NAZARBAYEV UNIVERSITY

MORPHO-FUNCTIONAL ANALYSIS OF BOVINE SPERMATOZOA USING CYTOMETRY WITH DEEP LEARNING ALGORITHMS

Anel Umirbaeva

BSc. in Biological Sciences, Nazarbayev University

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF SCIENCE IN BIOLOGICAL SCIENCES DEPARTMENT OF BIOLOGY SCHOOL OF SCIENCE AND HUMANITIES NAZARBAYEV UNIVERSITY

2026

Student	Anel Umirbaeva	24/04/2026
Student's Primary Supervisor	Professor Natalie Barteneva	24/04/2026
Student's Secondary Supervisor	Professor Ivan Vorobyev	24/04/2026

DECLARATION

I hereby declare that the thesis is my original work, and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

_____

Anel Umirbaeva
23.04.2026

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my research supervisors and mentors, Professor Natalie Barteneva and Professor Ivan Vorobyev, for their invaluable guidance, continuous support, and trust throughout this work. I am also grateful to my colleague and dear friend, Andrey Kurenkov, for developing the deep learning algorithm and co-authoring the resulting manuscript that has been published in Frontiers *in Veterinary Sciences* (see Appendix). To all my friends and family, who stayed strong alongside me and invigorated me throughout this challenging path, I say *Rakhmet.*

I would also like to extend my appreciation to Dr. Olena Filchakova for providing Fluo-4 fluorescent dye and to Dr. Vadim Mustyatsa for his assistance with live imaging of spermatozoa using fluorescence microscopy, and to Ms. Aigul Kussanova for her technical support during microscopy experiments. I am also thankful to Dr. Bolat Seisenov and the Republican Center "Assyl-Tulik" for providing fresh and cryopreserved bovine sperm samples, as well as CASA-derived motility data.

TABLE OF CONTENTS

SUMMARY

This study presents a high-throughput approach for single-cell morpho-functional analysis of bovine spermatozoa by integrating imaging flow cytometry with deep learning. A dataset of over 400,000 sperm images was used to train a ConvNeXt-Tiny convolutional neural network for automated classification into ten morphological categories. Functional analysis was performed on approximately 60,000 images obtained from cryopreserved semen, resulting in a final dataset of 40,613 high-quality single-cell events annotated with mitochondrial membrane potential (MMP) and intracellular calcium levels, stained with tetramethylrhodamine ethyl ester perchlorate (TMRE) and Fluo-4, respectively. Statistical analysis using the Kruskal–Wallis test revealed significant differences in calcium levels across morphological groups ($p < 0.001$), with post-hoc Dunn's test identifying multiple group-level differences. Results demonstrated that abnormal midpiece, coiled tail, and distal cytoplasmic droplet groups exhibited distinct functional profiles, including altered mitochondrial activity and calcium signaling. This study represents the first large-scale morpho-functional analysis of spermatozoa from Kazakhstani cattle breeds at single-cell resolution. The developed pipeline provides a foundation for improved fertility assessment and highlights the potential of AI-driven multi-parametric analysis in reproductive biology.

LIST OF TABLES

LIST OF FIGURES AND ILLUSTRATIONS

predicted labels on horizontal axis. Correct predictions are depicted on the diagonal cells.

Figure 6. Gating procedure of focused single cells; within focused cells, events were represented as channel 3 (TMRE staining) intensity histogram with gating for Low and High MMP groups.

Figure 7. Images of spermatozoa stained with Fluo-4 and TMRE dyes for intracellular calcium and mitochondrial membrane potential measurements with IFC (images taken with 40x magnification).

Figure 8. Distribution of Fluo-4 intensity across sperm morphological groups

Figure 9. Images of spermatozoa stained with DNA dye (Hoechst 33342) and acquired with Imagestream X Mark II (40x magnification).

ABBREVIATIONS

AI	Artificial intelligence
AM	Abnormal midpiece
ART	Assisted reproductive technology
AT	Abnormal tail
CASA	Computer-assisted sperm analysis
CI	Confidence interval
CNN	Convolutional neural network
CTM	Coiled tail and midpiece
DCD	Distal cytoplasmic droplet
DsRed	Discosoma sp. Red fluorescent protein
FBS	Fetal bovine serum
FC	Flow cytometry
GFP	Green fluorescent protein
IACUC	International animal care and use committee
IFC	Imaging flow cytometry
IHS	Irregular head shape
IVF	In vitro fertilization
LED	Light-emitting diode
LOBO	Leave-one-breed-out
LP	Linear probing
LP-FT	Linear probing and fine tuning
MMP	Mitochondrial membrane potential
mPTP	Mitochondrial permeability transition pore
NM	Normal morphology
PBS	Phosphate buffered saline

PCD	Proximal cytoplasmic droplet
PM	Progressive motility
sCMOS	Scientific complementary metal-oxide semiconductor
TEH	Twisted and elongated head
TM	Total motility
VAP	Average path velocity

1. INTRODUCTION

1.1. Bovine semen quality assessment for assisted reproduction technologies

Artificial insemination is a technology used in cattle reproduction for controlled genetic breeding, prevention of disease spread, and improved production efficiency. This technology requires the selection of semen of the highest quality for successful fertilization and pregnancy rates in cows which is largely dependent on sperm quality. Gold standards for semen evaluation include motility, morphology, and viability assays coupled with other measurements, such as seminal plasma pH, cell count, cell concentration, and ejaculate volume (Perry, 2021; Bollwein&Malama, 2023). While these are well-established and critical features, to have a higher fertilization prediction accuracy and more precise diagnostics of diseases, it is important to evaluate functional parameters of sperm health as well, which are prone to variations due to changing external factors (De Lima Rosa et al., 2023; Odinius et al., 2024).

Bovine semen quality might exhibit considerable heterogeneity, not only among breeds and individual bulls but also across ejaculates from the same bull between seasons. Factors, such as age, breed, collection interval, environmental conditions, and nutritional changes influence sperm morphology and functional parameters including motility, viability, and MMP. Seasonal and individual variability further complicate semen assessment, necessitating emergence of techniques capable of single-cell evaluation of spermatozoa heterogeneity (Vincent et al., 2014; Murphy et al., 2018; Gloria et al., 2023).

Another factor, cryopreservation, serving as a common practice for transport and genetic selection, introduces additional challenges by reducing sperm viability and motility as freezing and thawing can damage the plasma membrane and acrosome, leading to cell death (Ozkavukcu et al., 2008; Hezavehei et al., 2018). Conventional staining techniques for morphological

assessments, especially those requiring fixation, can mask these subtle changes (Rodriguez-Martinez, 2006; Hanson et al., 2023). In addition, classic microscopic techniques are only capable of providing such information for about 100-200 cells per slide, requiring expert labeling, extensive staining, and manual counting, which pose challenges for an accurate sperm analysis (Dai et al., 2021). Together, all these limitations underline the necessity of high-throughput, multi-parametric approaches for simultaneous evaluation of morphological and functional indicators of sperm quality.

1.2. Morphological and functional analysis of sperm cells

Bovine spermatozoon consists of three major regions: head, middle piece, and tail, all of which are surrounded by a plasma membrane; a detailed description of the spermatozoon is provided by Perry (2021). Despite the widespread use of cryopreservation in bovine reproduction, the mechanisms underlying sperm cryosurvival are still not completely understood (Hezavehei et al., 2018; Grötter et al., 2019). The freeze–thaw process induces multiple structural and biochemical alterations that affect sperm functionality, including damage to the plasma membrane (Kumar et al., 2024; Garcia-Oliveros et al., 2025), acrosome (Ahmad et al., 2014; Park et al., 2023), and chromatin (Khalil et al., 2017; Palazzese et al., 2020), as well as ionic and osmotic imbalances and activation of proteases leading to further degradation of cellular components (Lum and Blobel, 1997; Gurupriya et al., 2014; Bogle et al., 2017). Disruption of the acrosome renders a spermatozoon incapable of penetrating the zona pellucida and affects the capability of the spermatozoon to fertilize the oocyte (Kumar et al., 2024). The middle piece of the spermatozoa contains the mitochondria, which participate in conversion of sugars, primarily glucose, into ATP. Destruction of the mitochondria by improper cooling or freezing can render the spermatozoa unable to generate energy necessary for oocyte fertilisation.

Damage to the spermatozoa tail will alter the ability of spermatozoa to swim to the oocyte, penetrate the zona pellucida, and fertilize the oocyte. Moreover, increased production of reactive oxygen and nitrogen species during cryopreservation places spermatozoa under oxidative stress, impairing viability, DNA integrity, and motility (Aitken et al., 1989; Lewis et al., 1996; Chatterjee et al., 2001; Gadea et al., 2008; Pintus et al., 2018; Kadlec et al., 2020). As a result, thawed samples show a high incidence of morphological abnormalities, including detached or abnormal heads, bent or coiled tails, damaged midpieces, and other structural deformities (Palacin et al., 2020). Additionally, some spermatozoa undergo "cryo-capacitation," a premature capacitation-like activation of cellular processes that reduces their lifespan and fertilizing potential (Bailey et al., 2003).

Cell morphology changes are closely related to the functional impairments. For example, defects in the sperm head have been linked to chromatin re-organization, potentially compromising nuclear membrane integrity (Prisant et al., 2007), and altered expression of proteins that are crucial for capacitation and oocyte-interaction processes (Saadi et al., 2013). At the same time, abnormalities of the tail and midpiece regions are associated with decreased motility and mitochondrial dysfunction, with severe flagellar deformations resulting in complete loss of motility and fertility (Thundathil et al., 2014; Mai et al., 2024). Other structural changes, including excess residual cytoplasm, may indicate incomplete maturation and mechanical damage due to high osmotic pressure, leading to reduced viability (Meyers, 2005; Rengan et al., 2012). However, morphology alone is insufficient to assess sperm quality, as cells with normal morphology may still harbor functional abnormalities (Prisant et al., 2007; Saadi et al., 2013; Bulkeley et al., 2023). Therefore, there is a need for simultaneous evaluation of morphological and functional parameters of individual cells that reflect their physiological state.

Mitochondrial membrane potential (MMP) is a major indicator of mitochondrial function, reflecting the electrochemical gradient across the inner mitochondrial membrane that drives ATP synthesis (Zhang et al., 2025). In spermatozoa, high MMP is directly associated with motility and fertilizing capacity (Thundathil et al., 2014; Mai et al., 2024; Zhang et al., 2025). However, during cryopreservation, increased production of reactive oxygen and nitrogen species leads to lipid peroxidation, opening of the mitochondrial permeability transition pore (mPTP), and dissipation of MMP (Aitken et al., 1989; Lewis et al., 1996; Chatterjee et al., 2001; Gadea et al., 2008; Pintus et al., 2018; Kadlec et al., 2020). These damages lead to reduced ATP production and flagellar movement, potentially initiating cell death pathways (Upadhyay et al., 2021; Hai et al., 2024, 2025). Mitochondrial stains (JC-1, TMRE, and others) (Marchetti et al., 2004) can differentiate between poorly functional and highly functional mitochondria in spermatozoa.

Calcium signaling is another essential functional parameter that regulates sperm motility, hyperactivation, capacitation, and acrosome reaction (Vicente-Carrillo et al., 2023; Antonouli et al., 2024; Kumar et al., 2025). Cryopreservation also affects intracellular calcium homeostasis through membrane damage and premature capacitation-like changes (Bailey et al., 2003; Antonouli et al., 2024; Kumar et al., 2024; Garcia-Oliveros et al., 2025). MMP and calcium regulation are tightly interconnected, as mitochondria buffer cytosolic Ca^{2+} via the mitochondrial calcium uniporter depending on membrane potential (Zhang et al., 2025). While moderate Ca^{2+} uptake supports ATP production, excessive accumulation induces mPTP opening, MMP decrease, and release of pro-apoptotic factors (Upadhyay et al., 2021; Zhang et al., 2025; Hai et al., 2024, 2025), negatively affecting the viability of cells. Therefore, assessing MMP and intracellular calcium together, alongside morphological classification, provides a more integrated

understanding of how structural abnormalities relate to functional impairment (Prisant et al., 2007; Saadi et al., 2013; Bulkeley et al., 2023).

1.3. Imaging flow cytometry and artificial intelligence algorithms for high-throughput single-cell analysis

Imaging flow cytometry (IFC) is a hybrid technology that combines flow cytometry with imaging, allowing analysis of tens of thousands of single cells in one run (Basiji, 2016; Barteneva et al., 2012; Barteneva & Vorobjev, 2018). For each cell, IFC simultaneously captures brightfield and darkfield (side scatter) images for morphology and fluorescence signals for functional parameters. This technique extracts hundreds of features per cell, which is especially useful for spermatozoa given their complex structure and the multiple coexisting abnormalities within the same cell (Buckman et al., 2009; Barth et al., 2025; Teves et al., 2025).

Though conventional morphological assessment under a light microscope remains a gold standard, it has several limitations, including subjectivity, low-throughput, excessive reliance on experts, and inability to accurately categorize subtle or complex defects. (Douglas et al., 2021; Javadi & Mirroshandel, 2019; Riordon et al., 2019). IFC addresses these limitations by generating large image datasets of single cells. For rapid analysis and testing of emerging data, convolutional neural network (CNNs) models can be applied. CNNs are a class of deep learning algorithms that learn discriminative features directly from raw images, eliminating the need for manual analysis (Cherouveim et al., 2023; Guahmich et al., 2023; Ghayda et al., 2024; Keller et al., 2025). An expert-annotated dataset is required for CNN training to classify spermatozoa into morphological groups with high consistency and accuracy (Ghayda et al., 2024).

A combination of IFC with CNN models enables detailed classification of morphological anomalies in spermatozoa and the evaluation of relationships between morphological and

functional parameters of the same cells in the tens of thousands (Bulkeley et al., 2023). This method provides a possibility to estimate the translation of a specific structural defect to its functional impairment, offering a rapid, multi-parametric morpho-functional analysis of semen quality (Bulkeley et al., 2023; Saadi et al., 2013; Prisant et al., 2007).

Despite these advances, existing studies on CNN-based sperm analysis have been limited in several key respects. Most have focused exclusively on human spermatozoa or, in animal models, on a single breed such as Holstein-Friesian or Large White boar, with datasets typically comprising fewer than 50,000 images (Keller et al., 2025; Aktas et al., 2024). Furthermore, no study to date has trained and evaluated deep learning models on paired fresh and cryopreserved samples from the same individuals, nor has any investigated sperm morphology in native Kazakh cattle breeds. The present study addresses these gaps by: (i) assembling the largest reported dataset of high-resolution sperm images for deep learning training, comprising over 400,000 single IFC images; (ii) training and evaluating models on both fresh and frozen bovine semen from the same bulls, enabling direct assessment of cryopreservation-induced domain shifts; and (iii) including, for the first time, to the best of our knowledge, native Kazakh breeds (the Kazakh Whitehead and Auliekol) which are economically vital to regional livestock production yet genetically distinct from breeds that dominate global reproductive research. This combination of dataset scale, cross-condition experimental design, and inclusion of underrepresented breeds establishes a new benchmark for morpho-functional sperm analysis using IFC and artificial intelligence.

