Advancing Fertility Analysis: A Novel Clustering-Based Approach for Classifying Sperm Motility Patterns in Reproductive Medicine

Julianna Lamm, Department of Computer Science, Fordham University, New York, United States, jlamm@fordham.edu

***Abstract*— This study presents a novel approach to analyzing sperm motility using advanced clustering techniques, namely hierarchical clustering and K-means clustering, applied to a recent, comprehensive dataset of bounding box coordinates from sperm video recordings. The primary objective was to classify sperm movement patterns to enhance the understanding of fertility-related characteristics. We first preprocess the video dataset to extract quantifiable features of sperm kinematic metrics. K-means clustering is then applied to discern natural groupings within the data, providing an initial insight into distinct movement patterns. Subsequently, hierarchical clustering was used to identify and cluster participants with similar sperm populations. employed to refine these classifications into clearly defined categories. The results demonstrate a significant correlation between the identified movement patterns and key velocity metrics and fertility indicators, offering potential applications in clinical diagnostics and fertility treatments. This research contributes to the broader field of biological data analysis by showcasing the effective application of machine learning techniques in interpreting complex, real-world biological phenomena. The findings also open avenues for further research in automated classification systems in reproductive medicine. The study's implications extend beyond academic interest, offering tangible benefits in clinical settings by enhancing the accuracy of fertility assessments and treatment plans.**

Keywords—sperm, motility, andrology, kinematics, computer vision, fertility, reproduction

I. Introduction

Approximately 17.5% of individuals of reproductive age are affected by infertility in their lifetime - and this number is projected to rise [1]. Each year, over 2.5 million cycles of assisted reproductive technologies (ART) are performed globally, but the success rate continues to remain at ~33%. There are two main ART procedures currently employed: In vitro fertilization (IVF) and Intracytoplasmic sperm injection (ICSI). During IVF, a selected subpopulation of ~25,000 to ~150,000 sperm is placed near an oocyte to initiate fertilization. By contrast, in ICSI, an individual sperm is carefully selected and subsequently injected directly into the oocyte [2]. The fertility industry has experienced a notable shift towards ICSI over IVF. This trend is attributed to ICSI's effectiveness in addressing issues related to sperm motility and quality, which are critical barriers in traditional sperm-egg fusion processes [3]. However, the safety of ICSI is increasingly scrutinized, as the technique, by circumventing motility and sperm-quality barriers, permits the fertilization of the egg by potentially low-quality sperm. In fact, the expanded use of ICSI is associated with an increased risk of congenital abnormalities and autism when compared to conventional IVF [2].

Identifying the highest quality sperm for ICSI is a critical challenge. Clinicians currently follow World Health Organization (WHO) guidelines, which recommend sperm selection based on morphology, motility, and DNA integrity. However, notable discrepancies between these methods have been reported, requiring manual analysis to support Computer Assisted Sperm Analysis (CASA) findings [4]. Motility parameters and classifications generated by CASA are developed using proprietary and poorly understood “black-box” algorithms, and systems are expensive and highly variable. While single-cell level sperm assessment is crucial, CASA currently only provides data at the sample level. Even further, the parameters provided by CASA lack descriptiveness and show limited predictive accuracy for fertility outcomes. Due to time constraints, individually examining thousands of sperm is impractical, and the selection of the "best" sperm often varies by clinician, making it highly subjective [5]. Consequently, there is an urgent need for more descriptive metrics to assess an individual's fertility status accurately.

Advancements in the field of machine learning have instigated a vast array of uses in areas like computer vision, pharmaceutical research, language interpretation, among numerous other applications. In the fertility industry, machine learning technologies have shown efficacy in identifying embryos with the highest likelihood of successful blastocyst formation. Traditional methods of blastocyst evaluation by embryologists, which rely on static images, tend to be subjective and require significant time [5]. The use of machine learning in sperm selection therefore shows significant potential to improve ART by classifying sperm motility tracks and correlations between sperm movement and health.

Researchers from the Simula Metropolitan Center for Digital Engineering in Oslo, Norway have expanded upon a

previous multi-modal dataset consisting of 20 annotated video clips featuring twenty 20 participants, each with associated metadata, including age, BMI, fatty-acid composition, and CASA motility data [6]. These annotations were used to train a Deep Learning (DL) model for predicting bounding boxes and identifying sperm in unlabeled videos.

The dataset's publication specifically invites its use in generating kinematic features and advancing sperm tracking algorithms. This targeted call is a central aspect of this current project, emphasizing the dataset's potential of validating the utility of DL models to enable a wider range of machine learning applications in sperm tracking.

Recognizing the growing significance of comprehending sperm motility patterns in Assisted Reproductive Technology (ART), this project aims to extract tracks and motility features from the bounding box data in the dataset. Utilizing K-Means and Hierarchical clustering algorithms, along with Random Forest, this study visualizes, categorizes, and deciphers the kinematic parameters associated with distinct human sperm subpopulations. This approach is intended to provide a deeper understanding of sperm behavior and its implications in ART.

