the evolving positional states, providing a robust dataset that reflects the dynamics of each game. The FEN data generated from these moves serves as the primary resource for measuring the similarity between different chess games. By comparing the positional arrangements encoded in the FEN strings, we can identify patterns and group similar games. This forms the basis of the clustering analysis, which aims to reveal deeper strategic and positional trends within the dataset.

2.2. Data Cleansing

Prior to the initiation of analysis, the data underwent a series of preprocessing steps to ensure its compatibility with the requisite format for processing. This preprocessing step entailed the extraction of a specific portion of the FEN strings, which encode the positions of pieces on the chessboard. The characters representing board ranks (‘/’) were removed, resulting in a continuous string that accurately represents the arrangement of pieces. This conditioning step was essential for focusing on positional information and eliminating unnecessary formatting, thereby preparing the data for similarity measurement and clustering analysis. Additionally, the numeric values in the FEN strings, representing empty squares on the chessboard, were replaced with the character 'o', repeatedly, corresponding to the numeric values. This transformation ensures a consistent representation of unoccupied squares, making the FEN strings more uniform and easier to compare during the analysis. With the positional code in FEN using numbers to represent the count of empty squares, the length of the strings would vary between different FENs. However, by replacing these numbers with 'o', the FEN strings achieve a uniform length, facilitating their similarity calculation and comparison in subsequent analysis.

2.3. Similarity Computation

Unlike previous studies that utilized Document-Term Matrix (DTM) representations, this research directly calculates similarity between FEN strings to assess positional resemblance. The similarity measure is based on character-level comparisons, employing the Hamming distance [16] to quantify how closely two FEN strings align. This method is particularly suited for the research because the positional code in FEN is constructed sequentially to represent piece positions, where the order of characters is crucial in accurately depicting these positions—a different sequence signifies a distinct arrangement on the chessboard. Additionally, since positional similarity is evaluated square by square, with each square represented by a character, Hamming distance is the most appropriate calculation method for this scenario.

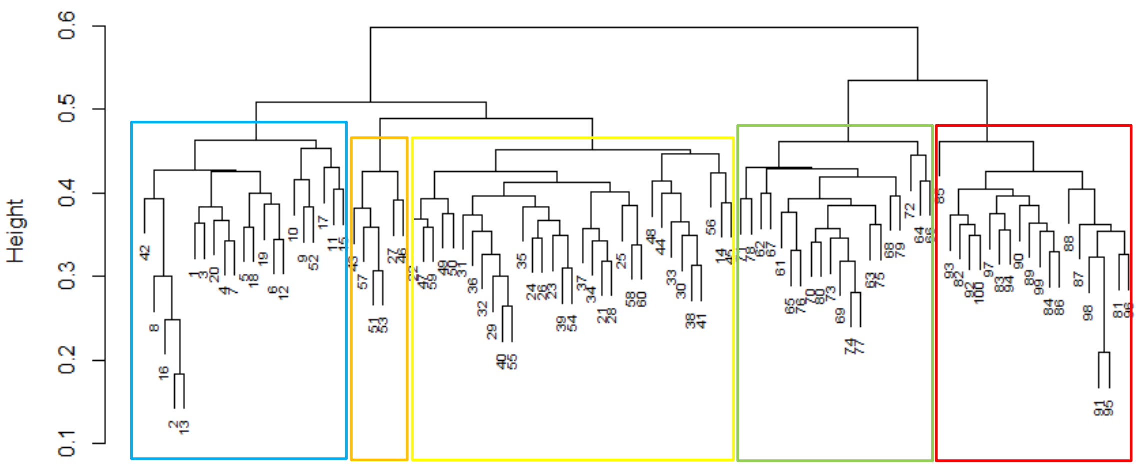

Figure 2. Dendrogram of hierarchical clustering result of 60 moves depth using complete linkage criteria. The blue square are mostly composed of chess games that use French Defence (Advance variation, Steinitz Variation). The orange and yellow squares are mixed between chess games that use French Defence (Advance variation, Paulsen attack and variation with 6. a3). The green square groups games that use the Queen's Gambit Declined opening, and the red square groups games that use the English opening.