2. AIMS OF THE PROJECT

The primary goal of this thesis is to establish a high-throughput, multi-parametric pipeline for the morpho-functional analysis of bovine spermatozoa by using imaging flow cytometry and deep learning-based artificial intelligence. This pipeline can be further applied to characterize the relationships between morphological abnormalities and key functional parameters (mitochondrial membrane potential and intracellular calcium levels) of single sperm cells of cryopreserved bovine semen.

Aim 1: Create an image dataset using imaging flow cytometry and apply a deep learning algorithm for morphological classification of single cell images;

Aim 2: Measure functional parameters of live, cryopreserved bovine semen using imaging flow cytometry;

Aim 3: Integrate the deep learning-based morphological categorization algorithm with functional data from individual sperm cells and perform statistical analysis.

3. MATERIALS AND METHODS

3.1. Animals and semen collection

Fresh and frozen sperm samples from six beef bull sires (four Kazakh Whitehead, one Simmental, and one Auliekol) were used in this study. All bulls were maintained under similar management conditions at the Republican Center "Asyl-Tulik" for breeding in livestock. Two fresh ejaculates per bull were collected during the spring and fall seasons. Cryopreserved semen samples of six bulls, with total motility above 60%, were obtained during the spring-summer season and diluted with OptiXcell® (IMV Technologies, France) cryoprotectant media in a 1:1 ratio. The resulting semen was stored in 0.25 ml straws with an average concentration of approximately 0.5 billion cells/ml, resulting in a final concentration of about 250 million cells/ml after dilution in Dulbecco's Phosphate Buffered Saline (DPBS) (Thermo Fisher Scientific, USA). Frozen samples were then thawed at 37°C in the water bath for 30 seconds prior to analysis.

The fresh sperm samples were not collected for a specific research project, but rather were byproducts of the standard production process and were exempt from the oversight of Nazarbayev University Institutional Animal Care and Use Committee (IACUC). All the procedures in the breeding company were conducted in accordance with relevant animal procedural policies to ensure animal wellbeing and compliance with internal ethical regulations.

3.2. CASA analysis of motility

Computer-assisted sperm analysis (CASA) was performed for both fresh and frozen sperm motility evaluation using the HTM-2030 Motility Analyzer (Hamilton Thorne LTD., USA). Parameters assessed included cell count, progressive motility (PM, %), total motility (TM, %), average path velocity (VAP um/s), straight line velocity (VSL, um/s), curvilinear velocity (VCL, um/s), amplitude of lateral head (ALH, um), beat cross frequency (BCF, Hz),

straightness (STR, %), linearity (LIN, %), elongation (Elong, %), and area of a sperm head (Area, um^2). Approximately 100 spermatozoa per sample were analyzed across five fields of view. Concentration of fresh spermatozoa was determined using a photometer FEK-M (HV-Lab, RF), conducted by a technician at the livestock breeding center for routine assessments and further preparation of frozen semen straws. Results of CASA analysis for each of the semen samples obtained are provided in Appendix Table A1.

3.3. Data acquisition from imaging flow cytometry for deep-learning model training

Each semen sample, both fresh and frozen cells, for all bulls was used for IFC-based morphology assessment. Frozen samples were taken as 0.25 ml straws, thawed at 37°C in the water bath, and analyzed immediately under IFC with a maximum of 10 million cells/ml concentration. Fresh semen was also diluted with phosphate buffered saline (PBS) and kept at 37°C all the time during the data acquisition process. Spermatozoa images were acquired using the ImageStreamX Mark II imaging flow cytometer (Amnis-Cytek, USA) equipped with 405, 488, 561, and 642 nm lasers, using a 40× objective. For each sample, two files containing 50,000 events were recorded. Initial datasets contained approximately 1.8 million images, and gating was applied to select focused singlets. The average number of extracted brightfield images (from Channel 1) was about 25,000 per sample, with additional details given in Appendix Table A2. The next step included preparation of extracted images for deep-learning model training and final analysis to reduce the chances of data overfitting and the emergence of noise. Center cropping and filtering were applied, and images with shorter edges of less than 50 pixels were discarded. The choice of this specific threshold was because frames at or below this scale were mostly cellular debris or lacked detectable morphological cues that would allow quality labeling.

The resulting dataset of images after filtering out debris and non-identifiable events contained 401,535 images (257,844 for fresh and 143,691 for frozen) and was utilized for all the following steps.

3.4. Morphology assessment

Images were manually annotated into eight morphological categories: normal morphology (NM), irregular head shape (IHS), abnormal tail (AT), abnormal midpiece (AM), twisted and elongated head (TEH), proximal cytoplasmic droplet (PCD), distal cytoplasmic droplet (DCD), and coiled tail and midpiece (CTM) (WHO, 2021). Additional residual categories for debris and multiple events were included for dataset quality control (making 10 categories in total). Annotation was performed by two independent raters. The following rules were applied for minimal overlap of the groups: 1) “Multiple” and “Debris” were marked out for a clear dataset and future application models; 2) TEH and IHS: TEH was set for spermatozoa with twisted (rotated) or elongated heads, and IHS was assigned for other head defects; 3) CTM substitutes AT group when coiling of a tail or midpiece region is obvious. IDEAS vs. 6.2 (Amnis-Cytek, USA) software was utilized for data browsing and filtering. Image labeling was conducted by two independent raters, with disagreements resolved by consensus. Stratified random sampling relative to bull samples was applied until 200 labeled images per class (total 10 classes) per breed per condition were obtained, resulting in 12,000 total number of images in the model training dataset with a uniform distribution across classes, bulls, breeds, and season conditions. Representative single spermatozoa images of morphological classes are given in Figure 1.

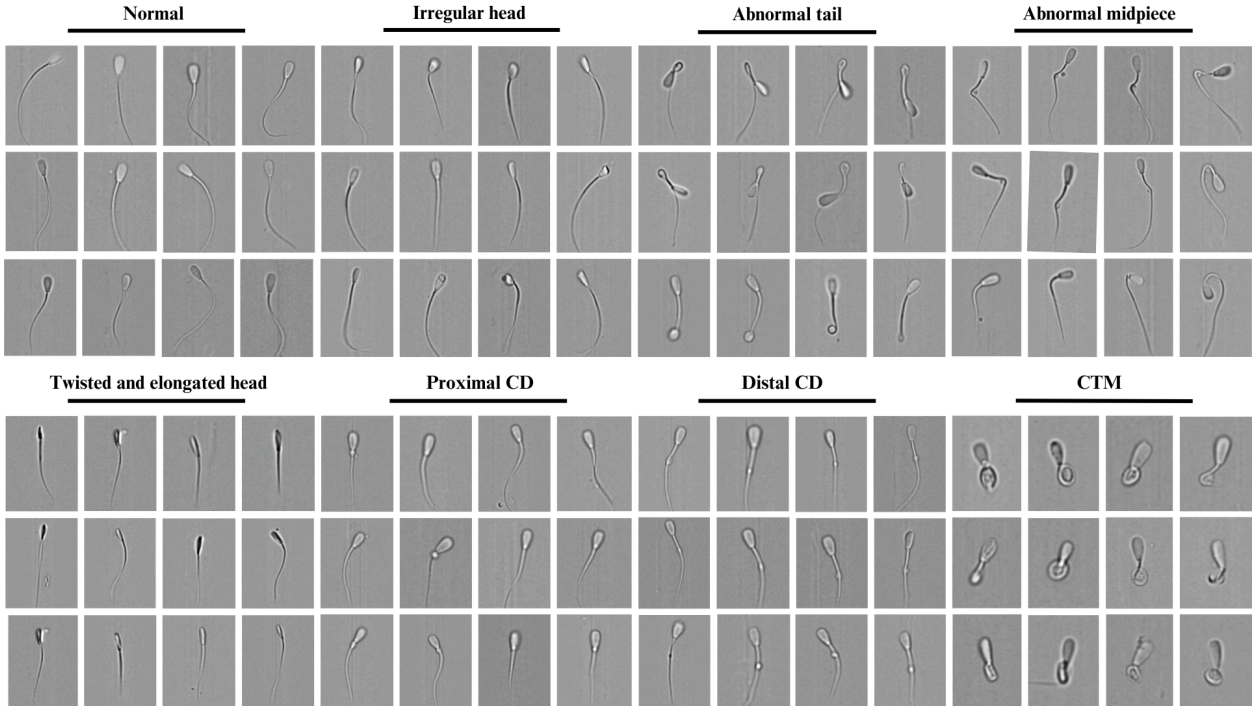


Figure 1. Brightfield images of bovine spermatozoa acquired with ImageStream X Mark II (Amnis-Cytek, USA) using 40x magnification. The image dataset was divided into eight distinct morphological groups: normal morphology (NM), irregular head shape (IHS), abnormal tail (AT), abnormal midpiece (AM), twisted and elongated head (TEH), proximal cytoplasmic droplet (PCD), distal cytoplasmic droplet (DCD), and coiled tail and midpiece (CTM). Other two groups - debris and multiple cells were not included in the figure.

3.5. Model training pipeline

3.5.1. Data preparation

Brightfield images were converted to RGB format and resized to 224×224 pixels. Data augmentation was applied to the training dataset, including random horizontal flips, small rotations (±10°), and random resized cropping with minor translations. Figure 2 depicts the analysis pipeline of data selection and processing.

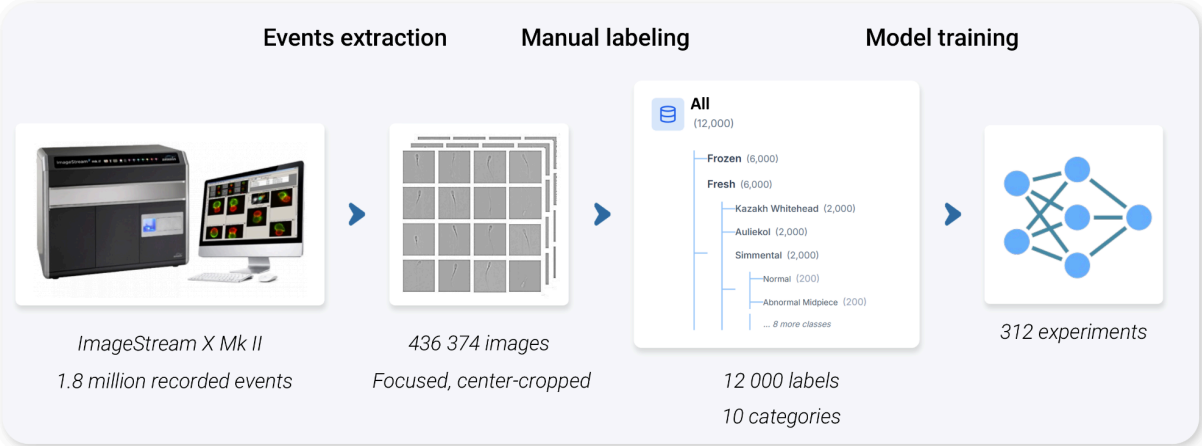


Figure 2. General overview of the analysis pipeline for the data selection, processing, and analysis stages. Images acquired with ImageStream X Mark II (Amnis-Cytek, USA) were gated for single, focused events and preprocessed with center-cropping to a square ratio for uniform resizing. The next step included manual labeling of images by two independent experts into ten morphological categories. Afterwards, multiple deep learning algorithms were trained and evaluated in the series of experiments, resulting in a final selected model used for further image classifications.

3.5.2. Training and validation strategy

Model training was performed using class-balanced sampling and stratification by bull and breed on all available images, with online augmentation to ensure fair comparisons across experiments. Per-breed and per-bull caps were also applied to prevent data volume from standing as a confounder without discarding annotated data. Identical curation was utilized each time experiments were limited to a single breed or condition for consistent class distribution and training budget across scenarios. Multiple experiments were performed: 1) to select efficient architectural models and training strategies; 2) to estimate the generalization of the chosen model across all conditions and breeds.

Model generalization was assessed using two protocols: cross-condition transfer (one model trained on all frozen images, the other on all fresh images) and cross-breed transfer within

a single condition (a model trained on three single breeds per condition). Then, each model was evaluated both within training conditions and across the others. For unbiased domain-shift analysis, the same training budget and class stratification were used.

The Leave-one-breed-out (LOBO) method was used for model selection treating the held-out breed as a test and reserving 20% of the data for per-epoch validation. The same model selection procedure was run at different labeled-data fractions $f \in \{0.1, 0.2, 0.5, 1.0\}$. Fixed stratified train/validate/test splits (80/10/10) were used for generalization experiments, without any cross-validation. Explanation of label usage for all experiments is given in the Appendix Table A3. All the results are noted over three seeds (unless mentioned otherwise), and there is a report of 95% confidence intervals for accuracies via a hierarchical cluster bootstrap (resampling bulls, then samples, then images, B=5,000) to account for intra-bull dependence. Differences are considered significant when confidence interval CI does not include 0.

Four commonly used in biological imaging and computer vision CNN backbones were used including MobileNetV3-Large, EfficientNetV2-S, ResNet-50, ConvNeXt-Tiny (He et al., 2016; Liu et al., 2022; Tan and Le, 2021; Howard et al., 2019) under two regimes: linear probe with fine-tuning (LP-FT) and standard linear-probing (LP) (Kumar et al., 2022; Kurenkov et al., 2024). The LP-FT regime includes training the classifier first with the backbone frozen at ImageNet-pretrained weights, followed by end-to-end fine-tuning with a reduced learning rate. PyTorch with GPU acceleration (NVIDIA GeForce RTX 4080) was used for models, trained with the AdamW optimizer and cross-entropy loss. Batch size was 64, and learning rates were 1×10^{-3} for linear probing and 1×10^{-4} for fine-tuning. Performance was scanned on the validation split at each epoch to record the best checkpoint based on validation accuracy and ensure the early-stopping callback to avoid stagnating training. The F1-score balanced the trade-off between

false positives and false negatives; the macro average calculated the F1 for each class independently and averaged classes, treating them all equally.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Because the evaluation splits were stratified and performance across classes was consistent, macro-F1 values were comparable to accuracy. Therefore, accuracy is presented in Table 1, while macro-F1 scores are reported in the Appendix Table A4.

3.6. Fluorescence microscopy analysis of MMP and intracellular calcium levels

Frozen bovine sperm straws were used for fluorescence microscopy image generation. One straw per experiment was thawed for 30 sec at 37ºC and diluted in Hanks' Balanced Salt Solution, HBSS, (Sigma-Aldrich, USA) supplemented with 5% Fetal Bovine Serum, FBS, (Thermo Fisher Scientific, USA) at a 1:4 ratio, pre-warmed to 37°C in a water bath. Calcium staining was performed by adding Fluo-4 (MedChemExpress, USA), prepared in DMSO (Sigma-Aldrich, USA), and diluted in the sample to a final concentration of 1 μM, followed by incubation for 30 min in the dark at 37°C. After incubation, cells were washed once by centrifugation (600 g, 6 min, 4ºC), the supernatant was removed, and the pellet was resuspended in fresh medium, followed by an additional 10 min incubation. The second staining step was performed using TMRE (Thermo Fisher Scientific, USA) at a final concentration of 100 nM. Cells were incubated for 15 min and washed by centrifugation (600 g, 4 min, 4ºC). The final sample preparation step included dilution in FBS-free medium and incubation for an additional 10 min at 37°C for semen stabilization.

