Sperm Hunter - Neural Network Model for Sperm Identification in Azoospermic Males

BY

SHRIKANT DILIPKUMAR PANDYA

B.S., University of Illinois at Chicago, 2014

THESIS

Submitted as partial fulfillment of the requirements

for the degree of Doctor of Philosophy in Bioengineering

in the Graduate College of the

University of Illinois at Chicago, 2021

Chicago, Illinois

Defense Committee:

Craig S. Niederberger, Chair and Advisor

Rodrigo Pagani, Urology

Thomas Royston

Cristian Luciano

Xincheng Yao

This thesis is dedicated to my wife, Sanika, and my parents. Some people are blessed with a loving and supporting family that provide support in challenging times. Others have families that push them to do more than they thought possible. I count myself among the lucky few blessed with both.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

A. Introduction

Azoospermia is defined as a lack of spermatozoa in the ejaculate (Berookhim and Schlegel 2014). This condition can be separated into two categories, obstructive azoospermia (OA), lack of sperm due to a physical blockage in the ejaculatory pathway, and azoospermia due to spermatogenic dysfunction (ASD). The latter category comprises 60% of azoospermic cases and is treated by surgically extracting tubules containing sperm directly from the testis (Practice and Medicine 2008). These seminiferous tubules are visualized under microscopy; the widest and most opaque tubules are 2 fold more likely to contain sperm and are preferentially sampled in the Microdissection Testicular Sperm Extraction (MicroTESE) procedure (Schlegel 1999).

B. Problem Statement

The acquisition of sperm dense seminiferous tubules requires a steep learning curve: novice surgeons who have performed fewer than 50 procedures achieve a sperm retrieval rate (SRR) of only 32%. Between cases 50 and 150, SRR improves to 48% (Ishikawa et al. 2010). After 792 procedures, a surgeon was able to reach an average SRR of 60% (Ramasamy et al. 2009). The limiting factor for successful sperm retrieval is the lack of clearly differentiated sperm-containing loci in the testes. Without these, surgeons are forced to perform random biopsies. There is also no clear point after which surgeons should stop searching for sperm. Studies have shown sperm retrieval is highest in cases lasting between 0 and 2 hours (88.6% SRR), but as many as 30.7% of cases lasting 4-7 hours can produce sperm (Ramasamy et al. 2011).

C. Specific Aims

There are no solutions commercially available to solve this problem. Theoretical solutions to the problem of sperm identification, including optical computed tomography (OCT) and Raman Spectroscopy (RS), require time consuming image capture and processing, and provide sperm density information proportional to their tissue invasiveness. New developments in the field of computer vision (CV) including convolutional neural networks (CNNs) show promise for applications that require fine detail identification in subjects that are superficially homogenous, such as automatic identification of cancer on digitized histopathological imagery (Mosquera-Lopez et al. 2015). This research aims to train a neural network model on seminiferous tubule biopsies to autonomously identify locations of higher sperm likelihood to maximize sperm retrieval.

The specific aims of this project are:

- Aim 1 – Develop an animal model of ASD to analyze artificially induced ASD on healthy Sprague-Dawley rats
- Aim 2 – Train a neural network to identify sperm dense locations using rodent ASD model biopsies
- Aim 3 – Develop a neural network application for intraoperative sperm identification
- Aim 4 – Evaluate model predictions on rodent ASD models

D. Significance

Nearly 2 million males in the United States suffer from ASD. Their only option for procreation is in-vitro fertilization (IVF). Before attempting IVF, these men must first undergo a surgical procedure to extract sperm from the testes. Sperm extraction is unsuccessful for around 40% of patients with ASD (Ramasamy et al. 2009). Even in successful extraction, patients may experience lower testosterone production due to loss of testosterone producing Leydig cells in the seminiferous tubules. If the extraction is traumatic, this decreased testosterone level may become chronic, especially in men with diminished androgen production such as those with Klinefelter syndrome (Sandro C. Esteves et al. 2013). In addition, these patients may experience testicular pain for days after the procedure and may have an inflammation and swelling of the testes (Sandro C. Esteves et al. 2013). These significant complications highlight the importance of initial extraction success.