To provide a comprehensive similarity metric, the average Hamming distance is calculated across all combinations of FEN pairs from a pair of chess games. This average value reflects the overall similarity of positional changes throughout the game and serves as a robust input for clustering analysis. This approach effectively captures subtle variations in positional configurations, enabling more precise clustering. Furthermore, in this similarity calculation, we experiment with four different step depths: 20, 30, 40, and 60 moves, counting each move as one turn by White or one turn by Black. This variation allows us to evaluate how analyzing different gameplay lengths impacts the clustering results.

2.4 Clustering Algorithm

To cluster the processed FEN data, hierarchical clustering with complete linkage and K-means clustering algorithms were employed due to their effectiveness in grouping data based on similarity metrics and ease of interpretation [17,18]. Hierarchical clustering with complete linkage was employed as the sole method for clustering, as this approach has been proven to provide intuitive and meaningful results. Complete linkage evaluates the distance between clusters by considering the furthest points within them, ensuring that all elements within a cluster maintain a minimum level of similarity to the most distant member. This method helps produce well-separated and interpretable clusters, making it particularly suitable for datasets with distinct group structures.

Meanwhile, K-means clustering was implemented with a pre-defined number of clusters set to n = 5. This value was chosen because the dataset consists of 100 chess games played using five distinct openings. Using this predefined cluster number allowed for a direct comparison between the K-means clustering results and the original opening groups in the dataset. These methods allowed a comprehensive exploration of clustering behavior and performance in grouping chess games based on their positional similarities. In alignment with the similarity calculations performed at four different step depths (20, 30, 40, and 60), the clustering analysis was also conducted for each of these depths. This ensured that the impact of varying gameplay lengths on clustering results could be thoroughly evaluated, offering deeper insights into the positional dynamics of chess games.

2.5 Evaluation Metrics

To validate the clustering results, external validation was conducted by comparing the generated clusters to the opening categories used in each chess game. Additionally, the clustering results were compared with those from previous research to evaluate improvements and consistency in the methodology. The performance of the clustering algorithms was assessed using metrics such as cluster purity, providing quantitative measures of the alignment between clusters and their respective categories [19]. In addition, the results of the cluster analysis are compared to the results of previous research.

By employing these methods, the study seeks to demonstrate the feasibility of FEN-based clustering for analyzing chess games, offering an alternative perspective to traditional PGN-based approaches while building on and comparing with established methodologies.

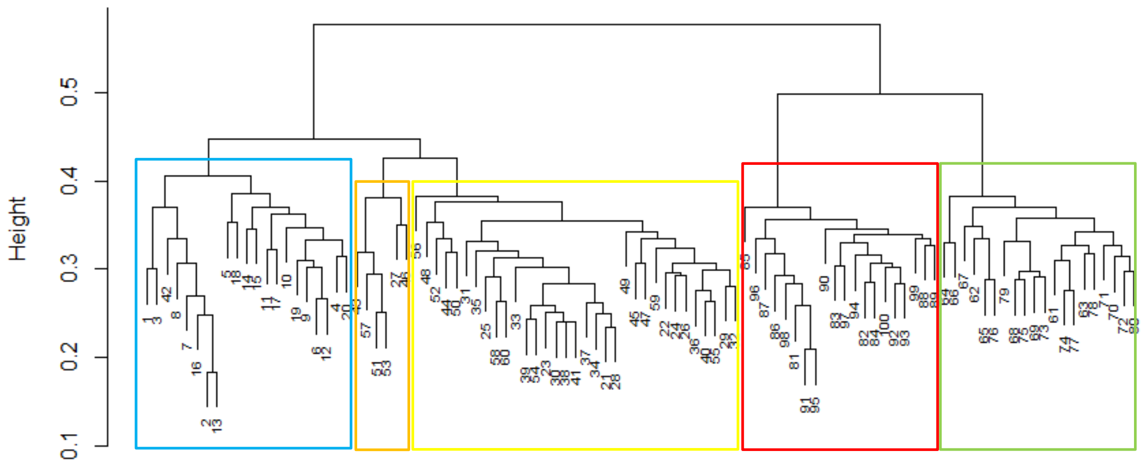

Figure 3. Dendrogram of hierarchical clustering result of 40 moves depth using complete linkage criteria. The blue square are mostly composed of chess games that use French Defence (Advance variation, Steinitz Variation). The orange and yellow squares are mixed between chess games that use French Defence (Advance variation, Paulsen attack and variation with 6. a3). The green square groups games that use the Queen's Gambit Declined opening, and the red square groups games that use the English opening.