Prepared samples of live cells were poured onto glass slides covered with 0.1 mg/ml poly-l-lysine in PBS (Elabscience, USA) for spermatozoa immobilization and incubated for 10 min at 37ºC in the microscope chamber with the cover slip on. Fluorescence imaging was performed using an inverted fluorescence microscope (Axio Observer 5, ZEISS, Germany) with

a 40x oil immersion objective. Fluo-4 and TMRE signals were detected using GFP and DsRed filter sets, respectively. Illumination was provided by a LED light source at appropriate excitation wavelengths. Images were acquired using a sCMOS camera (Zeiss Axiocam, Germany), sequentially for each fluorescence channel, with the excitation light source switched off between acquisitions to minimize photobleaching and spectral crosstalk. Only fluorescence channels were recorded, without brightfield imaging. The microscope chamber and stage were maintained at 37°C throughout image acquisition.

3.7. IFC analysis of MMP, calcium levels, and morphology of frozen sperm

Frozen bovine semen (n=3, straws from a Kazakh Whitehead breed, cryopreserved in January 2026) was used for IFC analysis, and the staining procedure with Fluo-4 and TMRE was identical to one described in section 3.6. Cells were thawed at 37ºC, with the final cell concentration of about 10 million cells/ml. Spermatozoa images were acquired using ImageStreamX Mark II (Amnis-Cytek, USA) imaging flow cytometry, using a 40x objective. Brightfield images were collected while fluorescence signals were acquired using the 488 nm and 561 nm lasers for excitation of Fluo-4 and TMRE stains, respectively, with emission detected in the corresponding fluorescence channels. Spectral compensation was performed to correct for spectral overlap between channels. A compensation matrix was generated using a single-stained control sample for each fluorophore and applied to all datasets prior to downstream analysis using IDEAS software (Amnis-Cytek, USA).

A small proportion of events displayed negative fluorescence intensity values in the TMRE channel as a result of background subtraction and spectral compensation applied during data processing in IDEAS software. As these events could not be reliably assigned to either high or low MMP populations, they were excluded from further MMP-based analysis. Therefore, the

combined proportions of high and low MMP cells do not sum to 100% for all morphological groups, with the unassigned fraction accounting for approximately 5-6% of events per group.

3.8. Deep learning model for multi-parametric sperm evaluation

For enabling an integrated morpho-functional analysis at the single-cell level, the trained convolutional neural network (ConvNeXt-Tiny) was applied to imaging flow cytometry datasets containing fluorescence information. These datasets were generated from three (n=3) cryopreserved semen straws, with four independent acquisition files per straw, each containing exactly 5,000 events. In total, there were 20,000 extracted images from one sample and 60,000 overall. The acquired images underwent preprocessing to ensure data quality and consistency. This included center cropping of brightfield images to standardize cell positioning, removal of low-quality or unfocused frames based on image features, and exclusion of debris and non-single-cell events using gating strategies in IDEAS software. Subsequently, fluorescence intensity data corresponding to intracellular calcium (Fluo-4, Channel 2) and mitochondrial membrane potential (TMRE, Channel 3) were extracted for each event. Only cells meeting predefined quality and fluorescence criteria were retained for downstream morpho-functional analysis.

Morphological classification of each individual spermatozoon was performed using the pretrained ConvNeXt-Tiny model, categorizing cells into ten morphological categories. The resulting dataset combined morphological labels with functional parameters (MMP and Fluo-4 intensity) at single-cell resolution. Statistical analysis was performed in Python (scipy) and RStudio (version 2026.01.2). Differences in Fluo-4 intensity across morphological groups were evaluated using the Kruskal-Wallis test, followed by Dunn’s post-hoc pairwise comparisons with

Bonferroni correction for multiple testing. This non-parametric approach was selected due to the non-normal distribution of fluorescence intensity data and unequal group sizes.

4. RESULTS

4.1. Fluorescence microscopy image acquisition

Initial experimental design for morpho-functional analysis of spermatozoa included observation and analysis of cells using fluorescence microscopy. However, as can be seen on Figure 3, there is a limited number of images per slide, with photobleaching occurring rapidly under light exposure. Therefore, for further data acquisition and semen assessment, it was determined to utilize IFC for high-throughput data.

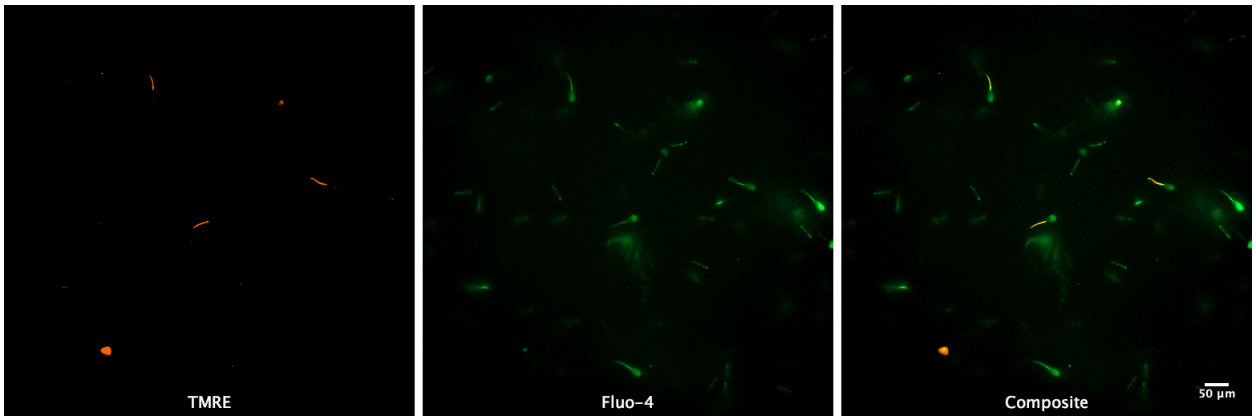


Figure 3. Frozen bovine double staining with TMRE and Fluo-4 stains, analysed using fluorescence microscope with a 40x oil immersion objective.

4.2. Model selection

Four various deep learning architectures were compared with LP and LP-FT strategies on the prepared datasets and repeated training at 0.1, 0.2, 0.5, and 1.0 fractions of the labeled images to assess sample efficiency, with the results given in Figure 4. The same fixed, stratified test split per seed was used for each fraction for reflecting differences of training data volume rather than test drift. During LP application, accuracy increased uniformly with the data, but the spread between models remained substantial. ConvNeXt-Tiny consistently led in every fraction with other backbones staying below the approximately 70% even at full datasets, depicting

underfitting and unstable convergence when only the classifier is trained. In case of LP-FT application, all backbones improved considerably with the narrowing gaps. However, ConvNeXt-Tiny remained best across fractions with 91.1% highest accuracy as shown on Figures 4 for LP-FT strategy and Figure 5 for the Multiple group. The graphs represent a steep gain from 0.1 to 0.2, followed by progressively smaller improvements; by point 0.5, the models approached their asymptotes. This pattern shows that 1) end-to-end adaptation is high-performing in learning robust representations for IFC images, and 2) ConvNeXt-Tiny is the most efficient model among the tested ones. Considering the model's consistent high performance across data regimens, strong stability under LP-FT, and computational efficiency, ConvNeXt-Tiny with LP-FT was chosen as the main architecture for further analyses.

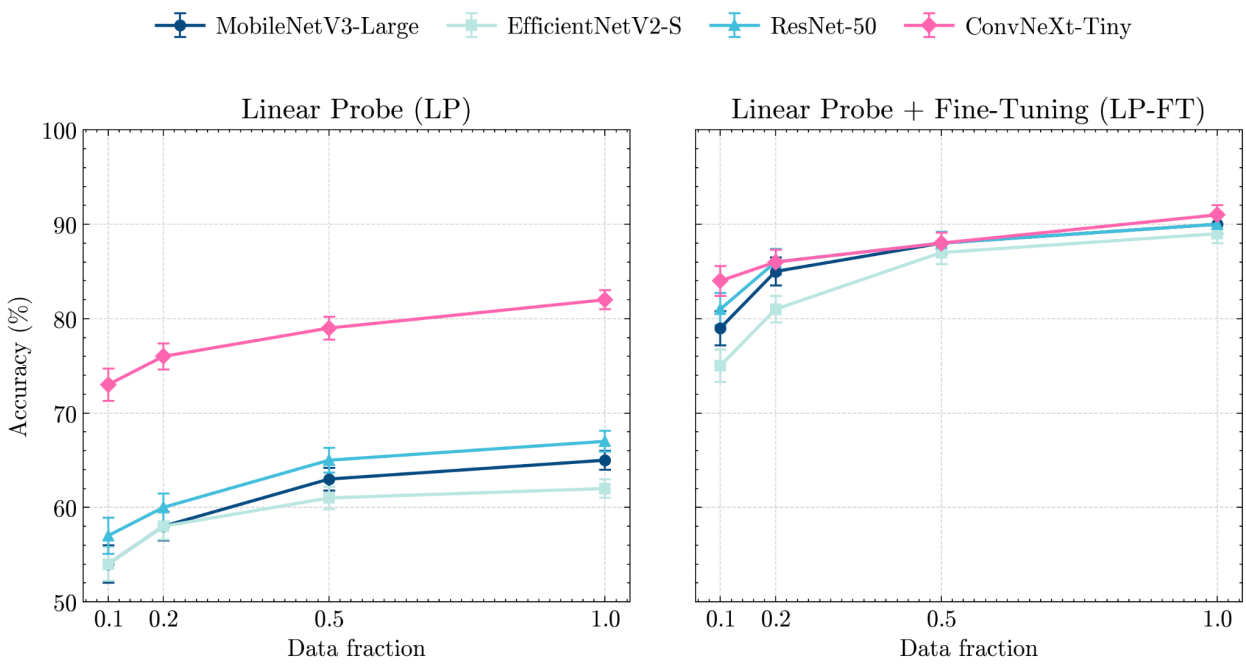


Figure 4. Accuracy results summary of tested model architectures and training strategies over four dataset fractions. Each graph shows the mean value across three seeds per LOBO fold, with error bars as ±SD across LOBO folds (k=3).

The best-performing model was analysed with the confusion matrices, given in Figure 5, and demonstrated classification accuracy across the majority of categories. Some misclassifications were observed between morphologically similar classes, particularly CTM and AT, as well as TEH and IHS. Minor confusion was also present between the "debris" and "multiple cells" categories; however, this did not affect downstream biological interpretation, as both classes were excluded from further analysis. Additionally, the implementation of the ConvNeXt-Tiny model enabled rapid inference, supporting its suitability for high-throughput applications.

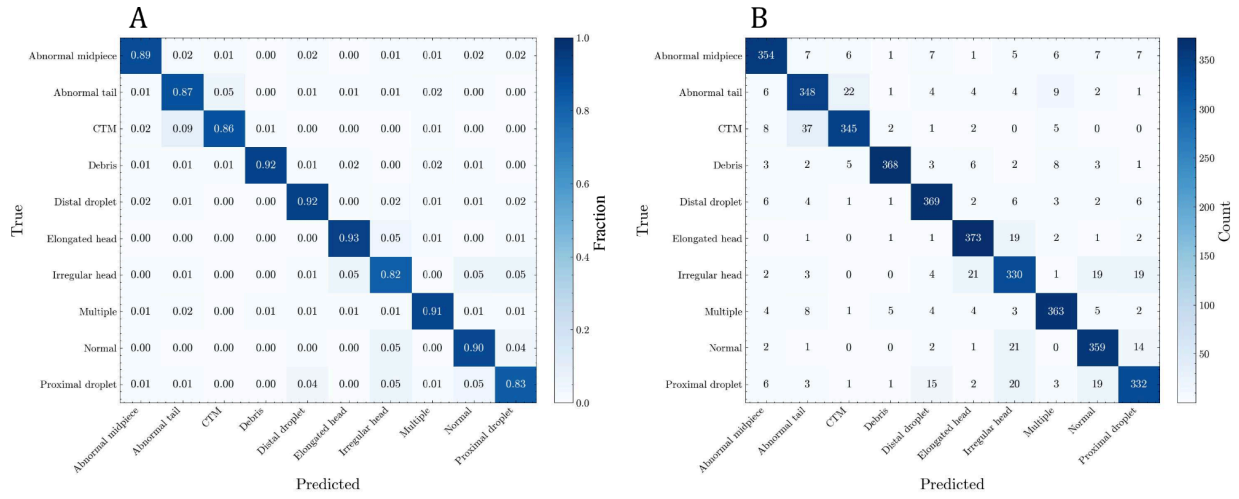


Figure 5. Confusion matrices illustration of the classification performance of the ConvNeXt-Tiny model across morphological groups. **A)** Normalized confusion matrix displaying the same results as proportions. **B)** Non-normalized confusion matrix showing absolute counts of predictions for each category, with true labels shown on vertical axis and predicted labels on horizontal axis. Correct predictions are depicted on the diagonal cells.

4.3. Generalization analysis results

To assess the generalizability of the model across different sample preparations (fresh versus frozen) and cattle breeds, as well as to evaluate potential training-related biases, a series of generalization analyses was performed.

4.3.1. Cross-condition performance

Models trained on aggregated data from a single condition achieved the highest performance when evaluated on the same condition. Accuracy reached 83.58% for the frozen→frozen setting and 76.39% for fresh→fresh (shown in Table 1). Cross-condition evaluation resulted in a decrease in performance of approximately 10-15 percentage points, with frozen→fresh dropping to 68.75% (−14.83 pp) and fresh→frozen to 66.12% (−10.27 pp). These findings suggest a pronounced domain shift between fresh and cryopreserved samples that is not fully mitigated by standard data augmentation strategies.

4.3.2. Cross-breed performance

Within each condition, models trained on a single breed consistently achieved the highest accuracy when evaluated on the same breed, with reduced performance on unseen breeds (as shown in Table 1). In the frozen subset, self-evaluation accuracies were 74.60% for Kazakh Whitehead, 83.49% for Simmental, and 75.75% for Auliekol. Cross-breed performance ranged from 59.86% (Kazakh Whitehead to Auliekol) to 81.67% (Simmental to Kazakh Whitehead), indicating relatively strong generalization of Simmental-trained models to Kazakh Whitehead, while Auliekol remained more challenging as a target for models trained on other breeds. In the fresh subset, self-evaluation accuracies reached 71.27% (Kazakh Whitehead), 85.55% (Simmental), and 75.03% (Auliekol). Cross-breed transfer was again lower, ranging from 60.08% (Simmental to Auliekol) to 68.43% (Kazakh Whitehead to Simmental). Overall,

cross-breed generalization trailed self-breed performance by approximately 5-15 percentage points, with the largest decreases observed when transferring to Auliekol. These findings suggest the presence of breed-specific morphological and/or imaging differences, along with dataset limitations, that constrain generalization from single-breed training.

Table 1. Summary of generalization analysis results.