II. METHODS

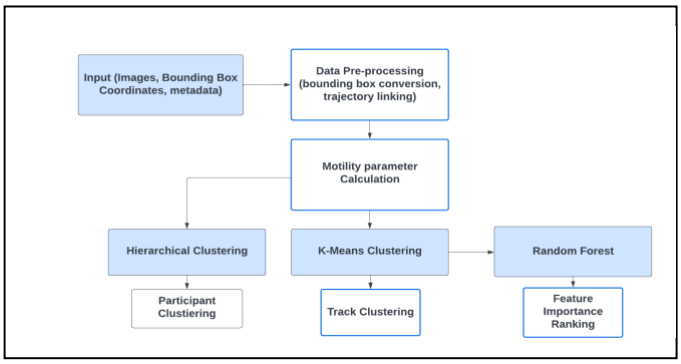

Fig. 1: Summary of workflow

A. VISEM Dataset

The VISEM-tracking dataset was obtained and downloaded from the following address: (https://zenodo.org/record/7293726) [6].The dataset contains 20 videos from 20 different participants, each with a fixed duration of 30 seconds with corresponding bounding boxes. In total, the researchers provided 656,334 bounding box coordinates corresponding to 1,121 unique sperms with tracking IDs. An additional 336 frames from unlabeled videos were provided for the purposes of exploring more advanced methods such as unsupervised learning, although they were not utilized in the current study's scope.

The dataset is organized into individual folders for each participant, containing extracted video frames, bounding box labels for each frame, and text files detailing bounding box coordinates with unique tracking identifiers for each spermatozoon. Additionally, the dataset includes a .csv file with general participant information, encompassing standard semen analysis, serum levels of sex hormones, measured from blood samples, serum levels of the fatty acids in phospholipids, and fatty acid levels of spermatozoa.

B. Data Pre-Processing

The text files containing bounding box coordinates and tracking identifiers ('fid') for each spermatozoon were imported into a Jupyter notebook and merged to form a comprehensive DataFrame. This DataFrame included data from all participants with their corresponding sperm tracking IDs. Initial exploratory and descriptive analyses were conducted. Bounding box coordinates were converted from YOLOv5 format to measurements in micrometers (µm) [7]. A summary of workflow is found in Fig. 1.

C. Track and Feature Generation

Seven motility features were crafted in accordance with the definitions from OpenCASA, a sperm tracking software. *[8]*. These features are:

1. Curvilinear velocity (VCL)
2. Linear velocity (VSL)
3. Average path velocity (VAP)
4. Linearity coefficient (LIN)
5. Wobble coefficient (WOB)
6. Straightness coefficient (STR)
7. Mean and maximum amplitude of lateral head displacement (ALH)
8. Fractal Dimension (FD)

These parameters were selected to mirror those provided CASA systems and are commonly utilized in clinical settings to assess fertility. Detailed descriptions and the formulas for these parameters are presented in Table 1. Fig. 2 provides a visual illustration of some of these calculated parameters.