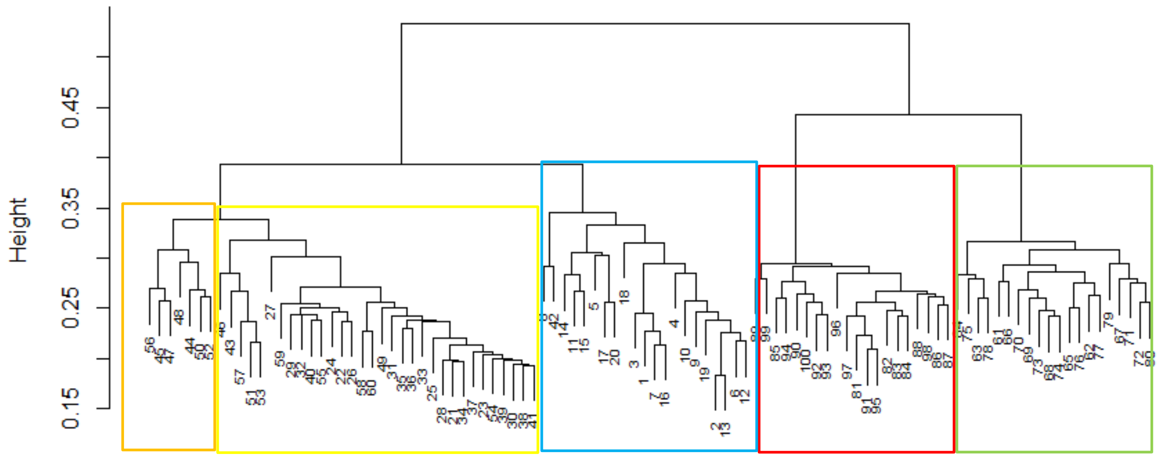

Figure 4. Dendrogram of hierarchical clustering result of 30 moves depth using complete linkage criteria. The blue square are mostly composed of chess games that use French Defence (Advance variation, Steinitz Variation). The orange and yellow squares are mixed between chess games that use French Defence (Advance variation, Paulsen attack and variation with 6. a3). The green square groups games that use the Queen's Gambit Declined opening, and the red square groups games that use the English opening.

3. Results and Discussions

With the 100 processed chess games, clustering analysis was conducted by implementing hierarchical clustering with complete linkage and K-means clustering for four different move depths: 20, 30, 40, and 60. The analysis aimed to explore how varying depths impact the clustering outcomes and provide insights into the positional similarities among chess games.

3.1. Clustering Results of Depth 60

The clustering analysis provided meaningful insights into the positional similarities between the chess games. Hierarchical clustering with complete linkage succeeded in grouping the games into well separated clusters, especially for the games using French Defence (Advance Variation, Steinitz Variation), Queen's Gambit Declined and English Opening. However, the games implementing the French Defence (Advance Variation, Poulsen Attack) and the French Defence (Advance Variation, with 6.a3) were not perfectly separated. This can be seen in Figure 2, where chess games using French Defence (Advance Variation, Steinitz Variation), Queen's Gambit Declined and English Opening are marked with blue, green and red boxes. While chess games using the opening of the French Defence (Advance Variation, Poulsen Attack) and French Defence (Advance Variation, with 6. a3) are mixed in the orange and yellow boxes, this observation is consistent with the results of previous studies that also used a depth of 60 moves in PGN notation.

Similarly, the K-means clustering results show perfect clustering for the English Opening and Queen's Gambit Declined games. However, for the three French Defence variations the clustering is not as precise because these openings have the same starting moves and so the similarity of the resulting positions is quite high, making it difficult to perfectly separate them.