Train	Test	Accuracy (%) [95% CI]
Frozen Condition		
Kazakh Whitehead	Kazakh Whitehead	74.60 [73.9–75.3]
	Simmental	66.80 [66.0–67.6]
	Auliekol	59.86 [59.0–60.7]
Simmental	Kazakh Whitehead	81.67 [80.9–82.4]
	Simmental	83.49 [82.8–84.2]
	Auliekol	75.76 [74.9–76.6]
Auliekol	Kazakh Whitehead	73.76 [72.9–74.6]
	Simmental	70.13 [69.2–71.0]
	Auliekol	75.75 [74.9–76.6]
Fresh Condition		
Kazakh Whitehead	Kazakh Whitehead	71.27 [70.5–72.0]
	Simmental	68.43 [67.6–69.2]
	Auliekol	67.99 [67.2–68.8]
Simmental	Kazakh Whitehead	66.51 [65.7–67.3]
	Simmental	85.55 [84.9–86.2]
	Auliekol	60.08 [59.2–60.9]
Auliekol	Kazakh Whitehead	65.09 [64.2–65.9]
	Simmental	65.12 [64.3–66.0]
	Auliekol	75.03 [74.2–75.8]
Aggregated		

Frozen	Frozen	83.58 [82.9–84.2]
	Fresh	68.75 [67.9–69.6]
Fresh	Frozen	66.12 [65.3–66.9]
	Fresh	76.39 [75.7–77.1]

Note: Accuracy given as % with 95% confidence interval (CI). The Train column demonstrates data source used for the model training stage, and data source of test data is in the Test column.

4.4. Categorization of sperm images into morphological groups using trained CNN model

Following CNN model training and cross-validation, all 401,535 images were automatically classified using the best-performing ConvNeXt-Tiny checkpoint to enable morphological profiling. The classification results are presented in the Appendix Table A5, with frozen and fresh samples reported separately. The fresh semen group was further stratified by season, including spring and fall.

Head shape abnormalities, particularly irregular head morphology, were prevalent in both frozen and fresh samples. Additional morphological defects, including distal cytoplasmic droplets (DCD), proximal cytoplasmic droplets (PCD), and abnormal midpiece (AM), were consistently observed across all bulls and groups. These results highlight the combined influence of preservation method and seasonal variation on sperm quality, emphasizing the importance of considering these factors in reproductive assessments.

4.5. Imaging flow cytometry MMP and intracellular calcium measurements

Frozen ejaculate (n=3) of a Kazakh Whitehead bull breed was taken for a final morpho-functional data acquisition using IFC-based MMP, intracellular calcium level, and morphology analysis. A total of four independent datasets were generated from three (n=3) frozen semen straws, each containing approximately 5,000 events, resulting in a total of 60,000

images across all ejaculates. After preprocessing, including gating for focused singlets and exclusion of debris and non-focused events, a total of 40,613 high-quality single-cell events were retained for analysis. Morpho-functional data were acquired from cryopreserved semen using imaging flow cytometry following dual staining with TMRE and Fluo-4. The gating strategy used to isolate biologically relevant events is presented in Figure 6.

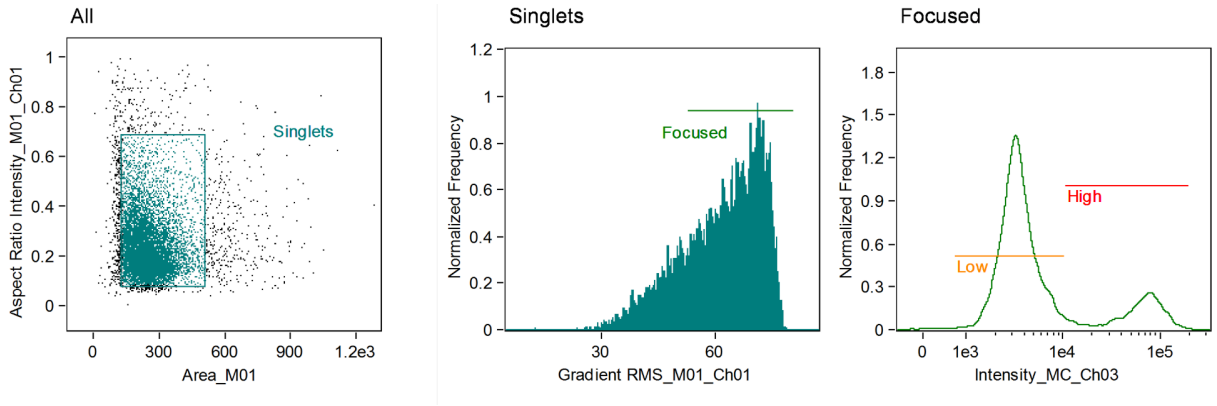


Figure 6. Gating procedure of focused single cells; within focused cells, events were represented as channel 3 (TMRE staining) intensity histogram with gating for Low and High MMP groups.

Initially, brightfield features were used to select focused single spermatozoa, excluding debris and aggregates. Fluorescence intensity in Channel 3 (TMRE) was used to define mitochondrial membrane potential (MMP) states, separating cells into high and low MMP populations based on intensity thresholds. This approach enabled stratification of spermatozoa according to mitochondrial functionality at the single-cell level.

Representative images of spermatozoa stained with Fluo-4 and TMRE and acquired by imaging flow cytometry are shown in Figure 7. These images illustrate the simultaneous detection of intracellular calcium levels and mitochondrial membrane potential within individual cells, demonstrating the capability of IFC to capture multi-parametric functional information alongside morphological features.

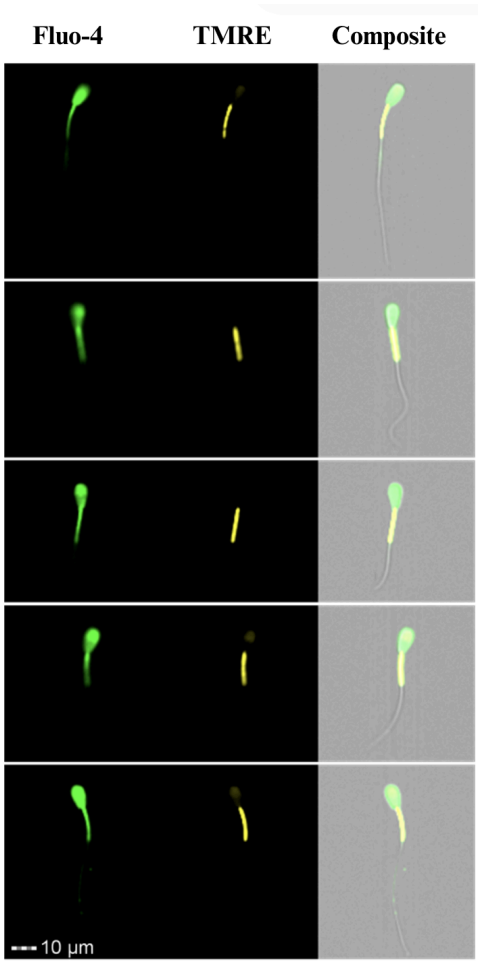


Figure 7. Images of spermatozoa stained with Fluo-4 and TMRE dyes acquired with Imagestream X Mark II (images taken with a 40x objective).

Fluo-4 fluorescence intensity (Channel 2) was used as a quantitative measure of intracellular Ca^{2+} levels. Only Fluo-4–positive events were included in further analysis, with positivity rates of 93.23% on average across all datasets (given in Appendices Table A6), confirming efficient staining and data quality. The distribution of high and low MMP events across morphological groups is summarized in the Appendices Table A6 as well. Normal spermatozoa were predominantly characterized by low MMP (79-85% range across bulls),

whereas abnormal midpiece (AM), coiled tail (CTM), and abnormal tail (AT) groups exhibited higher proportions of high MMP cells, indicating altered mitochondrial function associated with structural abnormalities.

4.6. Analysis of functional parameters in morphological groups

Morpho-functional integration was performed by combining CNN-based morphological classification with fluorescence-derived functional measurements at the single-cell level. The final dataset consisted of 40,613 spermatozoa categorized into ten morphological groups and annotated with MMP status and Fluo-4 intensity values as provided in Table 2.

Table 2. Morpho-functional analysis of mitochondrial membrane potential (MMP) and intracellular Ca^{2+} across sperm morphological groups for n=3 frozen bull ejaculates.

Groups	Bull 1					Bull 2					Bull 3				
	n	**Mean (%) ± SD**	**High MMP mean (%) ± SD**	**Low MMP mean (%) ± SD**	**Fluo-4 intensity median [Q1-Q3]**	**n**	**Mean (%) ± SD**	**High MMP mean (%) ± SD**	**Low MMP mean (%) ± SD**	**Fluo-4 intensity median [Q1-Q3]**	**n**	**Mean (%) ± SD**	**High MMP mean (%) ± SD**	**Low MMP mean (%) ± SD**	**Fluo-4 intensity median [Q1-Q3]**
Normal	4881	34.13 ± 0.60	9.14 ± 0.38	84.88 ± 0.89	20,465 [15425 – 27617]	3506	27.025 ± 0.77	16.90 ± 0.73	82.90 ± 0.69	14175 [10565 – 20017]	3600	27 ± 1.76	19.68 ± 2.70	78.71 ± 3.61	11195 [8347 – 17226]
AT	1095	7.65 ± 0.66	54.58 ±3.08	43.00 ± 3.01	20756 [15335 – 28967]	1015	7.8 ± 0.39	34.15 ± 3.28	65.47 ± 3.46	15502 [11510 – 25566]	611	4.6 ± 0.22	37.45 ± 2.80	60.58 ± 3.63	13916 [9292 – 19534]
AM	484	3.38 ± 0.16	70.23 ± 7.67	27.50 ± 8.08	25153 [17328 – 42323]	513	3.95 ± 0.39	47.48 ± 3.71	51.42 ± 4.06	24842 [13488 – 39335]	326	2.45 ± 0.31	48.47 ± 4.86	49.83 ± 5.06	16373 [9829 – 25507]
CTM	224	1.56 ± 0.31	63.42 ± 5.40	34.59 ± 3.61	20372 [15851 – 25385]	203	1.55 ± 0.13	28.92 ± 2.33	69.61 ± 2.56	14548 [10752 – 22402]	154	1.175 ± 0.10	57.96 ± 4.86	40.70 ± 6.10	14958 [11501 – 20343]
TEH	580	4.06 ± 0.34	10.64 ± 3.61	78.71 ± 4.99	18643 [14145 – 25750]	493	3.825 ± 0.74	17.69 ± 1.87	82.08 ± 1.62	14110 [10163 – 19634]	654	4.925 ± 0.30	16.55 ± 1.90	81.79 ± 1.48	10460 [8073 – 15628]
IHS	998	6.99 ± 0.87	10.83 ± 2.87	79.88 ± 2.23	19615 [15236 – 26369]	769	5.925 ± 0.38	14.77 ± 2.04	84.70 ± 2.30	13093 [9311 – 17958]	903	6.8 ± 0.59	17.16 ± 3.66	82.05 ± 3.50	10995 [8348 – 16568]
PCD	38	0.27 ± 0.06	10.52 ±7.37	81.05 ± 12.90	18193 [14412 – 24234]	30	0.225 ± 0.05	24.16 ± 9.00	75.85 ± 9.00	16103 [11028 – 26513]	26	0.2 ± 0.08	21.35 ± 12.97	72.30 ± 8.97	12383 [8219 – 14719]

DCD	309	2.16 ± 0.19	19.25 ± 4.50	76.01 ± 6.71	19566 [14988 – 25833]	334	2.575 ± 0.39	25.31 ± 4.80	74.17 ± 4.44	14950 [10330 – 22333]	372	2.775 ± 0.22	25.71 ± 4.60	73.18 ± 3.74	11883 [8668 – 17368]
Debris	5608	39.20 ± 0.73	21.95 ±1.73	70.08 ± 2.63	20095 [15047 – 29202]	6051	46.6 ± 1.73	33.48 ± 2.23	64.93 ± 2.33	15649 [10942 – 24894]	6644	49.875 ± 0.88	38.10 ± 3.13	59.62 ± 3.35	11883 [8508 – 18337]
Multiple	87	0.61 ± 0.22	34.52 ±14.05	62.63 ± 12.45	38408 [24614 – 51706]	71	0.55 ± 0.13	46.32 ± 9.64	53.68 ± 9.64	26461 [16403 – 36236]	34	0.225 ± 0.10	65.91 ± 12.56	34.09 ± 12.56	17963 [12224 – 23611]
Total	**14304**					**12985**					**13324**				

Morphological composition was consistent across bulls, with debris being the most abundant group (ranging from 39% to 50%), followed by normal spermatozoa (27-34%). IHS was consistently the third most represented abnormal morphology across all animals.

Analysis of mitochondrial membrane potential revealed that normal spermatozoa were predominantly characterized by low MMP (79-85%) across all three bulls. In contrast, groups associated with structural abnormalities consistently showed elevated proportions of high MMP cells, with AM, CTM, and AT exhibiting the highest proportions across all animals, suggesting that structural defects in the midpiece and tail are associated with altered mitochondrial function.

Intracellular calcium levels, assessed by Fluo-4 intensity, showed consistent patterns across all three bulls. The Multiple and AM groups exhibited high calcium levels in all animals, while TEH and IHS groups consistently showed the lowest values, indicating that functional heterogeneity in intracellular calcium signaling is reproducibly linked to sperm morphology across individuals.

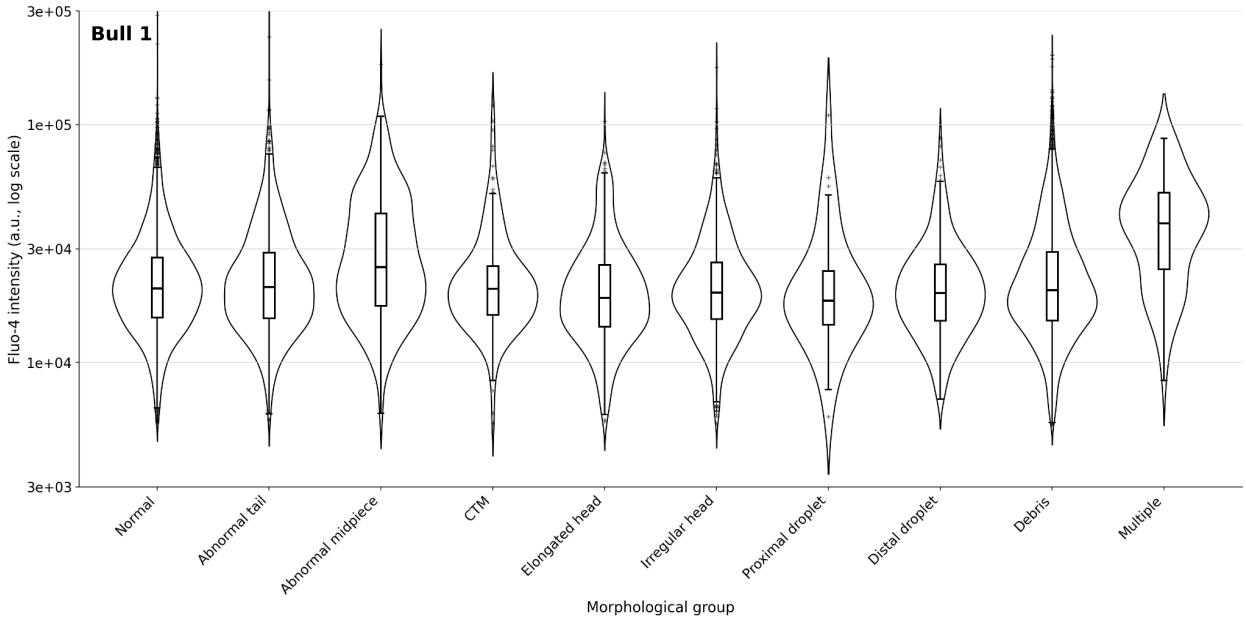
Statistical analysis using the Kruskal–Wallis test demonstrated significant differences in Fluo-4 fluorescence intensity among morphological groups in all three bulls, indicating that within the analyzed dataset, intracellular calcium levels were not uniformly distributed across morphological categories. The results are given in Table 3.