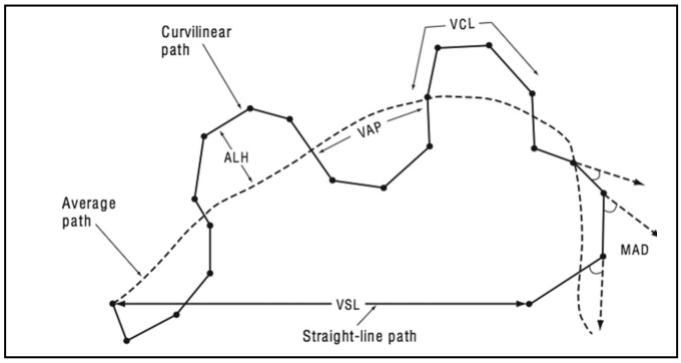

Fig. 2: Visualization of sperm track trajectory parameters

Parameter	Motility Parameter Definitions	
	Mathematical Definition	***Meaning***
Straight-line velocity (VSL)	$D(p1,pN)*. FrameRateN-1*\mu$, where $D(p1,pN)$ is distance between first and last point of trajectory, *FrameRateN-1* is the Frame rate in frames per second(fps), N is number of frames, μ is scaling factor	VSL is determined by finding the straight-line distance between the first and last points of the trajectory and correcting for time. This value then gives the net space gain within the observation period.
Curvilinear velocity (VCL)	$\sum t=1N-1D(pt,pt+1))*FrameRateN-1*\mu$, where D($pt$,$pt$+1)) is the distance between two consecutive points on trajectory. *FrameRateN-1* is the Frame rate in frames per second(fps), N is number of frames, μ is scaling factor	VCL is the distance traveled by the spermatozoon along its curvilinear path/s and is calculated by finding the sum of the distances along the trajectory then correcting for time. It refers to the total distance that the sperm head covers in the observation period.
Average-path velocity (VAP)	$(\sum t=1N-wD(qt,qt+1))*FrameRateN-w*\mu$ Where $D(qt,qt+1)$) is distance between two points on the average path (moving average), w is specified window for moving average, and μ is unit scaling factor	VAP is the distance the spermatozoon has traveled in the average direction of movement in the observation period. It is calculated by finding the length of the average path and correcting for time.
Linearity (LIN)	$VSL/VCL*100$	LIN is a comparison of the straight-line and curvilinear paths. It is an expression of the relationship between the two-dimensional projection of the three-dimensional path taken by the spermatozoon (i.e., curvilinear path) and its net space gain.
Wobble (WOB)	$VAP/VCL*100$	WOB is the expression of the relationship between the average and curvilinear paths.
Straightness (STR)	$VSL/VAP*100$	STR is a comparison of the straight-line and average paths and gives an indication of the relationship between the net space gain and the general trajectory of the spermatozoon.
Amplitude of Lateral Head Displacement (ALHmean)	ALHmean = Scale Factor × ∑(Distances between St and S) / Count ALHmax = Scale Factor × max(Distances between St and S) Where: 'St' is a segment of the actual trajectory. 'S' is a segment of the smoothed trajectory. 'Scale Factor' ('μ') is the microns per pixel conversion factor. 'Count' is the number of segments analyzed for each trajectory.	The amplitude of lateral head displacement (ALH) is used as an approximation of the flagellar beat envelope. ALHmean is the mean of all of the ALH values along the trajectory.
ALHmax	Using the above definition.	ALHmax is the maximum ALH found along the trajectory.
Fractal Dimension	FD=log(n)[log(n)+log(dL]	The fractal dimension is an expression of the degree to which a line fills a plane. It may be considered that the fractal dimension of a curve indicates its regularity. A curve with a low fractal dimension would be regular and predictable. Similarly, a curve with a high fractal dimension would have irregularly spaced changes in direction, apparently at random.
Hyperactivation	In CASA analysis, most frequently: VCL ≥ 150 μm/s, LIN ≤ 50%, and ALH ≥ 7	Defined as covering less forward distance relative to their vigor, displaying large deviations of the head from the path of movement and angles >90º between adjacent point. [goodson]

TABLE I. Table of definitions for kinematic parameters describing sperm motility.

D. Clustering

The movement and trajectories of each sperm were categorized by clustering the tracks using Principal Component Analysis (PCA) enhanced K-Means and Hierarchical Clustering Algorithms implemented through the Scikit-Learn library [9]. To determine the optimal number of clusters for K-Means, *n_clusters*, both the elbow and silhouette methods were employed. Elbow and silhouette methods [9]. A silhouette value measures the quality of clusters in a dataset by measuring the similarity of an object to its own cluster compared to other clusters. The silhouette score ranges from -1 to +1, where a higher score signifies that the object is well integrated into its own cluster and distinct from other clusters. The calculation of the silhouette value involves measuring how close each point in one cluster is to points in the neighboring clusters. This metric is particularly useful for validating the consistency within clusters of data. The silhouette value is calculated as follows:

$$s(i) = \frac{b_i - a_i}{\max\{a_i b_i\}} \qquad (1)$$

where a_i stands for average distance from data point i to other data points within the same cluster, and b_i is the minimum distance from data point *i* to other data points in other clusters. If s_i is close to +1, it indicates that the data point is well clustered. If s_i is close to 0 it suggests that the data point is on or very close to the decision boundary between two neighboring clusters. If s_i is close to -1, it indicates the data point has been assigned to the incorrect cluster. Similarly, the elbow method determines the optimal number of k clusters by computing clustering for different k values and evaluating the sum of squared differences (SSD) between each point and the centroid of its cluster. "elbow point" is identified at the juncture where the rate of decrease in the Sum of Squared Distances (SSD) markedly shifts. This point is considered to indicate the optimal number of k clusters. A summary of cluster visualizations for different k clusters, along with corresponding silhouette and elbow scores is included in Fig. 4.

The *K-means* ++ algorithm, recognized by Scikit-Learn as an enhanced method for centroid initialization was employed. The number of initializations, *n_init,* is the number of times the k-means algorithm is run with different seeds. The study adhered to the default setting of running the k-means algorithm ten times with different centroid seeds, denoted by *n_init*=10. The maximum number of iterations per run, *max_iter*, was capped at 100, and the random seed, *random_state*, set to 42 by convention.

Given the high dimensionality of the feature set utilized for clustering, PCA was employed to scale and transform the data [10]. By setting *n_components* = 2, PCA reduced the dataset from its original high-dimensional feature space to a two-dimensional plane defined by the first two principal components, *PC1* and *PC2*. These components capture the most significant variance within the data, facilitating an improved clustering process. The bi-dimensional PCA-reduced data was used for both K-Means and hierarchical clustering visualization, enabling the observation of potential clusters in a simplified two-dimensional context.

A hierarchical clustering dendrogram was produced to group motility metrics at the population level for the participants. This dendrogram enables the grouping of participants with similar motility patterns. The optimal number of clusters was determined by visually examining the dendrogram's y-axis to identify the most suitable distance cut-off.