Treatment for ASD relies on specialist urological surgeons who require years of training to be proficient in performing MicroTESE. Novice surgeons may be 12% less successful and 25 minutes slower per procedure than experienced surgeons (Ishikawa et al. 2010).

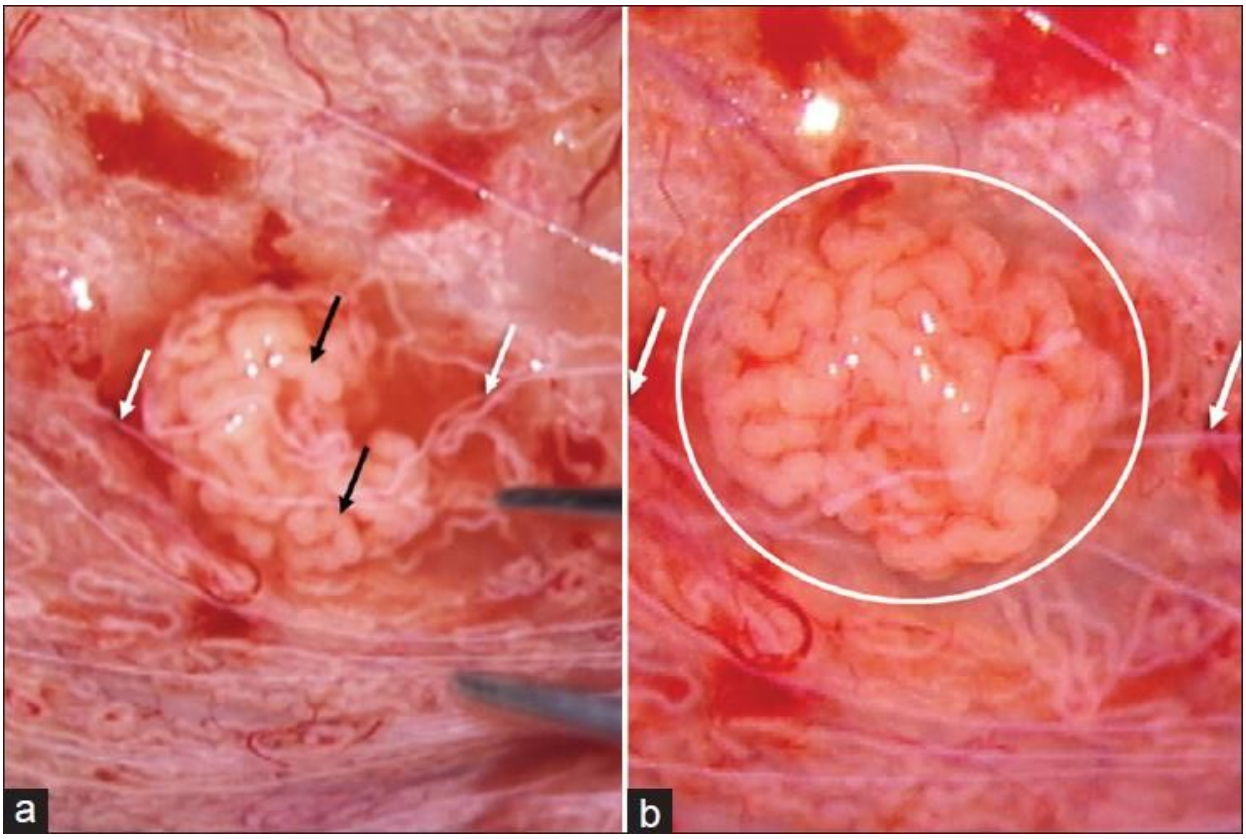


Figure 1. Clearly dilated seminiferous tubules (black arrows) surrounded by atrophied tubules (white arrows) under 25x magnification (Ashraf et al. 2013)

The purpose of this research is to create an experience-agnostic optical identification system that automatically identifies locations of high sperm likelihood during surgery. While imaging modalities exist that are able to identify sperm density in tissue post-extraction, none are available for active intra-operative guidance during sperm extraction procedures (Liu et al. 2014).

