**Awards issued prior to January 20, 2025, were funded under previous Administrations and may not reflect the priorities and policies of the current Administration.**

**United States Department of Agriculture**
**Project Initiation**

# Project Data for Accession Number 1025817

## Project Title

Improving herd reproductive efficiency utilizing artificial intelligence assisted sperm analysis & support apps

## Project Details

| | |
|---|---|
| **Sponsoring Institution**<br>National Institute of Food and Agriculture | **Project Status**<br>COMPLETE |
| **Funding Source**<br>NIFA Non Formula | **Grant Year**<br>2021 |
| **Grant No.**<br>2021-33530-34400 | **Cumulative Award Amt.**<br>$100,000.00 |
| **Proposal No.**<br>2021-00998 | **Multistate No.**<br>(N/A) |
| **Project Start Date**<br>Jul 01, 2021 | **Project End Date**<br>Feb 28, 2023 |
| **Program Code**<br>[8.3] Animal Production & Protection | |

## Project Director

Jensen, G.

## Recipient Organization

EMERY ANIMAL HEALTH

490 W HWY 29

CASTLE DALE, UTAH 84513

## Performing Department

(N/A)

## Non Technical Summary

Bulls are a major contributing factor to the reproductive efficiency and economic viability of beef and dairy businesses. However, current speriograms are incomplete, subjective,unreliable, and without a system for bull buyers to easily access and understand results. Sperm morphology is a critical component of fertility evaluation during a breeding soundness examination and freezing semen. Using machine learning, a form of artificial intelligence, we will develop a chute side semen analysis tool that will compile an objective complete differential spermiogram with consistent accuracy unmatched by humans. This chute side diagnostic tool will remove the subjective nature of semen analysis. Once completed will be an invaluable tool to quickly and accurately assess semen during a bull breeding soundness exam. Accompanying apps will help veterinarians and cow-calf producers understand bull testing results needed tobetter make management decisions for improving herd fertility. Accurate consistent results will be assured through defined processes, careful selection of microscopic equipment, reliable image data, and deep machine learning

technologies.Veterinarians, artificial insemination centers, researchers, and otherscan use this chute side semen evaluation toolchute sidefor quick accurateresults. Cow-calf producers will improveprofits through improvingreproductive efficiency by using bulls and semen with increasedpotential for higher pregnancy rates. The US beef industry can realize up to $1.2 billion from improved production for improved sustainability.

## Animal Health Component

Animal Health Component: 100%

## Research Effort Categories

| | |
|---|---|
| Basic | 0% |
| Applied | 0% |
| Developmental | 100% |

## Classification

| Knowledge Area (KA) | Subject of Investigation (SOI) | Field of Science (FOS) | Percent |
|---|---|---|---|
| 301 | 3399 | 1081 | 100% |

## Knowledge Area

[301] Reproductive Performance of Animals

## Subject Of Investigation

[3399] Beef cattle, general/other

## Field Of Science

[1081] Breeding

## Keywords

Computer Vision, Artificial Intelligence, Bull Breeding Soundness, Herd fertility, Sperm Morphology

## Goals / Objectives

Goals:Our final goal is to createchute side diagnostic tool that will objectively analyze bull semen for motility and morphology. This project phase will focus on two areas: 1) Obtaining a large data set of quality labeled sperm images. 2) Determining the best machine learning (ML) algorithms or sets of algorithms for optimal results.At completion we will have a large dataset of bull sperm labeled images, know where our data set weakness may be and what we need both in additional quantity of images and potential quality of labels and image quality for bull sperm morphological assessment.Have a functional set of processes and ML algorithms that will create reliable results for bull sperm morphological analysis.Technical ObjectivesCreate a high-quality data set with 20,000 - 30,000 labeled images of sperm normal and abnormal morphology for ML training.Identify the most promising open-source CNN architectures and use these to train a model, or multiple models, for the purposes of comprehensive bull sperm morphological

analysis.Develop standardized data processing flow for uploading images from the microscope camera into the application.Develop a beta, proof of concept, application that can deploy the model developed in objective #2 and obtain prediction results in a reasonable amount of time on a common mid-level laptop or mobile device.

## Project Methods

DatasetsTo compile an effective data set with the ability to create modules which will identify specific sperm abnormalities with a high percent of recall and precision we will start with a list and example images of the labels (morphological abnormalities) to use for labeling. This reference data set will represent the variations of morphological labels and may be adjusted as needed. The reference data set will be reviewed by a member of the Collage of Theriogenology.We will create labeled datasets to be used in Google's computer visions training. Our final data sets will have at least 1,000 samples of each label. The ML training creates a module, and our tested results will give us the recall and precision of the model created from the data set. The data set will then be reviewed to identify errors or needs which can improve recall and precision. These include changes like breaking down morphological labels into more distinct morphological characteristics or excluding images of sperm that are out of focus or being obscured by other objects.The refined datasets will then be used to compare different labeling techniques. Bonding boxes are used in labeling Google's computer vision ML, pixel labeling technics are used in APEER, Zeiss's annotation methods, which takes more time but may be a superior way to segment individual sperm cells from each other and other types of material within the microscopic field. We will consult with individuals who have experience in this area and run A/B tests where needed to compare methods. This can be accomplished by taking the exact same data set, labeling sperm using the different methods and then compare results.Machine learning architype and processes.Application development will utilize common code languages such as Python and Java Script, as well as open-source frameworks for ML model development and deployment, including the popular and widely used TensorFlow and Keras frameworks. We will also make use of open-source CNN architectures available on conde sharing platforms such as GitHub as we evaluate various network architectures for the specific problem of sperm morphology detection for practical usage in the field. We currently have sufficient images and labeled datasets to begin this process. Therefore, app development and dataset creation can occur concurrently.We will focus first on developing a model that can accurately identify readable sperm and tally the total sperm count in an image. We will then develop a model that will identify morphological abnormalities present in the individually identified sperm. We will be making use of the approach popularized by Andrew Ng in which we build the first model quickly and then iterate to identify the most successful hyperparameters and data processing strategies give the characteristics of the unique problem domain. We will start with a basic set of abnormalities and then iterate through this process as many times as needed. To extend the number of abnormalities that can be comprehensively predicted.In evaluating the resulting trained model(s) we will be looking for cross validation accuracy, recall, and precision, as well as the feasibility of using the model in an application. This includes analyzing the physical storage size of the saved model, including parameters and weights, its impact on application footprint, and model performance when running prediction locally on hardware not specialized for machine learning computation.We will run A/B trials to compare two ML architectures. The results can be compared to select the superior method. This can be repeated by making changes to the ML methods and architypes each time comparing results to select the superior process and achieve a high percent recall and precision of morphological labels.Once reasonable confidence in model architecture and feasibility has been achieved, we will begin laying the groundwork for the basic application, iterating, and adding features as time allows, including but not limited to:Identifying and building skeletal application code targeted to specific devices, starting with environments most likely to be commonly used by veterinary consumers of the application (i.e. Android, iOS, Windows/.NET)Setting up cloud-based application serving and centralized data aggregationSyncing prediction results obtained in offline mode to a centralized database.Increasing the quality of BBSE data gathered by the application.Increase/improve analysis generated from gathered data and insights from that data.Diversification of users to include client user views in addition to veterinary user views.A write-up or record of each iteration will be created including algorithms and changes made including results of each. We will also create a tabular record for easy comparison and analysis of results. These will be reviewed by the PI and AI engineer on a regular basis to make selections and changes as needed base on the results.

**United States Department of Agriculture**
**Project Initiation**