E. Random Forest

Cluster labels generated from K-Means clusters were used as class labels to rank each of the Kinematic Features by their importance in the clustering algorithm [9]. The standard parameters listed by Sci-Kit were used at *n_estimators* = 100 and *random_state* = 42. The top five (5) ranked outputs were then used in the hierarchical clustering algorithm

III. RESULTS

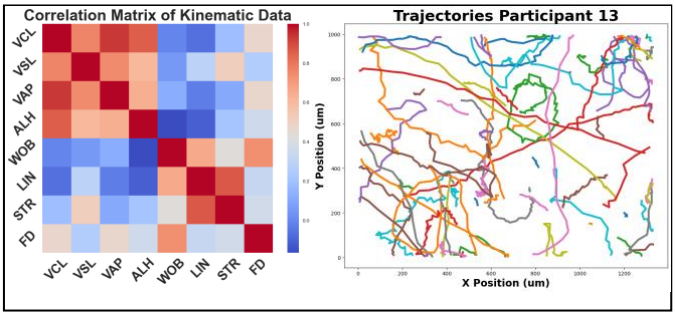

Fig. 2: (left) Correlation matrix for kinematic data. (right) Example of trajectory tracks from Participant 13.

A. Data Exploration and Feature Selection exploration and Feature Selection

The dataset was initially analyzed by counting the individual sperm 'fids' per participant and plotting the x and y coordinates for each sperm across the frames. Fig. 2 (right) presents the sperm movement trajectories for a single participant. Subsequently, kinematic features derived from these trajectories were represented in a correlation matrix, as shown in Fig. 2 (left), to understand the interrelations among kinematic metrics and to confirm the non-redundancy of features selected for clustering. High correlation among the velocity metrics (VCL, VSL, and VAP) was anticipated as they all describe movement over time. This correlation necessitates models capable of handling inter-feature dependencies, hence the future selection of PCA for dimensionality reduction. Motility metrics derived from velocity measurements also exhibited expected correlations, particularly LIN and STR, which relate straight-line paths to VSL and VAP and the directness of the trajectory, respectively.

B. Track Pattern Identification

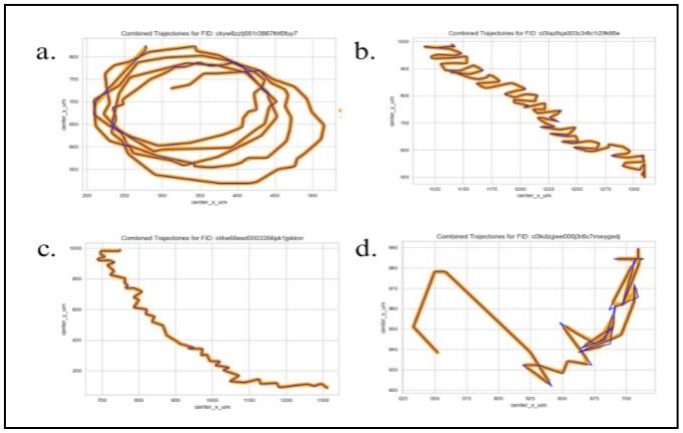

Fig. 3: Examples of three distinct track patterns of interest. a. Hyperactivated tracks circular track b. Hyperactivated track c. Linear Mean/Progressive track d. Weakly motile track

Before clustering, track patterns were identified using domain expertise and findings from related research. Three main track parameters characterizing sperm movement have been recognized in the literature: circular/hyperactivated, linear progressive, and weakly motile. [8, 10, 11]. These trajectories were also visible in the analyzed dataset, as illustrated in Figure 3, which plots the x and y coordinates of a single particle across frames. The tracks shown in Figure 3a and 3b correspond to the "star-spin hyperactivated" and "hyperactivated (transitional)" patterns described by CASAnova, a Multi-Class Support Vector Machine Model from researchers at the University of North Carolina at Chapel Hill. The patterns observed from the generated trajectories are consistent with those identified in recent studies [11, 10].

C. K-Means PCA Clustering

Commercial CASA systems typically report only the percentage of motile and immotile sperm without providing detailed statistics on swimming patterns, despite the significance of differentiating hyperactive, progressive, and immotile tracks for assessing fertility status. Through KMeans clustering, leveraging the kinematic features previously discussed, each of the three motility patterns was segregated into distinct clusters. The optimal cluster count was gauged using silhouette and elbow scores, with Fig. 4 showcasing the results for various numbers of n_clusters. Due to the ambiguity of the elbow point, the silhouette method was employed for a clearer determination of the ideal cluster number. A combination of elbow and silhouette scores, along with visual cluster examination, indicated that the most effective clustering occurred with n_clusters set to 5.

For K-Means clustering, 1,103 individual sperm swim tracks were observed. Motility parameters used for this clustering were VCL, VSL, VAP, WOB, LIN, STR, ALH Mean, ALH Max, and FD. The minimum, first quartile (Q1), median, third quartile (Q3), and maximum values calculated for each of the metrics in each cluster were summarized in a box-and-whisker plots shown in *Fig. 5.* Visually, each of the clusters segments the three identified tracks reasonably well.