Aim 1: Develop an animal model of ASD to analyze artificially induced ASD on healthy Sprague-Dawley rats

To train a neural network model on ASD pathologies with a high enough data volume to develop model accuracy, a rat model of azoospermia was developed. Staggered injections of busulfan were used to induce controllable spermatogenic dysfunction in Sprague-Dawley rats.

A recorded biopsy procedure was used to produce the training images for the neural network training as shown in Figure 2. Sperm samples were evaluated under high magnification phase contrast microscopy to determine exact sperm counts and were encoded into the images to train the neural network.

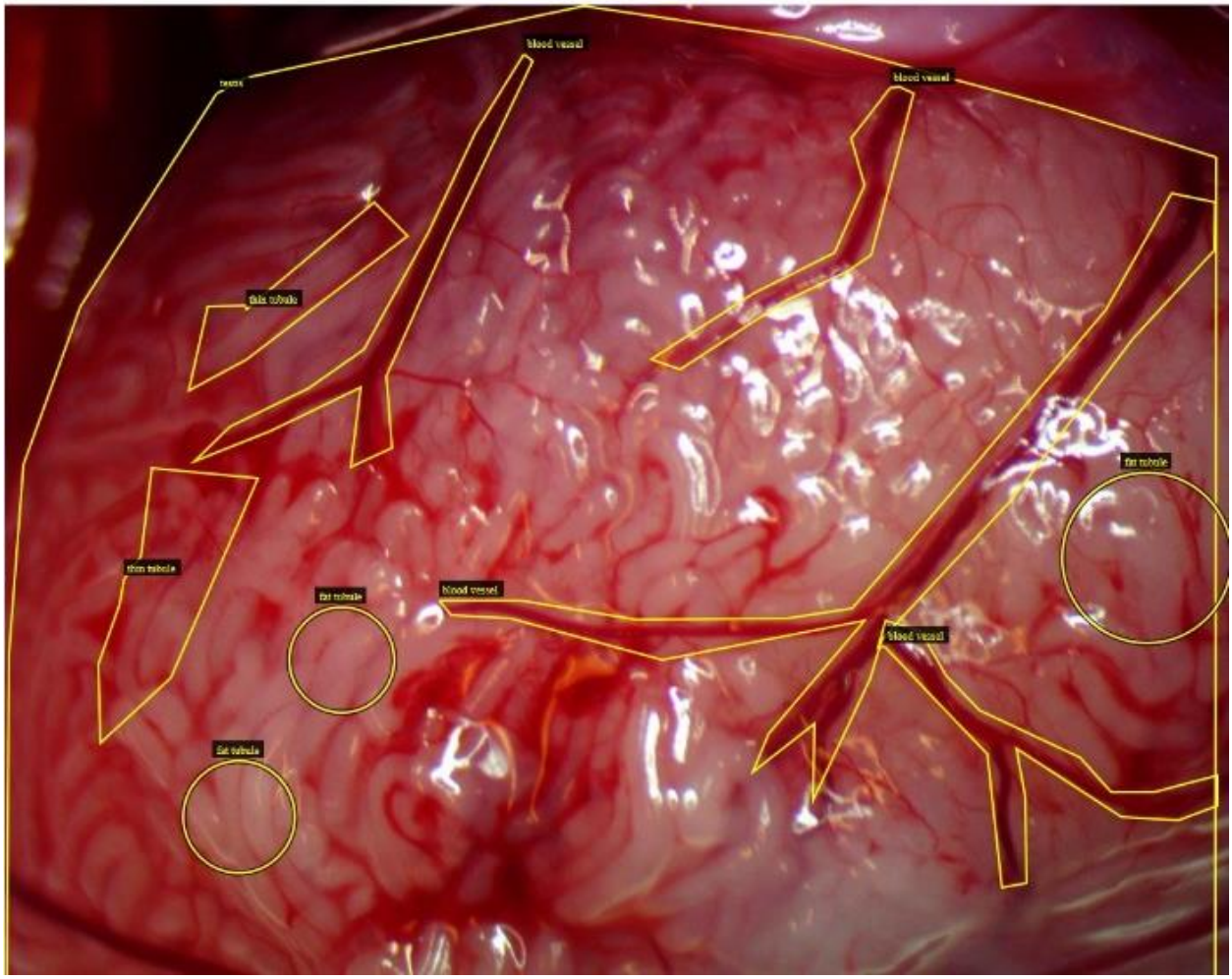


Figure 2. An annotated biopsy image identifying targeted tubules, important structures such as blood vessels are also identified.

Aim 2: Train a neural network to identify sperm dense locations using rodent ASD model biopsies

A neural network model based on the pyramid scene parsing neural network (PSP-Net) architecture was trained using the recorded and annotated biopsy images. Images were augmented to increase effective training data volume and provide non-biological variance to the data to limit the reliance on training specific factors, such as surgical lighting conditions, magnification, and camera properties.

Aim 3: Develop a neural network application for intraoperative sperm identification

Using the trained model, a real time detection application was created to be used intraoperatively as a guidance system during MicroTESE. This system was deployed on a laptop adjacent to the operative area and was operated by the surgeon during the procedure.

Aim 4: Evaluate model performance on rodent ASD models The model's predictions were evaluated by allowing the model to select 50% of the sampling locations. The other half of sampling locations were selected by a human operator in the regions not indicated by the system. A comparison of chosen vs. not chosen locations was used to evaluate the sensitivity and specificity of model predictions.

Figure 3 A prediction of high sperm yield locations projected on an azoospermic rat testis