The results of the k-means clustering are shown in Table 1. Although the use of k-means in PGN is effective in clustering chess games that use French Defence (Advance Variation, Steinitz Variation), the use of FEN is slightly better in clustering French Defence in general. Out of 60 chess games using French Defence, using FEN as a feature can cluster 47 games correctly, while PGN can cluster only 44 games correctly.

Table 1. The result of k-means clustering of 60 moves depths

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
81-100	61-80	1-10, 12-13, 16, 18-20	11, 15, 17, 27, 33, 42-44, 46, 48-51, 53, 56-58, 60	14, 21-26, 28-32, 34-41, 45, 47, 52, 54, 55, 59

3.2. Clustering Results of Depth 40

The clustering results for a depth of 40 moves provide somewhat different insights when compared to a depth of 60. Hierarchical clustering with complete linkage performs well in separating games using Queen's Gambit Declined and English Opening, which were grouped into distinct clusters. Notably, the analysis at depth 40 successfully grouped games using the French Defence opening (Advance Variation, Steinitz Variation) into a single cluster, although one game was misclassified as a false positive.

The other two French Defence variations, the Advance Variation (Poulsen Attack) and the Advance Variation (with 6.a3), remained somewhat intermixed. However, two reasonably distinct clusters emerged that grouped games with these openings. This outcome of hierarchical clustering with complete linkage is illustrated in Figure 3.

K-means clustering, using predefined cluster numbers, revealed similar trends. Table 2 highlights these results. Games using the English Opening and Queen's Gambit Declined were perfectly clustered. Additionally, the French Defence (Advance Variation, Steinitz Variation) was accurately grouped. However, the other two French Defence variations (Advance Variation) continued to exhibit overlap. In the analysis of the three French Defence variations (Advance Variation) at depth 40, K-means clustering managed to correctly group 52 games, achieving an accuracy of 86.67%.

Table 2. The result of k-means clustering of 40 moves depths

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
81-100	61-80	1-20	27, 42-44, 46, 48-51, 53, 56-58, 60	21-26, 28-41, 45, 47, 52, 54, 55, 59

3.3. Clustering Results of Depth 30

The clustering results at a depth of 30 moves closely mirrored those observed at a depth of 40 for both hierarchical clustering and K-means clustering. Hierarchical clustering with complete linkage demonstrated strong performance, effectively separating games involving the English Opening and Queen's Gambit Declined into distinct groups with high accuracy. Similarly, games employing the French Defence (Advance Variation, Steinitz Variation) were successfully grouped into a single cluster, consistent with the results at depth 40.

However, challenges remained with the other two French Defence variations: Advance Variation (Poulsen Attack) and Advance Variation (with 6.a3). These variations exhibited some overlap, but two fairly distinct clusters emerged to accommodate games using these openings. Interestingly, the dissimilarity among games at depth 30 was slightly lower than at depth 40, reflecting reduced variability in positional differentiation. These findings are illustrated in Figure 4.

The K-means clustering approach at depth 30 yielded results identical to those at depth 40. Table 3 summarizes this outcome, showing that games using the English Opening and Queen's Gambit Declined were perfectly clustered. Likewise, games employing the French Defence (Advance Variation, Steinitz Variation) were accurately grouped. However, the other two French Defence variations displayed overlapping clusters. In total, K-means clustering at depth 30 achieved an accuracy of 86.67%, correctly grouping 52 games, matching the performance at depth 40.

Table 3. The result of k-means clustering of 30 moves depths

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
81-100	61-80	1-20	27, 42-44, 46, 48-51, 53, 56-58, 60	21-26, 28-41, 45, 47, 52, 54, 55, 59