- *Cluster 0 (Hyperactived):* 107 tracks with high VCL and LIN values and low VSL values. High VAP and FD. Visual inspection of tracks in *Figure 5* indicate possible transition state that characterizes sperm in hyperactivated state.
- *Cluster 1 (weakly motile):* 371 tracks with very low velocity values and short tracks indicating very little movement.
- *Cluster 2 (immotile/dead cells):* 152 tracks -all metrics are 0 for this cluster and therefore not shown.
- *Cluster 3 ("circular" hyperactivated):* 237 tracks with less vigorous VCL than in cluster 0 but larger distribution of values. Visual inspection of tracks does not show much difference than in cluster 0, highlighting the importance of not relying solely on visual identification of tracks.
- *Cluster 4 (Linear mean/progressive tracks):* 236 tracks with moderate VCL and VSL values accompanying box-and-whiskers plot also confirms the forward moving trajectories characterizing Cluster 4, with a high degree of linearity.

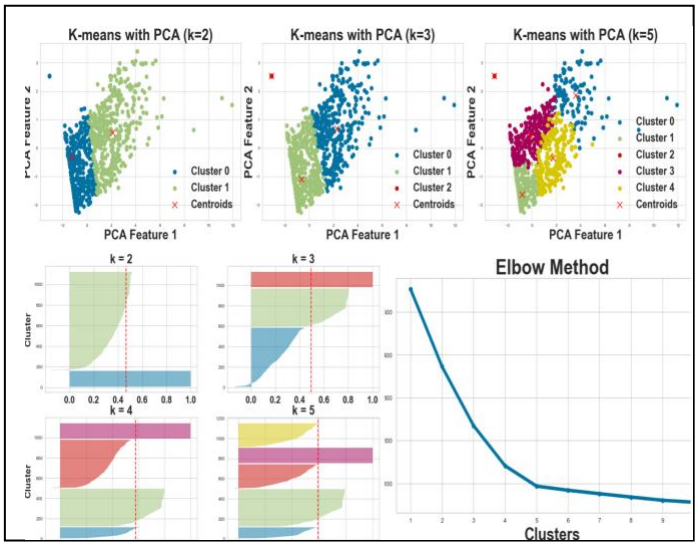

Fig. 4: (Top) K-Means Clusters for k=2, k=4, k=5. (Bottom) Elbow method and silhouette scores for determining optimal k.

D. Random Forest

In anticipation of using the dataset to set thresholds for classifying and predicting sperm tracks, a Random Forest Classifier was employed to determine the importance of features contributing to cluster formation. Table 2 displays the most significant features for differentiating the clusters. Consistently, VCL (Curvilinear Velocity), VSL (Straight-Line Velocity), and LIN (Linearity) emerged as the top-three