I. Clinical Background

A. Azoospermia

Azoospermia is defined as a lack of sperm in the ejaculate (Berookhim and Schlegel 2014). Diagnosis of this condition must be confirmed by 3,000 G centrifugation of a semen specimen for 15 minutes at room temperature and examination of the resulting pellet under high-powered microscopy (Cao et al. 2011). The condition of azoospermia has two causes; one is the case of obstructive azoospermia (OA) which comprises 40% of azoospermic cases, and the other is azoospermia due to spermatogenic dysfunction (ASD), also seen in literature as non-obstructive azoospermia (NOA), which comprises the remaining 60% of azoospermic cases (Practice and Medicine 2008; Jarow, Espeland, and Lipshultz 1989).

B. Spermatogenesis

Spermatogenesis progresses in 3 phases of cell differentiation. The process starts in the basement membrane of the seminiferous tubule where the peritubular myoid (PTM) cells cooperate with Sertoli cells to create niches for spermatogonial stem cells (SSCs), the precursors to germ cells that develop into sperm. The PTM cells surround the tubule and provide the contraction force needed to push spermatozoa along the tubule. The Sertoli cells provide the signaling and growth factors needed to allow for germ cell differentiation and proliferation. Between adjacent seminiferous tubules lie Leydig cells that are responsible for the production of testosterone. SSCs differentiate in the basement membrane into two cells, one maintains the stem cell line, the other becomes an undifferentiated spermatogonium that will develop into a spermatozoa. Undifferentiated spermatogonia undergo multiple divisions with incomplete cytokinesis to form chains. These chains transform en-mass into differentiated spermatogonia after chain lengths of 16

or 32 are reached. Another series of mitosis brings the differentiated spermatogonia to the stage of preleptotene spermatocytes. These cells undergo meiosis resulting in round haploid spermatids which further differentiate into elongated spermatids and finally spermatozoa (Smith and Walker 2014).