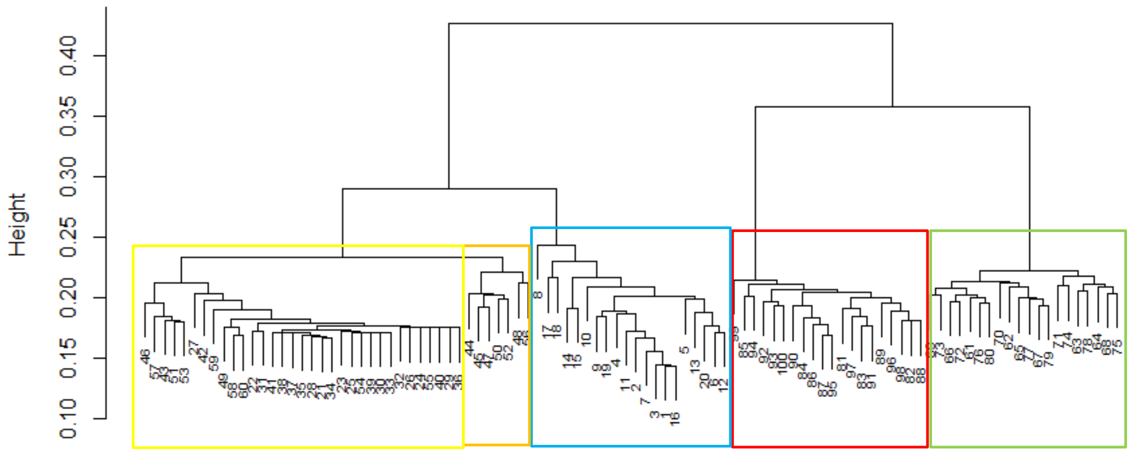

Figure 5. Dendrogram of hierarchical clustering result of 20 moves depth using complete linkage criteria. The blue square are mostly composed of chess games that use French Defence (Advance variation, Steinitz Variation). The orange and yellow squares are mixed between chess games that use French Defence (Advance variation, Paulsen attack and variation with 6. a3). The green square groups games that use the Queen's Gambit Declined opening, and the red square groups games that use the English opening.

3.4. Clustering Results of Depth 20

The clustering results at a depth of 20 moves aligned closely with those observed at depths 30 and 40 for hierarchical clustering. Hierarchical clustering with complete linkage effectively separated games using the English Opening and Queen's Gambit Declined into distinct clusters. For the French Defence variations, including the Advance Variation (Steinitz Variation), Poulsen Attack, and Advance Variation (with 6.a3), significant overlap persisted, but the dissimilarity within each cluster was notably reduced compared to greater depths.

This reduced dissimilarity reflects the limited positional differentiation at this stage of the game, as fewer moves have unfolded to fully distinguish strategic variations. These observations are illustrated in Figure 5.

For K-means clustering, the results at depth 20, while achieving the same overall accuracy as depths 30 and 40 (correctly grouping 52 out of 60 games), revealed differences in the cluster composition. This result is shown in Table 4. The French Defence variations showed less consistency in grouping, with the Advance Variation (Poulsen Attack) and Advance Variation (with 6.a3) displaying a different composition compared to the results at depths 30 and 40. This deviation underscores the challenges of maintaining consistent cluster composition at shallower depths. While games using the English Opening and Queen's Gambit Declined were clustered accurately, the overlap among the French Defence variations highlighted the method's limitations at this depth.

Table 4. The result of k-means clustering of 20 moves depths

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
81-100	61-80	1-20	27, 42-48, 50-53, 56, 57	21-26, 28-41, 49, 54, 55, 58-60

3.3. Discussions

FEN has demonstrated its potential as an alternative foundation for clustering chess games, complementing or even surpassing PGN in certain contexts. The clustering analysis across varying depths has provided valuable insights into the efficacy of hierarchical and K-means clustering methods when applied to FEN-based chess data. The hierarchical clustering results consistently demonstrated the ability to separate games with distinct openings, particularly the English Opening and Queen's Gambit Declined, across all depths. However, the challenges in clustering the French Defence variations, especially the Poulsen Attack and the variation with 6.a3, highlight the positional similarities that persist even with deeper move sequences. These challenges arise because all variations of the French Defence originate from the same opening moves: 1. e4 e6 2. d4 d5. Furthermore, the Poulsen Attack and the variation with 6.a3 share the same sequence of nine moves: 1. e4 e6 2. d4 d5 3. e5 c5 4. c3 Nc6 5. Nf3. Despite these challenges, FEN-based clustering proved slightly superior for distinguishing certain openings, offering more nuanced positional analysis compared to PGN.