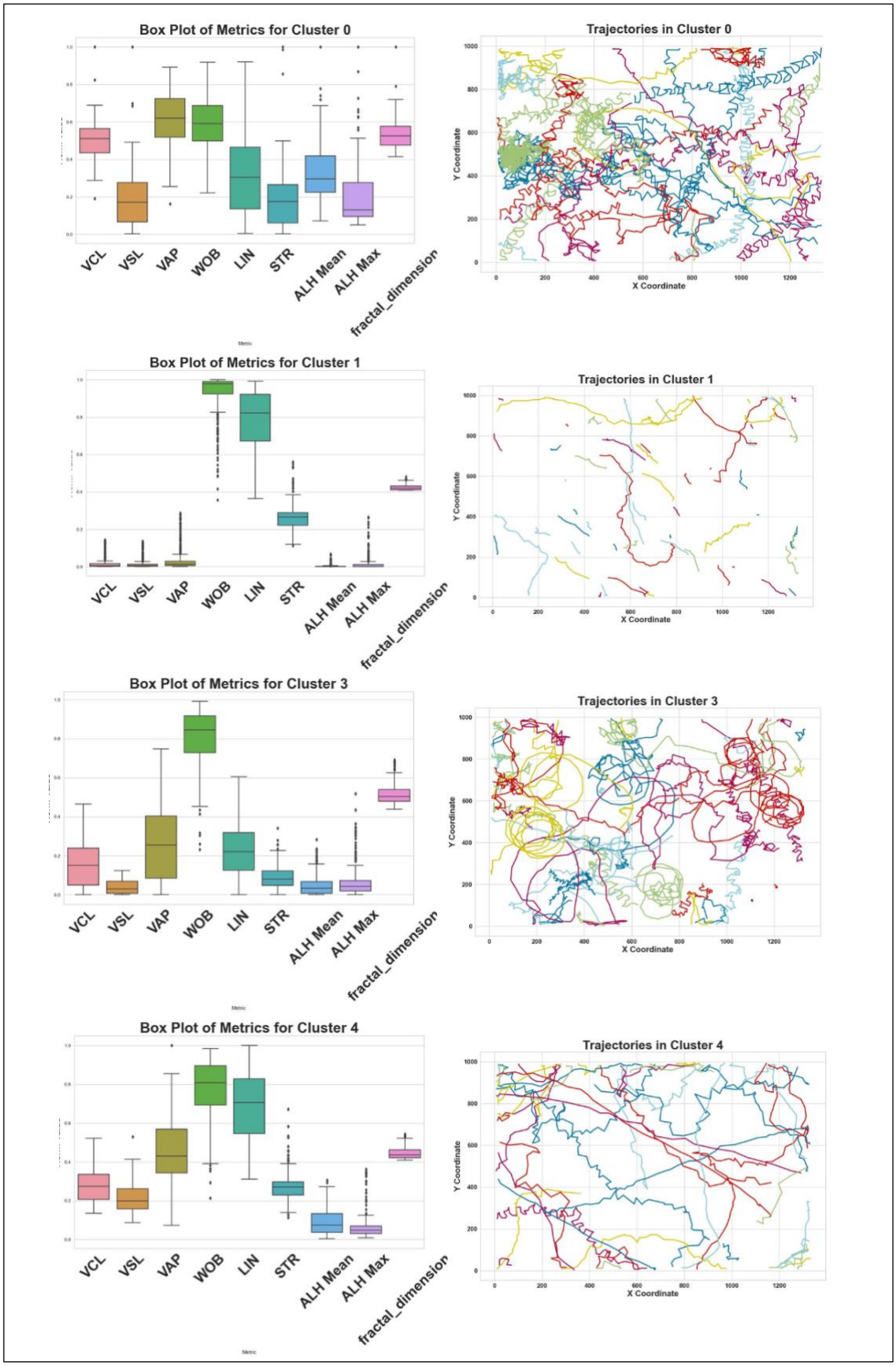

Fig. 5 Right Column: Box-and-whisker plot of the 9 kinematic parameters associated with each cluster. Left Column: Plotted trajectories of tracks in each cluster.

features across all clusters, underscoring velocity and linear movement as the key predictors in distinguishing between hyperactive, progressive, and immotile tracks.

This finding aligns with the inherent definitions of these tracks, as hyperactivity and progressiveness are predominantly differentiated by their linear movement. Linearity, which is calculated as the ratio of VSL to VCL, emerges as a critical predictor of efficient, nearly straight-line sperm movement towards the egg. The selection of these features also provides valuable insights into the motility metrics that most effectively characterize the three principal track patterns.

TABLE 2

Random Forest Feature Selection Ranking		
Rank	***Feature***	***Value***
1	VCL	0.200
2	VSL	0.175
3	LIN	0.150
4	ALH Mean	0.125
5	FD	0.100
6	STR	0.950
7	VAP	0.075
8	ALH Max	0.070
9	WOB	0.035

E. Hierarchical Clustering

The top 5 ranked metrics returned by Random Forest feature selection were used to cluster the 18 participants with similar sperm motility patterns using Hierarchical Clustering.

Cluster 1- (moderate progressive/weakly motile tracks)

Cluster 1 is composed of eight participants whose sperm exhibit motility with low velocity and some degree of forward progressivity. The trajectories for participants 60 and 52 are depicted in Figures 6a and 6b, illustrating linear movements that suggest moderate progressivity. This observation is corroborated by the supplementary videos. Further validation comes from commercial CASA system metadata, which provides the percentages of progressive, immotile, and non-progressive sperm for each participant, as seen in Fig. 8. Analysis of this data reveals that Cluster 1 participants generally have higher proportions of immotile and non-progressive sperm. However, there is a moderate level of progressive motility, averaging around 20% for most participants in this cluster, with the exception of participant 35, who is noted as an outlier.

Cluster 2 – Hyperactivated Tracks

Cluster 2, comprising two participants, is distinguished by high velocity, low linearity, and a high fractal dimension, indicative of hyperactivated sperm. The sperm tracks for these participants, as illustrated in Figures 6c and 6d, exhibit circular and hyperactivated patterns. The CASA data for participants 12 and 30 reflects a higher percentage of progressive and non-progressive motility, and a lower proportion of immotile sperm, compared to those in Cluster 1, as shown in Figure 8.

Hyperactivated sperm are known for their erratic swimming patterns, which cover less forward distance despite their vigorous velocity. Consequently, commercial CASA systems often classify this type of motility as “non-progressive” due to their lower overall displacement. Therefore, the use of hierarchical clustering based on descriptive kinematic features in this study provides a more nuanced perspective, potentially identifying valuable movement patterns that might be otherwise categorized as suboptimal by standard CASA evaluations.

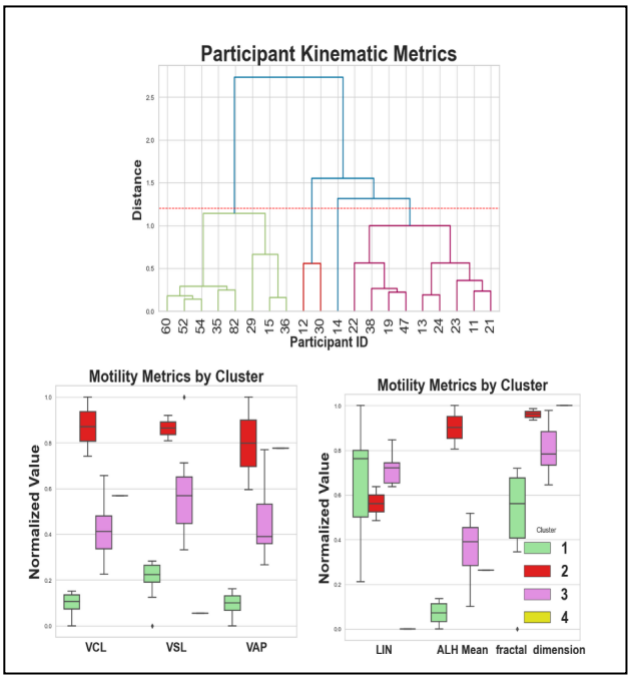

Fig. 7: (Top) Hierarchical Clustering of Participants based on motility patterns. (Bottom) Motility statistics by cluster colored corresponding to dendrogram branches.