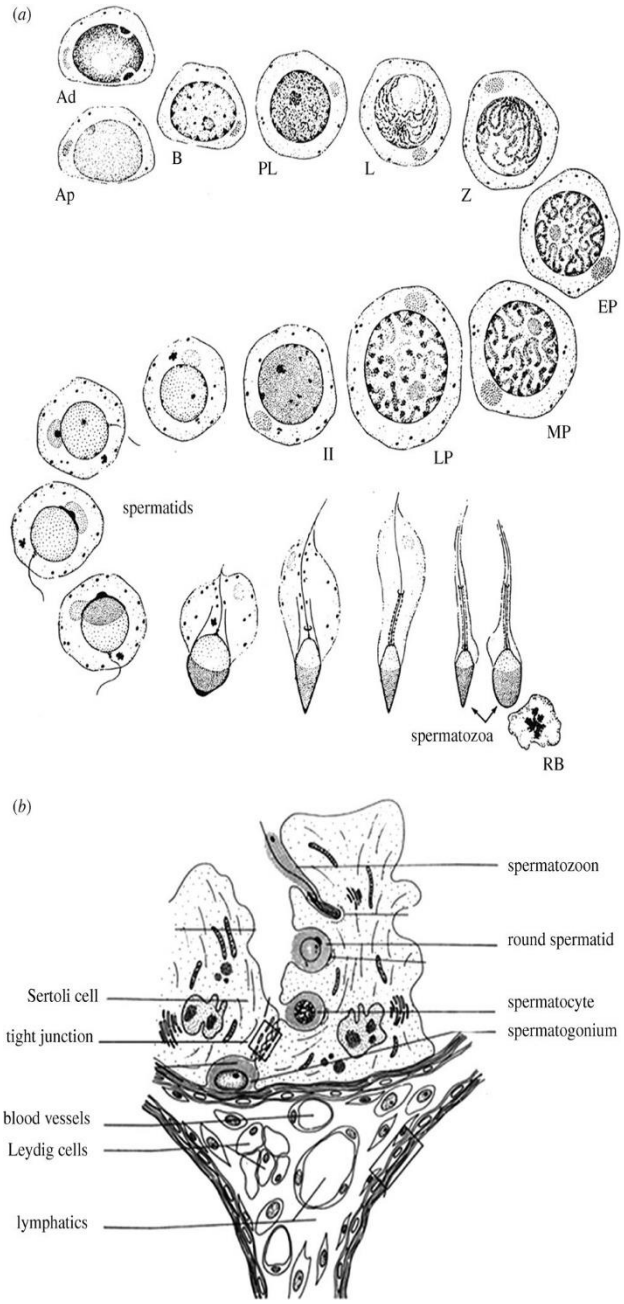


*Figure 4 (*a*) The progression of spermatogonia to sperm. Spermatogonial stem cells (SSCs) divide into preleptotene spermatocytes (PL). These further divide into leptotene, zygotene and pachytene spermatocytes. Secondary spermatocytes (II) are produced from meiotic division of the pachytene spermatocyte which divide again to form round haploid spermatids. The spermatids elongate and differentiate until they become mature spermatozoa. (*b*) A cross section of the space between three seminiferous tubules is shown. The interstitial area contains Leydig cells, blood vessels, and lymphatic tissue. A basement membrane containing PTM cells surrounds the seminiferous tubules. Sertoli cells grow from the basement membrane to the lumen of the tubule with niches where spermatogonial cells will mature into spermatozoa. (Image reproduced with permission) (Walker 2010)*