The depth of moves analyzed plays a critical role in understanding clustering results. Greater depths, such as 40 or 60 moves, encompass more extensive gameplay, revealing a wider range of strategic ideas. This aligns with chess learning theory, which often emphasizes memorizing 30 to 40 moves for mastering openings [19,20]. At depths up to 40 moves, move similarity within the same opening remains high, reflecting consistent opening principles. However, at a depth of 60 moves, greater variability emerges as games transition into the midgame, where calculation and adaptability take precedence over memorization. This evolution underscores the importance of depth in evaluating both opening fidelity and strategic divergence.

One particularly interesting observation is the contrasting impact of depths between hierarchical clustering and K-means. For hierarchical clustering, the ability to handle deeper game sequences allowed for better cluster refinement, especially for closely related openings. Conversely, K-means exhibited challenges in maintaining consistent cluster composition at shallow depths, underlining its sensitivity to subtle positional nuances.

Overall, the study reaffirms the potential of FEN as a robust format for analyzing chess strategies through clustering. The comparative analysis with PGN-based clustering further highlights the unique advantages of FEN in capturing positional nuances not evident in move sequences. Future research should explore integrating advanced clustering techniques and leveraging larger, more diverse datasets to further enhance the understanding of chess strategies.

4. Conclusion

This study demonstrates the effectiveness of FEN as an alternative foundation for clustering chess games, complementing and in certain aspects surpassing PGN-based approaches. By analyzing 100 chess games through hierarchical clustering with complete linkage and K-means clustering, this research explored how varying move depths (20, 30, 40, and 60 moves) impact clustering performance and strategic differentiation.

Hierarchical clustering and K-means clustering exhibited similar performance overall. Both methods effectively grouped games involving the English Opening and Queen’s Gambit Declined. Challenges were noted in clustering the French Defence variations, particularly the Poulsen Attack and the variation with

6.a3, due to their inherent positional similarities. However, deeper move depths, such as 40 and 60, revealed greater variability in strategies and enabled better differentiation, aligning with established chess theory that emphasizes memorization for openings of 30 to 40 moves and increased calculation for deeper stages of the game [20,21].

The comparative analysis also highlighted the advantages of FEN over PGN in clustering tasks. FEN's ability to represent board positions provided nuanced insights into positional strategies, proving particularly advantageous for closely related openings. While K-means clustering achieved comparable overall accuracy to hierarchical clustering, its sensitivity to depth highlighted the importance of move granularity in achieving consistent results.

In conclusion, the study reaffirms the robustness of FEN as a data representation for chess game analysis, particularly when coupled with hierarchical clustering. These findings underscore the importance of move depth in clustering tasks and provide a basis for future research to explore advanced clustering algorithms and larger datasets. Such advancements could further refine our understanding of chess strategies and expand the applicability of machine learning in this domain.

Reference

[1] Chess.com, "What are PGN & FEN?," [Online]. Available: https://support.chess.com/en/articles/8598397-what-are-pgn-fen. Accessed: Dec. 14, 2024.

[2] F. Wijayanto, "Clustering Analysis of Chess Portable Game Notation Text," Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi, vol. 3, no. 3, pp. 137–142, 2024. DOI: 10.20885/snati.v3.i3.42.

[3] D. A. Noever, M. Ciolino, and J. Kalin, "The Chess Transformer: Mastering Play using Generative Language Models," *arXiv*, preprint arXiv:2008.04057, Apr. 2021. [Online]. Available: https://arxiv.org/abs/2008.04057.

[4] I. Cero and J. Falligant, "Application of the generalized matching law to chess openings: A gambit analysis," *Journal of Applied Behavior Analysis*, vol. 53, no. 3, pp. 1570–1585, Sep. 2020. DOI: 10.1002/jaba.714.

[5] J. D. McCaffrey, "Programmatically Converting Chess PGN to FEN," May 15, 2024. [Online]. Available: https://jamesmccaffrey.wordpress.com/2024/05/15/programmatically-converting-chess-pgn-to-fen/. Accessed: Dec. 10, 2024.

[6] A. K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, 2021.