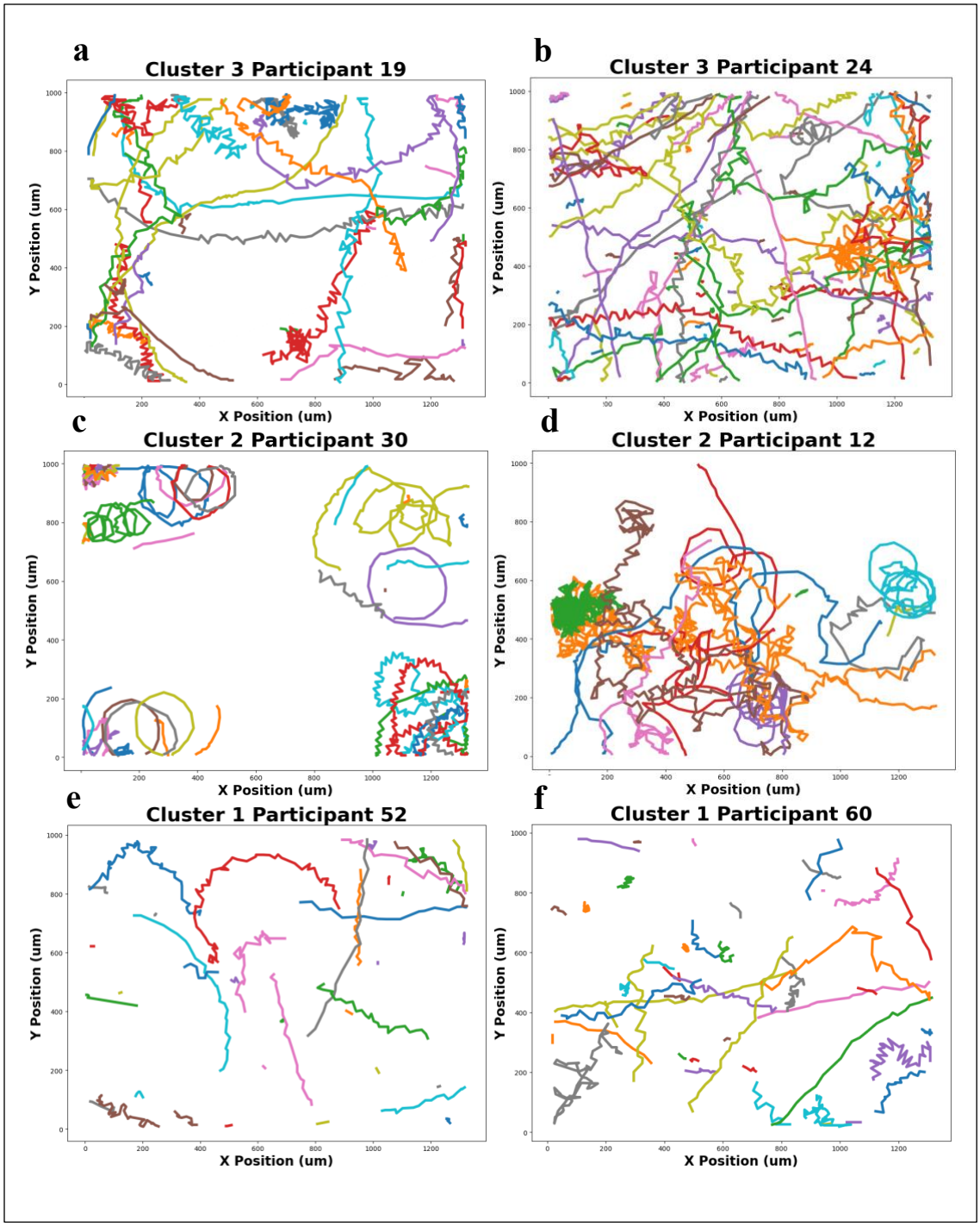

Fig. 6 (a-b) Participants in cluster 3 showing progressive trajectories. (c-d) Participants in Cluster 2 showing hyperactivated/circular trajectories. (e-f) Participants in Cluster 1 with showing weakly motile trajectories