C. Obstructive Azoospermia

Obstructive azoospermia (OA) results from a physical blockage of the male excurrent ductal system at any location between the rete testis and the ejaculatory ducts (Wosnitzer, Goldstein, and Hardy 2014). In this presentation, the sperm production capability of the testes is usually undiminished. Infertility is caused by the sperm never reaching the ejaculatory duct due to the blockage. OA may be acquired by vasectomy, failure of a vasectomy reversal, post-infectious diseases, masses, surgical procedures in the scrotal, inguinal, pelvic, or abdominal regions, and trauma (S C Esteves 2011). Congenital OA can be caused by cystic fibrosis, congenital absence of the vas deferens (CAVD), ejaculatory duct or prostatic cysts, or Young's syndrome (increased viscosity of mucosal fluid impeding spermatic motility)(S C Esteves 2011). OA is usually treated surgically by the way of microsurgical reconstruction (vasovasostomy-VV or vasoepididymostomy-VE) to establish an unobstructed path for sperm (Wosnitzer, Goldstein, and Hardy 2014). Success of this method ranges from 70% - 90% for VV and 30% - 90% for VE (Matthews, Schlegel, and Goldstein 1995; Silber and Grotjan 2004). In cases where VV or VE are not appropriate surgical interventions, such as men with congenital bilateral absence of the vas deferens (CBAVD), percutaneous or microsurgical aspiration of sperm from the epididymis or testis combined with an in-virto fertilization technique such as intracytoplasmic sperm injection (ICSI) is the recommended treatment (Anger et al. 2004).

D. Azoospermia Due to Spermatogenic Dysfunction

Azoospermia due to spermatogenic dysfunction (ASD) is caused by testicular failure resulting in an inability to produce sperm. ASD can be delineated into two categories, pre-testicular and testicular causes. Pretesticular causes, also known as secondary testicular failure, manifest as

endocrine disorders such as hypogonadotropic hypogonadism (HH), hyperprolactinemia, and androgen resistance. Testicular causes include structural or locational abnormalities, such as varicocele and undescended testes respectively, as well as virally contracted and genetic causes. Viral causes of ASD include mumps orchitis which may present in pubertal and postpubertal males as scrotal swelling. Approximately 13% of these patients suffer from infertility. Acquired ASD may also be induced by gonadotoxic drug effects such as antiandrogens, exogenous androgens, or chemotherapy agents (Cocuzza, Alvarenga, and Pagani 2013).

Genetic abnormalities may also cause azoospermia. Klinefelter syndrome (KS) is defined by an X-chromosome polysomy, the most common version being XXY on the 47th chromosome, and occurs in 0.1% to 0.2% males. These patients commonly present with low testosterone levels, a micropenis, and small, firm testes. The prevalence of azoospermia in KS patients is 45 times higher than the general population. Testicular sperm extraction procedures may yield spermatozoa in 69% of KS patients. Y-chromosome polysomy (47,XYY) occurs in 1:1000 men and causes azoospermia, or severe oligozoospermia. These males exhibit antisocial behavior, decreased intelligence, and a tall stature. However, unlike KS patients, serum testosterone levels are normal in these patients (Cocuzza, Alvarenga, and Pagani 2013).

Azoospermia can be divided into three categories. Sertoli cell only syndrome (SCO) describes patients presenting with no spermatogenesis in the testis; These patients are completely sterile. In some cases, patients may initially be diagnosed with SCO but will be found to have limited sperm production upon biopsy. Hypospermatogenesis (HS) describes the condition of limited, but non-arrested, spermatogenesis throughout the testis. Maturation arrest (MA) presents as an inability for spermatogonia to progress past certain stages of development. MA is separated into two stages, early maturation arrest (EMA) where spermatogonia may develop up to the stage of a primary spermatocyte, and late maturation arrest (LMA) where development may progress all the way to a

spermatid. Spermatids, unlike primary spermatocytes, are true haploid reproductive cells. Their development stops before they elongate and shed much of the cytoplasm to resemble normal spermatozoa (Weedin et al. 2008; Walker 2010). Some spermatogonia in the LMA case will have a chance to mature into spermatozoa, or at least a usable form of haploid spermatid, therefore viable sperm retrieval is possible in the LMA case. Spematogonia in the EMA case never progress to a viable spermatid so these patients are not candidates for sperm retrieval treatments. Figure 4 shows the progression of spermatogenesis from spermatogonia to spermatozoa.