Table 3. Kruskal-Wallis test results for Fluo-4 intensity across morphological groups in three bulls.

	Bull 1	**Bull 2**	**Bull 3**
Total cells	14304	12985	13324
Fluo-4 positive, %	98.2	90.35	91.14
H (χ^2)	186.72	308.74	166.46

df	9	9	9
p-value	2.00×10^{-35}	3.66×10^{-61}	3.37×10^{-31}

The distribution of intracellular calcium levels across sperm morphological groups and for each bull is shown in Figure 8. Fluo-4 intensity values were visualized on a logarithmic scale to account for the skewed distribution of fluorescence signals. Clear differences in calcium levels were observed between morphological categories, with the “Multiple” and abnormal midpiece groups exhibiting higher median and overall distributions compared to other groups. In contrast, head defect categories such as elongated and irregular head showed comparatively lower intensity ranges. These visual patterns are consistent with the statistical analysis.



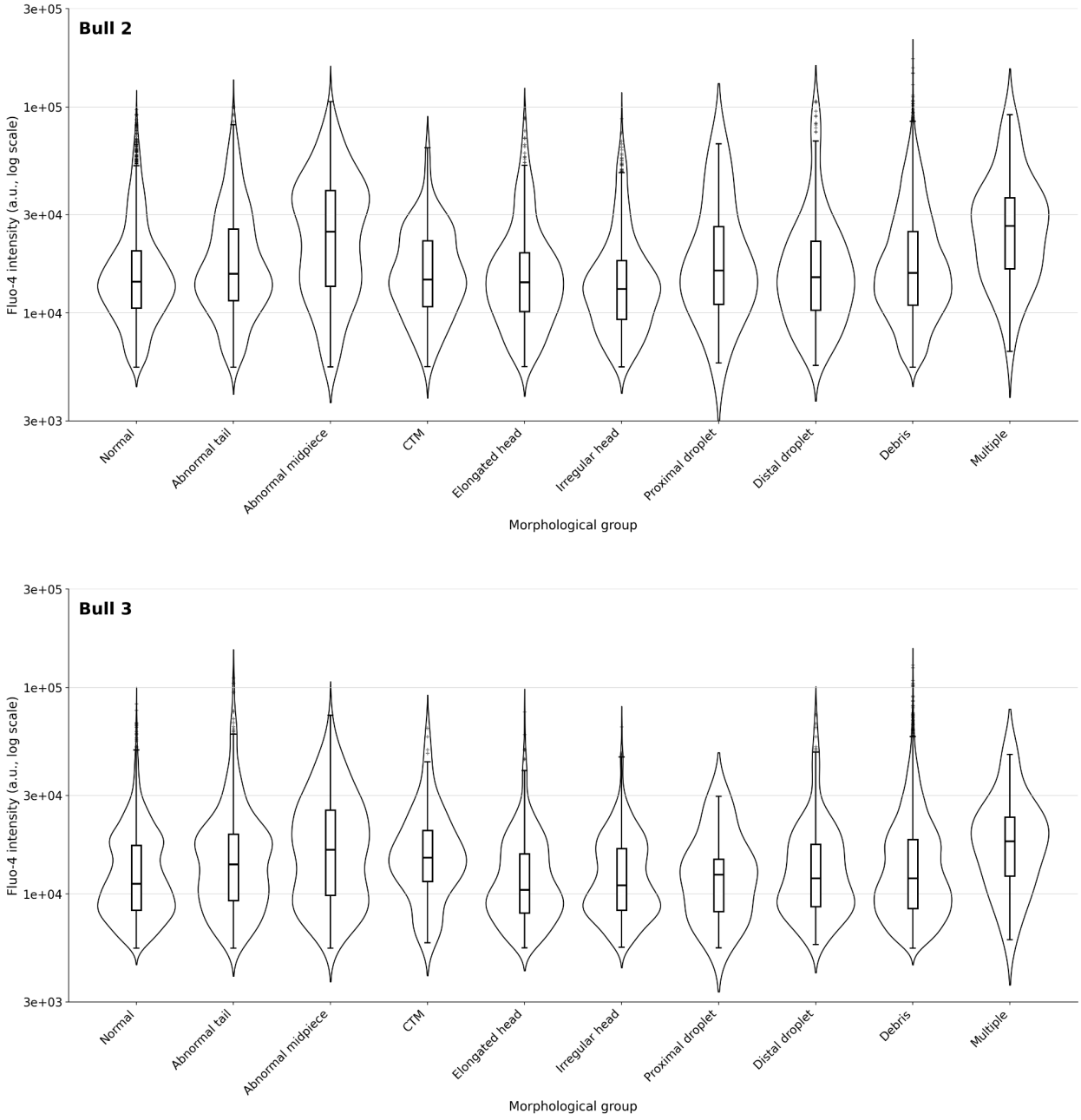


Figure 8. Distribution of Fluo-4 intensity across sperm morphological groups and bulls.

5. DISCUSSION

Conventional semen evaluation for assisted reproductive technologies relies on light microscopy and CASA systems, which, despite their utility in motility assessment, remain limited in evaluating morphological abnormalities and viability (Hidayatullah et al., 2021). These approaches are limited by low throughput, reliance on expert interpretation, low image resolution, and potential sampling bias due to the small number of analyzed cells (Javadi & Mirroshandel, 2019). In contrast, imaging flow cytometry enables high-throughput acquisition of single-cell images, allowing more robust statistical assessment of sperm morphology (Buckman et al., 2009; Barteneva et al., 2012; Bulkeley et al., 2023; Barth et al., 2025; Teves et al., 2025).

Morphological abnormalities in spermatozoa arise from multiple factors, including oxidative stress, mechanical damage, and incomplete maturation, and are closely associated with key functional parameters such as membrane integrity, mitochondrial activity, and DNA quality (Domínguez-Rebolledo et al., 2010; Gliozzi et al., 2017; Hezavehei et al., 2018; Grötter et al., 2019; Peris-Frau et al., 2019; Palazzese et al., 2020; Ferreux et al., 2021; Hai et al., 2024; Kumar et al., 2024; Pelzman et al., 2024; Zhang et al., 2025; Garcia-Oliveros et al., 2025). Accurate classification of these abnormalities is therefore essential for assessing fertilization potential and identifying underlying physiological alterations.

Deep learning approaches, particularly convolutional neural networks, provide an effective framework for automated and scalable classification of sperm morphology (Akinsulie et al., 2024; Burti et al., 2024). While traditional image analysis tools such as IDEAS software enable feature extraction and rule-based classification, CNNs allow direct learning from raw image data, improving consistency and reducing dependence on manual feature engineering. Previous studies have applied machine learning and computer vision techniques to sperm

analysis, including segmentation-based methods and CNN-based classification models (Matamoros-Volante et al., 2018; Javadi & Mirroshandel, 2019; Fraczek et al., 2022; Iqbal et al., 2020; Aktas et al., 2024; Maalej et al., 2025). However, many of these approaches have focused primarily on human spermatozoa and were often limited by small dataset sizes (Movahed et al., 2019; Marín & Chang, 2021; Hernández-Herrera et al., 2024). More recent work has demonstrated the feasibility of IFC-based classification in animal models, such as boar semen, though with a restricted number of morphological categories and manually annotated datasets (Keller & Kerns, 2022; Keller et al., 2025).

The present study extends these approaches by utilizing a large-scale dataset of 401,535 high-resolution IFC images of bovine spermatozoa across multiple breeds and conditions. Among the evaluated architectures, ConvNeXt-Tiny demonstrated the most consistent performance, achieving an accuracy of 91.1% on a mixed dataset. The model was capable of distinguishing multiple morphological classes, although some degree of confusion remained between visually similar categories, such as TEH and IHS or CTM and AT.

Analysis of model generalization revealed that performance decreased when training data were restricted to a single breed or individual, indicating sensitivity to dataset diversity. These findings highlight the importance of sufficient dataset size and variability for distinguishing true biological differences from artifacts related to sampling or model bias. The observed domain shifts emphasize the need for balanced and representative datasets when developing AI-based tools for biological image analysis.

The distribution of morphological classes showed that approximately 58% of spermatozoa in frozen samples exhibited normal morphology, compared to 38% and 56% in fresh samples collected during spring and fall, respectively. A reduction in the proportion of

morphologically abnormal cells was observed following cryopreservation, consistent with previous findings (Yurchuk et al., 2021; Milewski et al., 2024). Therefore, there is the influence of sample processing on sperm morphology that reinforces the importance of evaluating model performance across experimental conditions.

A key advantage of this approach lies in the integration of IFC with deep learning, enabling analysis of tens of thousands of individual cells per sample. This substantially increases statistical power compared to conventional microscopy, where only 100-200 cells are typically assessed per ejaculate. In addition, automated classification reduces observer bias and ensures consistent evaluation across datasets. Despite these strengths, several limitations remain. Certain morphological categories were underrepresented due to limited sample availability, particularly as only one ejaculate per bull was analyzed under each condition. Furthermore, label-free analysis may be insufficient for accurately identifying asymmetric or out-of-focus cells, which may require additional validation through staining or manual inspection (Lafouresse et al., 2021). As shown in Figure 9, some imaging artifacts, including angular acquisition of cells, can affect classification accuracy and may necessitate further methodological refinement.

Future work should focus on expanding dataset diversity, improving representation of rare morphological classes, and integrating additional data modalities, including fluorescence-based functional parameters. Such developments may enhance both the accuracy and biological interpretability of AI-driven sperm analysis and support its application in livestock reproductive technologies.

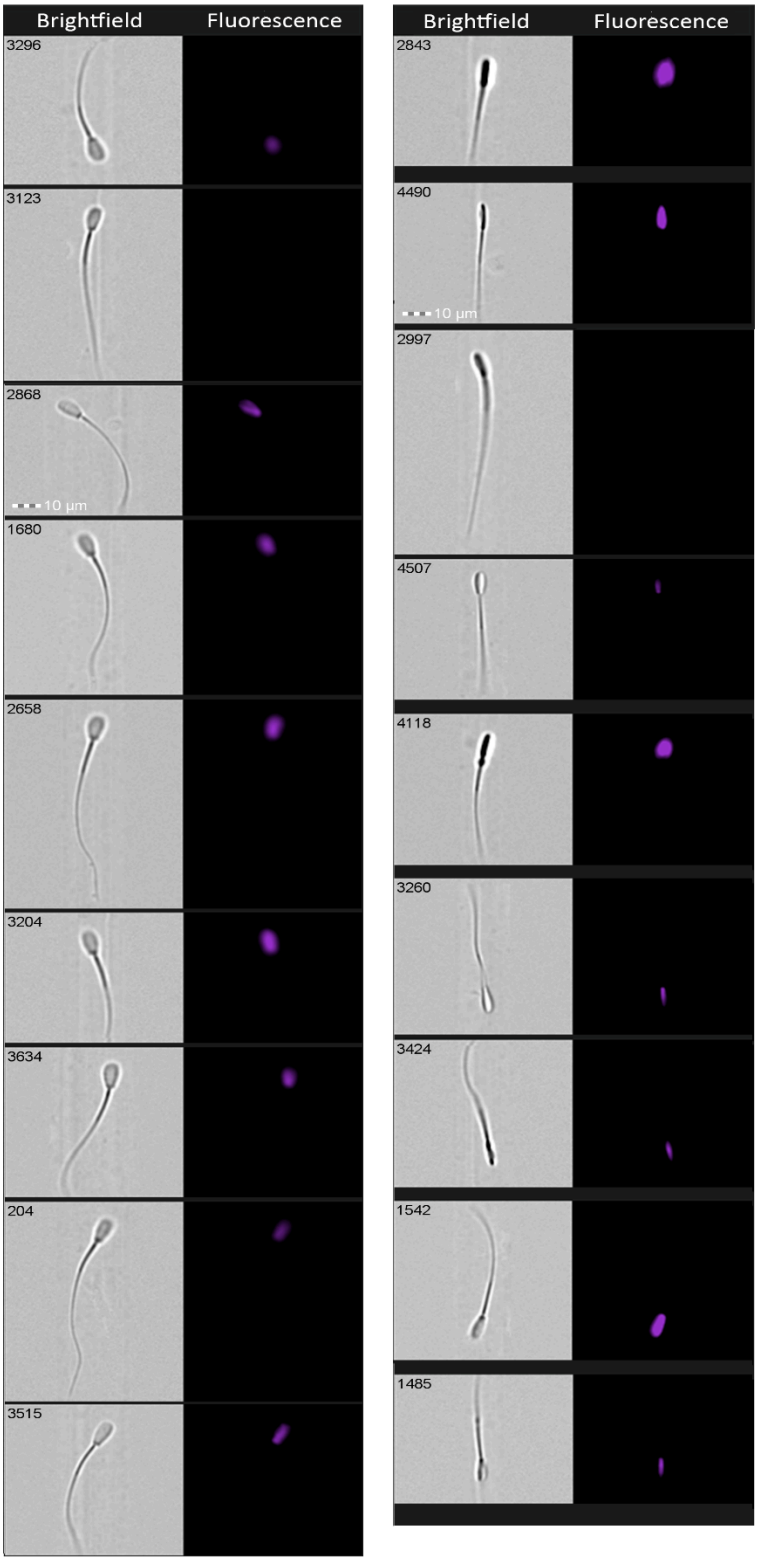

Brightfield
Fluorescence
3296
3123
2868
10 µm
1680
2658
3204
3634
204
3515
Brightfield
Fluorescence
2843
4490
10 µm
2997
4507
4118
3260
3424
1542
1485

Figure 9. Images of spermatozoa stained with DNA dye (Hoechst 33342) and acquired with Imagestream X Mark II (40x magnification).

A key contribution of this study is the integration of morphological classification with functional parameters at the single-cell level using a high-throughput pipeline. The observed differences in intracellular calcium levels across morphological groups, supported by highly significant statistical results, indicate that structural abnormalities are not only morphological descriptors but represent functionally relevant phenotypes. In particular, the abnormal midpiece (AM) group consistently showed elevated mitochondrial activity (high MMP) alongside significantly altered calcium levels, suggesting disruption of mitochondrial-calcium coupling mechanisms. This observation is supported by previous studies demonstrating that structural defects in the sperm midpiece are closely associated with impaired mitochondrial organization and function, which in turn affects sperm physiology and fertility potential (Sutovsky, 2015; Skowronek et al., 2025; Díaz‑Miranda et al., 2020). Midpiece abnormalities have been linked to alterations in mitochondrial morphology and bioenergetics, reinforcing the idea that this region plays a central role in regulating sperm function. Even in cases where motility parameters appear preserved, a high prevalence of midpiece defects may still reflect underlying functional dysregulation, particularly at the level of mitochondrial activity (Maggi et al., 2024).