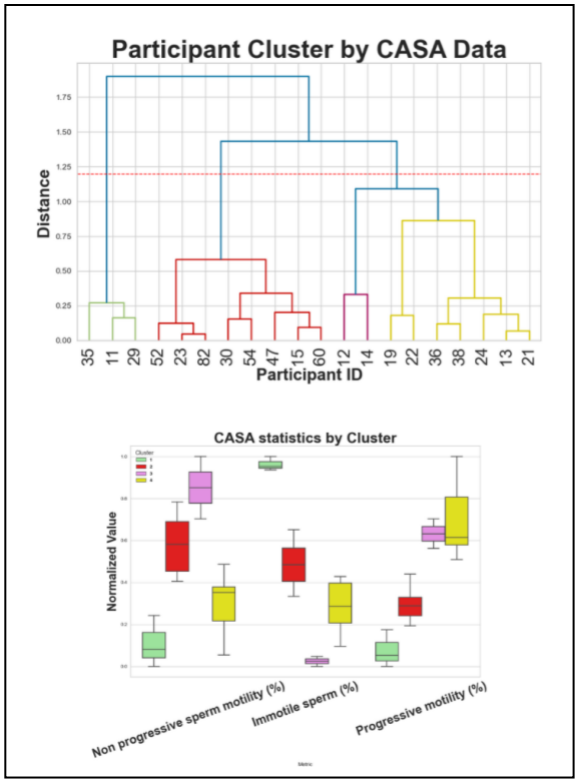

Fig. 8: (Top) Participant hierarchical clustering with CASA data features. Participants clustered based on distribution of motility patterns. (Bottom) CASA statistics by cluster. Box plots colored corresponding to dendrogram branches.

Cluster 3 - Linear Mean/Progressive Tracks

Cluster 3, marked in dark purple, comprises nine participants and is characterized by moderate to high velocity and linearity (LIN), suggesting the presence of linear progressive tracks. For the sake of conciseness, the study focused on the tracks of participants 19 and 24, who are positioned close together in the dendrogram (Fig. 6e-f). The trajectory plots for these participants show strong progressive and forward movement, bearing resemblance to Fig. 3b's trajectory.

The motility data from commercial CASA systems indicates high progressivity for these participants but fails to distinguish between the motility track patterns observed in Clusters 2 and 3. Interestingly, despite the high percentage of immotile sperm reported, which could imply similarity to the weakly motile tracks of Cluster 1, both the trajectory plots and supplementary videos demonstrate that this is not the case. This discrepancy highlights a potential limitation in commercial CASA systems' ability to accurately reflect the nuances of sperm motility patterns.

Cluster 4 – Outlier

Cluster 4 contains one participant and is therefore considered an outlier.

IV. References

[1] World Health Organization, 4 April 2023. [Online]. Available: https://www.who.int/news/item/04-04-2023-1-in-6-people-globally-affected-by-infertility. [Accessed December 2023].

[2] M. Balli, A. Cecchele, V. Pisaturo, G. Carullo, S. Makieva and E. Somigliana, "Opportunities and Limits of Conventional IVF versus ICSI: It Is Time to Come off the Fence," *J Clin Med.,* vol. 11, no. 19, 2022.

[3] R. Sciorio and S. C. Esteves, "Contemporary Use of ICSI and Epigenetic Risks to Future Generations," *Journal of Clinical Medicine,* vol. 11, no. 8, 2022.

[4] W. H. Cheon, H. J. Park, M. J. Park, M. Y. Lim, J. H. Park, B. J. Kang and N. C. Park, "Validation of a smartphone-based, computer-assisted sperm analysis system compared with laboratory-based manual microscopic semen analysis and computer-assisted semen analysis," *Investigative and Clinical Urology,* vol. 60, no. 5, 2019.

[5] J. B. You, C. McCallum, Y. Wang, J. Riordon, R. Nosrati and D. Sinton, "Machine learning for sperm selection," *Nature Reviews Urology,* vol. 18, pp. 387-403, 2021.

[6] V. Thambawita, S. A. Hicks, A. M. Storås, T. Nguyen, J. M. Anderson, O. Witczak, T. B. Haugen, H. L. Hammer, P. Halvorsen and M. A. Riegler, "VISEM-Tracking, a human spermatozoa tracking dataset," *Scientific Data,* vol. 10, 2023.

[7] G. Jocher, *YOLOv5 by Ultralytics,* 2020.

[8] C. Alquézar-Baeta, S. Gimeno-Martos, S. Miguel-Jimenéz, P. Santolaria, J. Yániz, I. Palacín, A. Casao, J. Á. Cebrián-Pérez, T. Muiño-Blanco and R. Pérez-Pé, "OpenCASA: A new open-source and scalable tool for sperm quality analysis," *PLoS Computat Biol.,* no. 1, 2019.

[9] F. Pedgrosa, Varoguax, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in {P}ython," *Journal of Machine Learning Research,* vol. 12, pp. 285-2830, 2011.

[10] J.-W. Choi, L. Urbano, P. Masson, M. VerMilyea and M. Kam, "Classification and Clustering of Human Sperm Swimming Patterns," *IEEE 3rd Internaltional Conference on Signal and Image Processing,* 2018.

[11] S. Goodson, S. White, A. Stevans, S. Bhat, K. Chia-Yu, S. Jaworski, T. Marlowe, M. Kohlmeir, L. McMillan, S. H. Zeisel and D. O. O'Brien, "CASAnova: a multiclass support vector machine model for the classification of

human sperm motility patterns," *Biology of Reproduction,* vol. 97, no. 5, pp. 698-708, 2017.