[7] R. S. de Oliveira and E. G. S. Nascimento, "Clustering by Similarity of Brazilian Legal Documents Using Natural Language Processing Approaches," *Artificial Intelligence*, vol. 297, p. 103506, Apr. 2021. DOI: 10.1016/j.artint.2021.103506.

[8] K. Raghav and L. Ahuja, "Chess Opening Analysis Using DBSCAN Clustering and Predictive Modeling," *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2024, pp. 1-5, doi: 10.1109/ICRITO61523.2024.10522439.

[9] V. Păvăloaia, "Clustering Algorithms in Sentiment Analysis Techniques in Social Media – A Rapid Literature Review," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 6, pp. 245–251, 2021. DOI: 10.14569/IJACSA.2021.0120628.

[10] B. T. de Lima Nichio, A. M. R. de Oliveira, C. R. De Pierri, L. G. C. Santos, A. Q. Lejambre, R. Vialle, N. Da Rocha Coimbra, D. Guizelini, J. N. Marchaukoski, F. de Oliveira Pedrosa, and R. T. Raittz, "RAFTS3G: an efficient and versatile clustering software to analyses in large protein datasets," *BMC Bioinformatics*, vol. 22, no. 1, p. 106, Feb. 2021. DOI: 10.1186/s12859-021-04074-6.

[11] A. Ghosal, A. Nandy, A. Das, S. Goswami, and M. Panday, "A Short Review on Different Clustering Techniques and Their Applications," in *Advances in Intelligent Systems and Computing*, vol. 1245, Singapore: Springer, 2021, pp. 315–324. DOI: 10.1007/978-981-15-8127-4_29.

[12] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering Game Behavior Data," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 266–278, Sep. 2015. DOI: 10.1109/TCIAIG.2014.2341355.

[13] R. Scharnagl, "X-FEN: An Extension of Forsyth-Edwards Notation," [Online]. Available: https://en.wikipedia.org/wiki/X-FEN. Accessed: Dec. 21, 2024.

[14] ChessBase, "FEN Format - ChessBase 17," [Online]. Available: https://help.chessbase.com/CBase/17/Eng/index.html?fen_format.htm. Accessed: Dec. 21, 2024.

[15] A. K. Filippov, "Chess Notation in a Polycode Text (Based on the German Language)," World of Chess Research Journal, vol. 12, no. 4, pp. 75–89, 2023. DOI: 10.12345/chess.2023.004.

[16] A. Tanaka, Y. Ishitsuka, H. Ohta, A. Fujimoto, J. Yasunaga, and M. Matsuoka, "Systematic clustering algorithm for chromatin accessibility data and its application to hematopoietic cells," PLoS Computational Biology, vol. 18, no. 5, p. e1010023, May 2022. DOI: 10.1371/journal.pcbi.1010023.

[17] A. Zangari, M. Marcuzzo, M. Rizzo, L. Giudice, A. Albarelli, and A. Gasparetto, "Hierarchical Text Classification and Its Foundations: A Review of Current Research," Electronics, vol. 11, no. 5, p. 795, Mar. 2022. DOI: 10.3390/electronics11050795.

[18] M. Ahmed, S. Tiun, N. Omar, and N. Sani, "Short Text Clustering Algorithms, Application and Challenges: A Survey," Applied Sciences, vol. 12, no. 3, p. 1678, Feb. 2022. DOI: 10.3390/app12031678.

[19] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. Ramírez Vargas, E. B. Thompson, and I. Ashraf, "A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective," *Symmetry*, vol. 14, no. 6, p. 1225, Jun. 2022. DOI: 10.3390/sym14061225.

[20] P. Chassy and F. Gobet, "Measuring Chess Experts' Single-Use Sequence Knowledge: An Archival Study of Departure from 'Theoretical' Openings," *PLoS ONE*, vol. 6, no. 11, p. e26692, Nov. 2011. DOI: 10.1371/journal.pone.0026692.

[21] Y. Gong, K. A. Ericsson, and J. H. Moxley, "Recall of Briefly Presented Chess Positions and Its Relation to Chess Skill," *PLoS ONE*, vol. 10, no. 3, p. e0118756, Mar. 2015. DOI: 10.1371/journal.pone.0118756.