The abnormal midpiece (AM) group exhibited the highest calcium intensities across all three bulls, which may reflect incomplete maturation or altered ionic homeostasis. Distal cytoplasmic droplet (DCD) cells also showed elevated intracellular calcium levels compared to morphologically normal spermatozoa, consistent with evidence indicating that retention of cytoplasmic droplets is associated with immature spermatozoa that have not fully acquired volume regulation and ion exchange mechanisms during epididymal maturation (Cooper &

Yeung, 2003). As a result, such cells may exhibit dysregulated calcium handling, supporting the elevated intracellular calcium levels observed in this group, however still possessing fertilizing potential.

Finally, the ability of the ConvNeXt-Tiny model to reliably classify sperm morphology enabled robust downstream statistical analysis on large datasets, overcoming limitations of manual annotation and allowing for the identification of subtle but biologically meaningful morpho-functional relationships at the single-cell level.

6. CONCLUSION

This study establishes a high-throughput, AI-assisted pipeline for single-cell morpho-functional analysis of bovine spermatozoa by integrating imaging flow cytometry with deep learning. A large-scale dataset of over 400,000 images was used to train a robust classification model, and a subset of approximately 60,000 images from cryopreserved semen enabled detailed functional analysis of 40,613 individual cells. The results demonstrate that morphological abnormalities are strongly associated with alterations in mitochondrial membrane potential and intracellular calcium levels. Significant differences in functional parameters across morphological groups confirm that morphology alone is insufficient to describe sperm quality and must be complemented with functional assessment.

This work is the first to our knowledge to represent the large-scale dataset and single-cell morpho-functional analysis of spermatozoa from Kazakhstani cattle breeds. The developed pipeline provides a reproducible framework for multi-parametric sperm evaluation and highlights the potential of AI-driven approaches for improving fertility assessment in livestock reproduction. Further directions include improvement of a dataset size and increasing representability of some rare classes for morphology assessment. Additional validation strategies via other methods for both functional and morphological assays are required before application of this approach in ART practices.

7. REFERENCES

Ahmad, M., Nasrullah, R., Riaz, H., Sattar, A., & Ahmad, N. (2014). Changes in motility, morphology, plasma membrane and acrosome integrity during stages of cryopreservation of buck sperm. Journal of the South African Veterinary Association, 85(1), 972. https://doi.org/10.4102/jsava.v85i1.972

Aitken, R. J., Clarkson, J. S., & Fishel, S. (1989). Generation of reactive oxygen species, lipid peroxidation, and human sperm function. Biology of reproduction, 41(1), 183-197. https://doi.org/10.1095/biolreprod41.1.183

Akinsulie, O. C., Idris, I., Aliyu, V. A., Shahzad, S., Banwo, O. G., Ogunleye, S. C., Olorunshola, M., Okedoyin, D. O., Ugwu, C., Oladapo, I. P., Gbadegoye, J. O., Akande, Q. A., Babawale, P., Rostami, S., & Soetan, K. O. (2024). The potential application of artificial intelligence in veterinary clinical practice and biomedical research. Frontiers in veterinary science, 11, 1347550. https://doi.org/10.3389/fvets.2024.1347550

Aktas, A., Serbes, G., Yigit, M. H., Aydin, N., Uzun, H., & Ilhan, H.O. (2024). Hi-LabSpermMorpho: a novel expert-labeled dataset with extensive abnormality classes for deep learning-based sperm morphology analysis. IEEE Access, 12, 196070-196091. https://doi.org/10.1109/ACCESS.2024.3521643

Antonouli, S., Di Nisio, V., Messini, C., Samara, M., Salumets, A., Daponte, A., & Anifandis, G. (2024). Sperm plasma membrane ion transporters and male fertility potential: a perspective under the prism of cryopreservation. Cryobiology, 114, 104845. https://doi.org/10.1016/j.cryobiol.2023.104845

Bailey, J., Morrier, A., & Cormier, N. (2003). Semen cryopreservation: successes and persistent problems in farm species. Canadian Journal of Animal Science, 83, 393-401.

https://doi.org/10.4141/A03-02

Barth, A., Perry, V.E.A., Hamilton, L.E., Sutovsky, P., & Oko, R. (2025). Assessing bovine male fertility in a technological age. Advances in Anatomy, Embryology and Cell Biology, 240, 297-329. https://doi.org/10.1007/978-3-031-70126-9_7

Barteneva, N.S., Fasler-Kan, E., & Vorobjev, I.A., (2012). Imaging flow cytometry: coping with heterogeneity in biological systems. Journal of Histochemistry & Cytochemistry, 60(10), 723-733. https://doi.org/10.1369/0022155412453052

Barteneva, N. S., & Vorobjev, I. A. (2018). Heterogeneity of metazoan cells and beyond: to integrative analysis of cellular populations at single-cell level. Cellular Heterogeneity: Methods and Protocols, 1745, 3-23. https://doi.org/10.1007/978-1-4939-7680-5_1

Basiji, D. A. (2016). Principles of Amnis imaging flow cytometry. In N. S. Barteneva & I. A. Vorobjev (Eds.), *Imaging flow cytometry: Methods and protocols* (Methods in Molecular Biology, Vol. 1389, pp. 13–21). Springer-Humana Press. https://doi.org/10.1007/978-1-4939-3302-0_2

Bogle, O. A., Kumar, K., Attardo-Parrinello, C., Lewis, S. E., Estanyol, J. M., Ballescà, J. L., & Oliva, R. (2017). Identification of protein changes in human spermatozoa throughout the cryopreservation process. Andrology, 5(1), 10–22. https://doi.org/10.1111/andr.12279

Bollwein, H., & Malama, E. (2023). Review: evaluation of bull fertility. Functional and molecular approaches. Animal : an international journal of animal bioscience, 17 Suppl 1, 100795. https://doi.org/10.1016/j.animal.2023.100795

Buckman, C., George, T. C., Friend, S., Sutovsky, M., Miranda-Vizuete, A., Ozanon, C., Morrissey, P., & Sutovsky, P. (2009). High throughput, parallel imaging and biomarker quantification of human spermatozoa by ImageStream flow cytometry. Systems biology in

reproductive medicine, 55(5-6), 244–251. https://doi.org/10.3109/19396360903056224

Bulkeley, E., Santistevan, A. C., Varner, D., & Meyers, S. (2023). Imaging flow cytometry to characterize the relationship between abnormal sperm morphologies and reactive oxygen species in stallion sperm. Reproduction in domestic animals, 58(1), 10–19. https://doi.org/10.1111/rda.14241

Burti, S., Banzato, T., Coghlan, S., Wodziniski, M., Bendazzoli, M., & Zotti, A. (2024). Artificial intelligence in veterinary diagnostic imaging: Perspectives and limitations. Research in Veterinary Science, 175, 105317. https://doi.org/10.1016/j.rvsc.2024.105317

Chatterjee, S., de Lamirande, E., & Gagnon, C. (2001). Cryopreservation alters membrane sulfhydryl status of bull spermatozoa: protection by oxidized glutathione. Molecular Reproduction and Development, 60(4), 498-506. https://doi.org/10.1002/mrd.1115

Cherouveim, P., Velmahos, C., & Bormann, C. L. (2023). Artificial intelligence for sperm selection-a systematic review. Fertility and sterility, 120(1), 24–31. https://doi.org/10.1016/j.fertnstert.2023.05.157

Cooper, T. G., & Yeung, C. H. (2003). Acquisition of volume regulatory response of sperm upon maturation in the epididymis and the role of the cytoplasmic droplet. Microscopy research and technique, 61(1), 28–38. https://doi.org/10.1002/jemt.10314

Dai, C., Zhang, Z., Shan, G., Chu, L. T., Huang, Z., Moskovtsev, S., Librach, C., Jarvi, K., & Sun, Y. (2021). Advances in sperm analysis: techniques, discoveries and applications. Nature reviews. Urology, 18(8), 447–467. https://doi.org/10.1038/s41585-021-00472-2

de Lima Rosa, J., de Paula Freitas Dell'Aqua, C., de Souza, F. F., Missassi, G., & Kempinas, W. G. (2023). Multiple flow cytometry analysis for assessing human sperm functional characteristics. Reproductive toxicology (Elmsford, N.Y.), 117, 108353.

https://doi.org/10.1016/j.reprotox.2023.108353

Díaz-Miranda, E. A., Maitan, P. P., Machado, T. P., Camilo, B. S., Lima, D. A., Okano, D. S., Penitente-Filho, J. M., Machado-Neves, M., de Oliveira, L. L., Guimarães, S. E. F., da Costa, E. P., & Guimarães, J. D. (2020). Disruption of bovine sperm functions in the presence of aplastic midpiece defect. Andrology, 8(1), 201–210. https://doi.org/10.1111/andr.12618

Domínguez-Rebolledo, A. E., Fernández-Santos, M. R., Bisbal, A., Ros-Santaella, J. L., Ramón, M., Carmona, M., Martínez-Pastor, F., & Garde, J. J. (2010). Improving the effect of incubation and oxidative stress on thawed spermatozoa from red deer by using different antioxidant treatments. Reproduction, fertility, and development, 22(5), 856–870. https://doi.org/10.1071/RD09197

Douglas, C., Parekh, N., Kahn, L. G., Henkel, R., & Agarwal, A. (2021). A novel approach to improving the reliability of manual semen analysis: a paradigm shift in the workup of infertile men. The world journal of men's health, 39(2), 172–185. https://doi.org/10.5534/wjmh.190088

Ferreux, L., Bourdon, M., Chargui, A., Schmitt, A., Stouvenel, L., Lorès, P., Ray, P., Lousqui, J., Pocate-Cheriet, K., Santulli, P., Dulioust, E., Toure, A., & Patrat, C. (2021). Genetic diagnosis, sperm phenotype and ICSI outcome in case of severe asthenozoospermia with multiple morphological abnormalities of the flagellum. Human reproduction (Oxford, England), 36(11), 2848–2860. https://doi.org/10.1093/humrep/deab200

Fraczek, A., Karwowska, G., Miler, M., Lis, J., Jezierska, A., & Mazur-Milecka, M. (2022). Sperm segmentation and abnormalities detection during the ICSI procedure using machine learning algorithms. In 2022 15th International Conference on Human System Interaction (HSI) (pp. 1-6). IEEE.

Gadea, J., Gumbao, D., Cánovas, S., García-Vázquez, F. A., Grullón, L. A., & Gardón, J. C.

(2008). Supplementation of the dilution medium after thawing with reduced glutathione improves function and the in vitro fertilizing ability of frozen-thawed bull spermatozoa. International journal of andrology, 31(1), 40–49. https://doi.org/10.1111/j.1365-2605.2007.00756.x

Garcia-Oliveros, L. N., Bozzi, A. D. R., Cardoso Silva, T. O., Baatsch Nascimento, F., Oliveira, T. R., Miguel, E. L., Alves, M. B. R., Rangel, R. B., Andrade, A. F. C., Arruda, R. P., Kerns, K., & Celeghini, E. C. C. (2025). Phosphatidylserine translocation, cholesterol spatial distribution, and acrosome reaction reliably distinguish sperm capacitation from cryoinjury in bovine sperm. Cryobiology, 121, 105302. https://doi.org/10.1016/j.cryobiol.2025.105302

Ghayda, R. A., Cannarella, R., Calogero, A. E., Shah, R., Rambhatla, A., Zohdy, W., Kavoussi, P., Avidor-Reiss, T., Boitrelle, F., Mostafa, T., Saleh, R., Toprak, T., Birowo, P., Salvio, G., Calik, G., Kuroda, S., Kaiyal, R. S., Ziouziou, I., Crafa, A., Phuoc, N. H. V., … Global Andrology Forum (2024). Artificial Intelligence in Andrology: From Semen Analysis to Image Diagnostics. The world journal of men's health, 42(1), 39–61. https://doi.org/10.5534/wjmh.230050

Gliozzi, T. M., Turri, F., Manes, S., Cassinelli, C., & Pizzi, F. (2017). The combination of kinetic and flow cytometric semen parameters as a tool to predict fertility in cryopreserved bull semen. Animal, 11(11), 1975-1982. https://doi.org/10.1017/S1751731117000684

Gloria, A., Bracco, C., Antenucci, E., & Contri, A. (2023). Validation of the volumetric flow cytometry for bovine sperm concentration. Scientific Reports, 13, 16560. https://doi.org/10.1038/s41598-023-43678-7

Grötter, L. G., Cattaneo, L., Marini, P. E., Kjelland, M. E., & Ferré, L. B. (2019). Recent advances in bovine sperm cryopreservation techniques with a focus on sperm post-thaw quality optimization. Reproduction in domestic animals 54(4), 655–665. https://doi.org/10.1111/rda.13409

Guahmich, N. L., Borini, E., & Zaninovic, N. (2023). Improving outcomes of assisted reproductive technologies using artificial intelligence for sperm selection. Fertility and sterility, 120(4), 729–734. https://doi.org/10.1016/j.fertnstert.2023.06.009

Gurupriya, V. S., Divyashree, B. C., & Roy, S. C. (2014). Cryogenic changes in proteases and antiprotease activities of buffalo (Bubalus bubalis) and cattle (Bos taurus) semen. Theriogenology, 81(3), 396–402. https://doi.org/10.1016/j.theriogenology.2013.10.010

Hai, E., Li, B., Zhang, J., & Zhang, J. (2024). Sperm freezing damage: the role of regulated cell death. Cell Death Discovery, 10(1), 239. https://doi.org/10.1038/s41420-024-02013-3

Hai, E., Li, B., Song, Y., Zhang, J., & Zhang, J. (2025). Ferroptosis emerges as the predominant form of regulated cell death in goat sperm cryopreservation. Journal of Animal Science and Biotechnology, 16(1), 26. https://doi.org/10.1186/s40104-025-01158-0

Hanson, R., Reddick, S., Thuerauf, S., Webb, K., Kasimanickam, V., & Kasimanickam, R. (2023). Comparison of bull sperm morphology evaluation methods under field conditions. Clinical Theriogenology, 15. https://doi.org/10.58292/ct.v15.9425

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

Hernández-Herrera, P., Abonza, V., Sanchez-Contreras, J., Darszon, A., & Guerrero, A. (2024). Deep learning-based classification and segmentation of sperm head and flagellum for image-based flow cytometry. Computación y Sistemas, 27(4). https://doi.org/10.13053/cys-27-4-4772

Hezavehei, M., Sharafi, M., Kouchesfahani, H. M., Henkel, R., Agarwal, A., Esmaeili, V., & Shahverdi, A. (2018). Sperm cryopreservation: a review on current molecular cryobiology and

advanced approaches. Reproductive biomedicine online, 37(3), 327–339. https://doi.org/10.1016/j.rbmo.2018.05.012

Hidayatullah, P., Mengko, T. L. E. R., Munir, R., & Barlian, A. (2021). Bull sperm tracking and machine learning-based motility classification. IEEE Access, 9, 61159-61170. https://doi.org/10.1109/access.2021.3074127

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 1314-1324). https://doi.org/10.48550/arXiv.1905.02244

Iqbal, I., Mustafa, G., & Ma, J. (2020). Deep learning-based morphological classification of human sperm heads. Diagnostics, 10(5), 325. https://doi.org/10.3390/diagnostics10050325

Javadi, S., & Mirroshandel, S. A. (2019). A novel deep learning method for automatic assessment of human sperm images. Computers in biology and medicine, 109, 182–194. https://doi.org/10.1016/j.compbiomed.2019.04.030

Kadlec, M., Ros-Santaella, J. L., & Pintus, E. (2020). The Roles of NO and H_2S in Sperm Biology: Recent Advances and New Perspectives. International journal of molecular sciences, 21(6), 2174. https://doi.org/10.3390/ijms21062174

Keller, A., & Kerns, K. (2022). Deep learning, artificial intelligence methods to predict boar sperm acrosome health. Animal Reproduction Science, 247, 107110. https://doi.org/10.1016/j.anireprosci.2022.107110

Keller, A., Maus, M., Keller, E., & Kerns, K. (2025). Deep learning classification method for boar sperm morphology analysis. Andrology, 13(6), 1615–1625. https://doi.org/10.1111/andr.13758

Khalil, W. A., El-Harairy, M. A., Zeidan, A. E. B., Hassan, M. A. E., & Mohey-Elsaeed, O. (2017). Evaluation of bull spermatozoa during and after cryopreservation: Structural and ultrastructural insights. International journal of veterinary science and medicine, 6(Suppl), S49–S56. https://doi.org/10.1016/j.ijvsm.2017.11.001

Kumar, A., Raghunathan, A., Jones, R., Ma, T., & Liang, P. (2022). Fine-tuning can distort pretrained features and underperform out-of-distribution. arXiv preprint, 2202.10054. https://doi.org/10.48550/arXiv.2202.10054

Kumar, A., Saxena, A., & Anand, M. (2024). Subtle membrane changes in cryopreserved bull spermatozoa when modified temperature drop rates are used during the first phase of freezing. Cryoletters, 45(4), 212-220. https://doi.org/10.54680/fr24410110312

Kumar, A., Prasad, J. K., Kumar, N., Verma, S., Anand, M., Behera, A., Kumar, A., Tiwari, J., Dhariya, R., Singhal, S., & Yadav, V. (2025). Quercetin modulates transcription of the apoptotic and CatSper genes and optimises post thaw viability and kinematics of buck spermatozoa. Scientific Reports, 15, 22818. https://doi.org/10.1038/s41598-025-88525-z

Kurenkov, A., Kussanova, A., & Barteneva, N. S. (2024). Advancing precision single-cell analysis of red blood cells through semi-supervised deep learning. In Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XXII (Vol. 12846, p. 1284602). SPIE. https://doi.org/10.1117/12.3008410

Lafouresse, F., Jugele, R., Müller, S., Doineau, M., Duplan-Eche, V., Espinosa, E., Puisségur, M. P., Gadat, S., & Valitutti, S. (2021). Stochastic asymmetric repartition of lytic machinery in dividing $CD8^+$ T cells generates heterogeneous killing behavior. eLife, 10, e62691. https://doi.org/10.7554/eLife.62691

Lewis, S. E., Donnelly, E. T., Sterling, E. S., Kennedy, M. S., Thompson, W., & Chakravarthy, U.

(1996). Nitric oxide synthase and nitrite production in human spermatozoa: evidence that endogenous nitric oxide is beneficial to sperm motility. Molecular human reproduction, 2(11), 873–878. https://doi.org/10.1093/molehr/2.11.873

Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11976-11986). https://www.computer.org/csdl/proceedings-article/cvpr/2022/694600l1966/1H1kqLpmaT6

Lum, L., & Blobel, C. P. (1997). Evidence for distinct serine protease activities with a potential role in processing the sperm protein fertilin. Developmental Biology, 191(1), 131-145. https://doi.org/10.1006/dbio.1997.8609

Maalej, R., Abdelkefi, O., & Daoud, S. (2025). Advancements in automated sperm morphology analysis: a deep learning approach with comprehensive classification and model evaluation. Multimedia Tools and Applications, 84, 27345-27378. https://doi.org/10.1007/s11042-024-20188-w

Maggi, G., Silveira, D., Vieira, M. B., da Silva, I. G. M., Barros-Cordeiro, K. B., Severo, N. C., Báo, S. N., & Mondadori, R. G. (2024). High percentage of midpiece defects in Brangus bull sperm with no reduction in sperm kinematics. Reproduction in domestic animals = Zuchthygiene, 59(5), e14585. https://doi.org/10.1111/rda.14585

Mai, Z., Yang, D., Wang, D., Zhang, J., Zhou, Q., Han, B., & Sun, Z. (2024). A narrative review of mitochondrial dysfunction and male infertility. Translational andrology and urology, 13(9), 2134–2145. https://doi.org/10.21037/tau-24-262

Marchetti, C., Jouy, N., Leroy-Martin, B., Defossez, A., Formstecher, P., & Marchetti, P. (2004). Comparison of four fluorochromes for the detection of the inner mitochondrial membrane

potential in human spermatozoa and their correlation with sperm motility. Human reproduction (Oxford, England), 19(10), 2267–2276. https://doi.org/10.1093/humrep/deh416

Marín, R., & Chang, V. (2021). Impact of transfer learning for human sperm segmentation using deep learning. Computers in Biology and Medicine, 136, 104687. https://doi.org/10.1016/j.compbiomed.2021.104687

Matamoros-Volante, A., Moreno-Irusta, A., Torres-Rodriguez, P., Giojalas, L., Gervasi, M. G., Visconti, P. E., & Trevino, C. L. (2018). Semi-automatized segmentation method using image-based flow cytometry to study sperm physiology: the case of capacitation-induced tyrosine phosphorylation. Molecular Human Reproduction, 24(2), 64-73. https://doi.org/10.1093/molehr/gax062

Meyers S. A. (2005). Spermatozoal response to osmotic stress. Animal reproduction science, 89(1-4), 57–64. https://doi.org/10.1016/j.anireprosci.2005.06.026

Milewski, A. J., Kuczyńska, A., Pawłowski, M., Martynowicz, I., Deluga-Białowarczuk, S., Sieczyński, P., Kuczyński, W., & Milewski, R., (2024). The influence of cryopreservation on sperm morphology and its implications in terms of fractions of higher-quality sperm. Journal of Clinical Medicine, 13 (24), 7562. https://doi.org/10.3390/jcm13247562

Movahed, R. A., Mohammadi, E., & Orooji, M. (2019). Automatic segmentation of Sperm's parts in microscopic images of human semen smears using concatenated learning approaches. Computers in biology and medicine, 109, 242–253. https://doi.org/10.1016/j.compbiomed.2019.04.032

Murphy, E. M., Kelly, A. K., O'Meara, C., Eivers, B., Lonergan, P., & Fair, S. (2018). Influence of bull age, ejaculate number, and season of collection on semen production and sperm motility parameters in Holstein Friesian bulls in a commercial artificial insemination centre. Journal of

animal science, 96(6), 2408–2418. https://doi.org/10.1093/jas/sky130

Odinius, T.S., Siuda, M., Lautner, M., Leiding, C., Neuner, S., Bollwein, H. & Malama, E. (2024). Sperm functional status: a multiparametric assessment of the fertilizing potential of bovine sperm. Veterinary Sciences, 11(12), 678. https://doi.org/10.3390/vetsci11120678

Ozkavukcu, S., Erdemli, E., Isik, A., Oztuna, D., & Karahuseyinoglu, S. (2008). Effects of cryopreservation on sperm parameters and ultrastructural morphology of human spermatozoa. Journal of Assisted Reproduction and Genetics, 25(8), 403-411. https://doi.org/10.1007/s10815-008-9232-3

Palacin, I., Santolaria, P., Alquezar-Baeta, C., Soler, C., Silvestre, M. A., & Yániz, J. (2020). Relationship of sperm plasma membrane and acrosomal integrities with sperm morphometry in Bos taurus. Asian journal of andrology, 22(6), 578–582. https://doi.org/10.4103/aja.aja_2_20

Palazzese, L., Anzalone, D. A., Turri, F., Faieta, M., Donnadio, A., Pizzi, F., Pittia, P., Matsukawa, K., & Loi, P. (2020). Whole genome integrity and enhanced developmental potential in ram freeze-dried spermatozoa at mild sub-zero temperature. Scientific Reports, 10, 18873. https://doi.org/10.1038/s41598-020-76061-x

Park, M., Yoon, H., Kang, B. H., Lee, H., An, J., Lee, T., Cheong, H. T., & Lee, S. H. (2023). Deep Learning-Based Precision Analysis for Acrosome Reaction by Modification of Plasma Membrane in Boar Sperm. Animals : an open access journal from MDPI, 13(16), 2622. https://doi.org/10.3390/ani13162622

Pelzman, D. L., & Sandlow, J. I. (2024). Sperm morphology: Evaluating its clinical relevance in contemporary fertility practice. Reproductive medicine and biology, 23(1), e12594. https://doi.org/10.1002/rmb2.12594

Peris-Frau, P., Álvarez-Rodríguez, M., Martín-Maestro, A., Iniesta-Cuerda, M., Sánchez-Ajofrín,

I., Garde, J. J., Rodriguez-Martinez, H., & Soler, A. J. (2019). Comparative evaluation of DNA integrity using sperm chromatin structure assay and Sperm-Ovis-Halomax during in vitro capacitation of cryopreserved ram spermatozoa. Reproduction in domestic animals, 54 Suppl 4, 46–49. https://doi.org/10.1111/rda.13519

Perry V. E. A. (2021). The role of sperm morphology standards in the laboratory assessment of bull fertility in Australia. Frontiers in veterinary science, 8, 672058. https://doi.org/10.3389/fvets.2021.672058

Pintus, E., Kadlec, M., Jovičić, M., Sedmíková, M., & Ros-Santaella, J. L. (2018). Aminoguanidine protects boar spermatozoa against the deleterious effects of oxidative stress. Pharmaceutics, 10(4), 212. https://doi.org/10.3390/pharmaceutics10040212

Prisant, N., Escalier, D., Soufir, J. C., Morillon, M., Schoevaert, D., Misrahi, M., & Tachdjian, G. (2007). Ultrastructural nuclear defects and increased chromosome aneuploidies in spermatozoa with elongated heads. Human reproduction (Oxford, England), 22(4), 1052–1059. https://doi.org/10.1093/humrep/del481

Rengan, A. K., Agarwal, A., van der Linde, M., & du Plessis, S. S. (2012). An investigation of excess residual cytoplasm in human spermatozoa and its distinction from the cytoplasmic droplet. Reproductive biology and endocrinology, 10, 92. https://doi.org/10.1186/1477-7827-10-92

Riordon, J., McCallum, C., & Sinton, D. (2019). Deep learning for the classification of human sperm. Computers in biology and medicine, 111, 103342. https://doi.org/10.1016/j.compbiomed.2019.103342

Rodríguez-Martínez H. (2006). Can we increase the estimated value of semen assessment?. Reproduction in domestic animals, 41 Suppl 2, 2–10. https://doi.org/10.1111/j.1439-0531.2006.00764.x

Saadi, H. A. S., van Riemsdijk, E., Dance, A. L., Rajamanickam, G. D., Kastelic, J. P., & Thundathil, J. C. (2013). Proteins associated with critical sperm functions and sperm head shape are differentially expressed in morphologically abnormal bovine sperm induced by scrotal insulation. Journal of Proteomics, 82, 64–80. https://doi.org/10.1016/j.jprot.2013.02.027

Skowronek, M. F., Pietroroia, S., De Cola, G., Ramos, M., Silvera, D., Casanova, G., Lecumberry, F., Cassina, A., & Sapiro, R. (2025). Mitochondrial morphology in fertile and infertile men: image processing and morphometric analysis of the sperm midpiece. Frontiers in Cell and Developmental Biology, 13, 1609081. https://doi.org/10.3389/fcell.2025.1609081

Sutovsky, P. (2015). New approaches to boar semen evaluation, processing and improvement. Reproduction in Domestic Animals, 50, 11–19. https://doi.org/10.1111/rda.12554

Tan, M., & Le, Q. (2021). EfficientNetV2: Smaller models and faster training. In Proceedings of the 38th International Conference on Machine Learning (PMLR) (Vol. 139, pp. 10096-10106). https://proceedings.mlr.press/v139/tan21a.html

Teves, M. E., Roldan, E. R. S., Agudo-Rios, C., Farnsworth, J., & Boyle, L. (2025). High-throughput analysis of postmeiotic male germ cells using image flow cytometry. Methods in molecular biology (Clifton, N.J.), 2897, 577–588. https://doi.org/10.1007/978-1-0716-4406-5_38

Thundathil, J. C., Dance, A. L., & Kastelic, J. P. (2014). Bovine sperm abnormalities: Prevalence, etiology and mechanisms leading to infertility. Clinical Theriogenology, 6(4), 525-532. https://doi.org/10.58292/CT.v6.11238

Upadhyay, V. R., Ramesh, V., Dewry, R. K., Kumar, G., Raval, K., & Patoliya, P. (2021). Implications of cryopreservation on structural and functional attributes of bovine spermatozoa: An overview. Andrologia, 53(8), e14154. https://doi.org/10.1111/and.14154

Vicente-Carrillo, A., Álvarez-Rodríguez, M., & Rodriguez-Martinez, H. (2023). The Cation/Calcium Channel of Sperm (CatSper): A common role played despite inter-species variation?. International journal of molecular sciences, 24(18), 13750. https://doi.org/10.3390/ijms241813750

Vincent, P., Underwood, S. L., Dolbec, C., Bouchard, N., Kroetsch, T., & Blondin, P. (2014). Bovine semen quality control in artificial insemination centers. In R. M. Hopper (Ed.), Bovine Reproduction (pp. 747-759). Wiley. https://doi.org/10.1002/9781118833971.ch74

World Health Organization. (2021). WHO laboratory manual for the examination and processing of human semen (6th ed.). World Health Organization. https://www.who.int/publications/i/item/9789240030787

Yurchuk, T., Petrushko, M., Gapon, A., Piniaiev, V., & Kuleshova, L. (2021). The impact of cryopreservation on the morphology of spermatozoa in men with oligoasthenoteratozoospermia. Cryobiology, 100, 117–124. https://doi.org/10.1016/j.cryobiol.2021.02.009

Zhang, R., Wang, X., Liu, R., Mei, Y., Miao, X., Ma, J., Zou, L., Zhao, Q., Bai, X., & Dong, Y. (2025). Proteomics and metabolomics analyses of mechanism underlying bovine sperm cryoinjury. BMC genomics, 26(1), 63. https://doi.org/10.1186/s12864-025-11258-w

8. APPENDICES

Article reference:

Umirbaeva A, Kurenkov A, Seisenov B, Zhumanov K, Tadziyev K, Mustafin M, Vorobjev IA and Barteneva NS (2026) Deep learning-enabled morphology analysis of bovine sperm for label-free imaging flow cytometry. Front. Vet. Sci. 13:1634224. doi: 10.3389/fvets.2026.1634224

Table A1. Photometric and kinematic data for fresh semen obtained in spring and fall seasons and frozen samples cryopreserved in summer for all 6 bulls.

Fresh, Spring														
Bull ID	**Date**	**Cell conc, 10^9/ml**	**Cell count**	**TM, %**	**PM, %**	**VAP, µm/s**	**VSL um/s**	**VCL, µm/s**	**ALH, µm**	**BCF, Hz**	**STR, %**	**LIN, %**	**Elong, %**	**Area, µm²**
Bull 1	23.05.2024	-	39	33,33	15,38	120.6±34.9	78.6±34.4	218.9±49.5	7.7±1.2	22.1±12.7	66±21	37±14	49±8.3	4.5±0.7
Bull 2	04.06.2024	189,0	12	25,00	16,67	90.4±15.6	65.8±5.1	152±19	5.2±0.4	22.1±16	74±9	44±5	47±11.1	4±0.3
Bull 3	04.06.2024	889,4	77	1,29	1,29	88.3±0	82.3±0	117.7±0	3.7±0	32.7±0	93±0	70±0	58±0	3.6±0
Bull 4	04.06.2024	1473,0	76	56,58	43,42	82.9±36.5	73.6±35	137.4±44.3	6.6±2.4	24±10.6	89±8	53±12	43±8.9	5±1.2
Bull 5	23.05.2024	-	38	60,53	44,74	141.3±28.9	114.7±33.8	221.3±47.3	7.4±3.4	26.7±20	81±17	53±14	49±11.5	4.3±0.3
Bull 6	04.06.2024	863,5	45	35,55	17,77	85.5±54.6	55.2±33.8	139.7±60.9	4.8±0.7	31±10.2	70±21	41±116	47±11.5	4.9±0.8
Fresh, Fall														
Bull 1	10.09.2024	0,960	111	60,36	50,45	172.4±57	147.6±54.8	253.9±74.8	8.7±3	15.9±19.8	85±15	59±17	46±11.8	4.7±0.8
Bull 2	06.09.2024	0,672	62	70,97	67,74	160.7±32.9	151.4±38.3	222.7±51	7.2±2.2	22.6±18.7	93±11	69±17	43±10.9	4.8±0.8
Bull 3	06.09.2024	1,092	273	71,79	41,39	65.9±35.4	63.5±34.1	76.9±50.2	2.7±2.5	29.9±12.5	97±7	88±15	47±13.3	5.5±5
Bull 4	06.09.2024	1,373	219	80,82	70,32	158.5±44.7	138.6±48.3	236.6±61	8.6±3.4	18.7±17.7	87±15	60±19	43±11.1	4.8±1.4
Bull 5	10.09.2024	1,015	134	90,30	76,12	152.7±54.3	134.9±56.8	214.2±71.7	7.5±2.8	18.3±18.6	88±15	65±20	42±13.1	5.5±8.1
Bull 6	06.09.2024	0,512	103	70,87	54,37	136.2±47.6	116.1±48.2	197.6±69	7.6±2.9	22.7±18.9	84±15	59±16	46±14.4	4.7±1.1
Frozen, Summer														
Bull	Date frozen	Date thawed	Cell count	TM, %	PM, %	VAP, µm/s	VSL µm/s	VCL, µm/s	ALH, µm	BCF, Hz	STR, %	LIN, %	Elong, %	Area, µm²
Bull 1	02.05.2024	10.09.2024	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Bull 2	06.08.2024	06.09.2024	671	84.2	37.7	62.5±39.1	54.3±32.4	83.3±62.7	3.9±3.5	20.9±13.6	90±16	75±23	47±15.8	7.6±7.6

Bull 3	10.09.2024	13.09.2024	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Bull 4	27.08.2024	06.09.2024	514	70.23	44.16	117.3±54.8	92.1±56	176.1±76.5	6.4±3.8	12.5±16	78±20	56±23	44±13.3	4.6±1.6
Bull 5	26.08.2024	10.09.2024	1038	82.85	45.57	96±54.4	73.3±50.4	149.9±76.4	6.9±4.2	17±16.1	76±20	52±23	45±15	5.6±4.2
Bull 6	20.08.2024	06.09.2024	450	92.44	66.67	86.8±38.8	74±36.4	122.2±56.5	4.9±3.1	18.3±15.1	85±17	65±24	47±12.9	4.7±1.8

Note: TM - total motility, PM - progressive motility, VAP - average path velocity, VSL - straight line velocity, VCL - curvilinear velocity, ALH - amplitude of lateral head, BCF - beat cross frequency, STR - straightness, LIN - linearity, Elong - elongation, Area - area of a sperm head

Table A2. Descriptive data of the number of IFC images extracted from each bull sample for morphological assessment and viability analysis (before additional filtering out).

Bull ID	Fresh, Spring	Fresh, Fall	Frozen, Summer
Bull 1	22614	18271	30592
Bull 2	25595	22008	31196
Bull 3	24267	23057	22387
Bull 4	31457	19562	22864
Bull 5	21905	20593	26251
Bull 6	26568	20118	27069

Table A3. Breakdown of split sizes of the labeled data for different experimental stages.

Experiment	Total Available	Train	Val	Test	Notes
Single-breed, single-condition	1 breed × 1 condition × 10 classes × 200 = 2,000	1,600	200	200	Used for within-breed self and cross-breed external tests (external tests use the target breed's 200-image test split in the same condition).
Condition-aggregated (Frozen)	3 breeds × 10 × 200 = 6,000	1,600 (capped)	200 (capped)	600	Train on Frozen; report Frozen→Frozen (self) and Frozen→Fresh (external).
Condition-aggregated (Fresh)	3 breeds × 10 × 200 = 6,000	1,600 (capped)	200 (capped)	600	Train on Fresh; report Fresh→Fresh (self) and Fresh→Frozen (external).

LOBO (per fold)	Train on 2 breeds × 2 conditions × 10 × 200 = 8,000; test on held-out breed (2 cond × 10 × 200 = 4,000)	6,400	1,600	4,000	k=3 folds (held-out breed rotates). 3 seeds per fold. Validation is a 20% hold-out from the training pool.
Data-fraction runs (LP/LP-FT, per LOBO fold)	same as LOBO training pool above (8,000 total, 6,400 train after 20% val)	f × 6,400	1,600	4,000	Fractions f∈{0.1,0.2,0.5,1.0} applied to the training split only; val/test kept fixed. So train counts are 640, 1,280, 3,200, 6,400 respectively.

Table A4. Summary of generalization analysis results. Data source specified in the Train column was used for model training, and data source specified in the Test column was used for the model evaluation. Accuracy and macro F1 are reported with 95% CI.

Train	**Test**	**Accuracy (%) [95% CI]**	**Macro F1 [95% CI]**
Frozen Condition			
Kazakh Whitehead	Kazakh Whitehead	74.60 [73.9-75.3]	0.74 [0.73-0.75]
	Simmental	66.80 [66.0-67.6]	0.67 [0.66-0.68]
	Auliekol	59.86 [59.0-60.7]	0.60 [0.59-0.61]
Simmental	Kazakh Whitehead	81.67 [80.9-82.4]	0.82 [0.81-0.83]
	Simmental	83.49 [82.8-84.2]	0.83 [0.82-0.84]
	Auliekol	75.76 [74.9-76.6]	0.76 [0.75-0.77]
Auliekol	Kazakh Whitehead	73.76 [72.9-74.6]	0.73 [0.72-0.74]
	Simmental	70.13 [69.2-71.0]	0.69 [0.68-0.70]
	Auliekol	75.75 [74.9-76.6]	0.75 [0.74-0.76]
Fresh Condition			
Kazakh Whitehead	Kazakh Whitehead	71.27 [70.5-72.0]	0.71 [0.70-0.72]
	Simmental	68.43 [67.6-69.2]	0.68 [0.67-0.69]
	Auliekol	67.99 [67.2-68.8]	0.68 [0.67-0.69]
Simmental	Kazakh Whitehead	66.51 [65.7-67.3]	0.67 [0.66-0.68]

	Simmental	85.55 [84.9-86.2]	0.85 [0.84-0.86]
	Auliekol	60.08 [59.2-60.9]	0.60 [0.59-0.61]
Auliekol	Kazakh Whitehead	65.09 [64.2-65.9]	0.65 [0.64-0.66]
	Simmental	65.12 [64.3-66.0]	0.65 [0.64-0.66]
	Auliekol	75.03 [74.2-75.8]	0.75 [0.74-0.76]
Aggregated			
Frozen	Frozen	83.58 [82.9-84.2]	0.83 [0.82-0.84]
	Fresh	68.75 [67.9-69.6]	0.68 [0.67-0.69]
Fresh	Frozen	66.12 [65.3-66.9]	0.66 [0.65-0.67]
	Fresh	76.39 [75.7-77.1]	0.76 [0.75-0.77]

Table A5. Number of images sorted into 10 morphological groups for all the six bulls both for the frozen and fresh, conditioned by season.

Frozen, Summer						
Class	**Bull 1**	**Bull 2**	**Bull 3**	**Bull 4**	**Bull 5**	**Bull 6**
AM	278	473	336	1200	936	1159
AT	157	637	270	2137	1135	993
CTM	129	171	71	317	275	166
DCD	389	1849	113	187	395	540
TEH	2269	2736	2464	2216	3098	3572
IHS	1974	2011	1719	1960	2814	2895
PCD	131	55	27	83	124	89
Multiple	99	119	129	77	110	87
Debris	4025	8031	6855	5259	6736	5839
Normal	12662	12860	9117	7926	9181	10029
Total*	22113	28942	21101	21362	24804	25369
% Normal**	70.02	61.50	63.99	49.46	51.12	51.58
Fresh, Spring						

AM	2637	611	137	89	1419	1400
AT	982	360	65	179	773	2174
CTM	439	436	46	197	466	868
DCD	469	705	257	604	936	1361
TEH	2512	3267	3313	3208	3173	2822
IHS	2761	1807	1525	2007	2029	1485
PCD	3023	3512	3712	5217	2385	1755
Multiple	211	441	695	725	210	402
Debris	4918	4658	3771	4815	5565	3939
Normal	3359	8519	9556	11268	4052	7837
Total*	21311	24316	23077	28309	21008	24043
% Normal**	20.49	44.33	51.35	49.49	26.24	38.98
Fresh, Fall						
AM	1520	1209	115	826	316	1489
AT	751	683	78	109	190	738
CTM	203	92	26	28	45	51
DCD	356	163	93	135	217	380
TEH	661	396	1731	318	366	384
IHS	777	1522	1802	1704	2135	1384
PCD	539	240	88	288	216	213
Multiple	287	560	270	490	345	389
Debris	7525	9698	6629	8623	11083	8760
Normal	4326	5948	10787	5884	4965	4634
Total*	16945	20511	21619	18405	19878	18422
% Normal**	45.92	55.01	71.96	60.15	56.45	47.96

Note: *Total number of images was calculated by excluding non-suitable images (as described in Section 2.5.1) from the total number of acquired images, as shown in Table S2.

**The percentage of cells with normal morphology was calculated by dividing the number of cells classified as "normal" by the total number of analyzed cells, excluding events labeled as “Multiple” and “Debris.”

Table A6. High and low MMP distribution across sperm morphological groups (per dataset).

	Bull 1				**Bull 2**				**Bull 3**			
	daf.1	**daf.2**	**daf.3**	**daf.4**	**daf.1**	**daf.2**	**daf.3**	**daf.4**	**daf.1**	**daf.2**	**daf.3**	**daf.4**
Total	3643	3600	3510	3551	3284	3259	3261	3181	3305	3236	3353	3430
Fluo-4 positive, %	97.80	98.30	98.4	98.3	89.59	91.04	90.92	89.85	90.26	90.82	92.04	91.43
High MMP, %	23.50	21.28	20.54	19.18	26.04	28.29	27.94	28.20	27.59	31.67	30.12	33.70
Low MMP, %	69.01	71.89	73.36	74.60	73.02	70.94	70.78	71.02	70.92	66.56	68.06	63.76

Table A7. Post-hoc pairwise analysis of intracellular Ca^{2+} levels between sperm morphological groups using Dunn’s test.

Bull 1										
	Normal	**AT**	**AM**	**CTM**	**TEH**	**IHS**	**PCD**	**DCD**	**Debris**	**Multiple**
Normal	—	ns	<0.001	ns	0.002	ns	ns	ns	ns	<0.001
AT	ns	—	<0.001	ns	0.002	ns	ns	ns	ns	<0.001
AM	<0.001	<0.001	—	<0.001	<0.001	<0.001	0.011	<0.001	<0.001	0.001
CTM	ns	ns	<0.001	—	ns	ns	ns	ns	ns	<0.001
TEH	0.002	0.002	<0.001	ns	—	ns	ns	ns	0.002	<0.001
IHS	ns	ns	<0.001	ns	ns	—	ns	ns	ns	<0.001
PCD	ns	ns	0.011	ns	ns	ns	—	ns	ns	<0.001
DCD	ns	ns	<0.001	ns	ns	ns	ns	—	ns	<0.001

Debris	ns	ns	<0.001	ns	0.002	ns	ns	ns	—	<0.001
Multiple	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	—
Bull 2										
	Normal	**AT**	**AM**	**CTM**	**TEH**	**IHS**	**PCD**	**DCD**	**Debris**	**Multiple**
Normal	—	<0.001	<0.001	ns	ns	<0.001	ns	ns	<0.001	<0.001
AT	<0.001	—	<0.001	ns	0.001	<0.001	ns	ns	ns	<0.001
AM	<0.001	<0.001	—	<0.001	<0.001	<0.001	ns	<0.001	<0.001	ns
CTM	ns	ns	<0.001	—	ns	ns	ns	ns	ns	<0.001
TEH	ns	0.001	<0.001	ns	—	ns	ns	ns	0.001	<0.001
IHS	<0.001	<0.001	<0.001	ns	ns	—	ns	0.023	<0.001	<0.001
PCD	ns	ns	ns	ns	ns	ns	—	ns	ns	ns
DCD	ns	ns	<0.001	ns	ns	0.023	ns	—	ns	<0.001
Debris	<0.001	ns	<0.001	ns	0.001	<0.001	ns	ns	—	<0.001
Multiple	<0.001	<0.001	ns	<0.001	<0.001	<0.001	ns	<0.001	<0.001	—
Bull 3										
	Normal	**AT**	**AM**	**CTM**	**TEH**	**IHS**	**PCD**	**DCD**	**Debris**	**Multiple**
Normal	—	<0.001	<0.001	<0.001	ns	ns	ns	ns	<0.001	0.003
AT	<0.001	—	ns	ns	<0.001	<0.001	ns	ns	0.001	ns
AM	<0.001	ns	—	ns	<0.001	<0.001	ns	<0.001	<0.001	ns
CTM	<0.001	ns	ns	—	<0.001	<0.001	ns	0.019	0.005	ns
TEH	ns	<0.001	<0.001	<0.001	—	ns	ns	ns	<0.001	<0.001
IHS	ns	<0.001	<0.001	<0.001	ns	—	ns	ns	0.007	0.001
PCD	ns	ns	ns	ns	ns	ns	—	ns	ns	ns
DCD	ns	ns	<0.001	0.019	ns	ns	ns	—	ns	0.028

Debris	<0.001	0.001	<0.001	0.005	<0.001	0.007	ns	ns	—	0.025
Multiple	0.003	ns	ns	ns	<0.001	0.001	ns	0.028	0.025	—

Note: ns = non-significant