E. Sperm Extraction Techniques

Surgical extraction of sperm for was first proposed as a treatment for azoospermia in 1985 (Temple-Smith et al. 1985). The technique proposed, microsurgical epididymal sperm aspiration (MESA), required the exteriorization of the testis and epididymis and harvesting of the fluid contained in the epididymis via blunted needle or a silicone tube. A percutaneous version of MESA, percutaneous epididymal sperm aspiration (PESA), was proposed by Craft and Shrivastav in 1994 for use in OA cases (Craft et al. 1995).

Testicular Sperm Extraction (TESE) describes a procedure to harvest tissue directly from the testis after exteriorization of the seminiferous tubules under naked eye observation and is viable for both OA and ASD treatment. This procedure was first used to harvest sperm from ASD patients for ICSI in 1995 and produced viable amounts of sperm in 13 of 15 cases. Of the 182 oocytes injected, 65.5% became viable embryos and were transferred or frozen. The resulting embryos were implanted normally in 18% of cases, similar to the implantation rate of ejaculated spermatozoa (Devroey et al. 1995).

Procedure	Sampling Location	Average SRR for ASD
MESA	Epididymis (aspiration)	35%
TESA	Testis (Aspiration)	10 – 30%
PESA	Epididymis (Aspiration)	35%
TESE	Testis (tubule extraction)	49.5%
MicroTESE	Testis (tubule extraction)	60%

Table 1 Sperm retrieval procedures and associated success rates

Figure 5. The process of MicroTESE - 1. the testis is exteriorized and the tunica albuginia is medially bisected exposing the seminiferous tubules. 2. Under a microscope, the widest and most opaque tubules are identified for extraction at 25x magnification. 3. The extracted tissue is dissected and examined for visible sperm under a compound microscope at 100x magnification. (Reproduced with permission) (Esteves 2015)

In 1999 Dr. Peter Schlegel proposed a modification to the TESE procedure that increased the likelihood of sperm retrieval in ASD cases while decreasing the amount of extracted testicular tissue, thereby limiting testicular trauma. The procedure, named MicroTESE, employed a microscope to assist the surgeon's acquisition of enlarged (or dilated) and opaque seminiferous tubules. Dilated and opaque tubules were shown to have a 100% higher likelihood of sperm presence than were atrophied and transparent tubules. In addition, an average of 9.4 mg per microdissection sample produced 160,000 ± 270,000 sperm compared to 64,000 ± 52,000 sperm for standard biopsies that harvested an average of 720 mg of testicular tissue. This represents an average 191.5 times greater sperm yield per mg for microdissection than traditional biopsy (Schlegel 1999).

F. Learning Curve of Sperm Extraction Procedures

For MicroTESE, success rate and surgical time varies greatly based on surgeon experience with the procedure. In a 2010 study on the topic, Ishikawa et al. suggested novice surgeons are 12% less successful at retrieving sperm and 25 minutes slower to complete the MicroTESE procedure than expert surgeons (Ishikawa et al. 2010). The sperm retrieval rate (SRR) in the first 50 cases was 32% and increased by 12% in the following 50 cases. The final 50 cases saw only a 4% increase in SRR demonstrating diminishing returns for SRR related to experience.

A larger cohort study of MicroTESE tracked a single urologist through 9 years and 792 procedures on men with ASD. This study produced an average SRR of 60%, which may represent an asymptote for MicroTESE success for expert surgeons (Ramasamy et al. 2009). Figure 6 shows the relationship between surgeon experience and SRR, the data is fit by a logarithmic curve with an R^2 = 0.943 representing the steep learning curve and decreased performance gain with experience after the first 100 cases.

Another study by Ramasamy et al. demonstrated the relationship between duration of a MicroTESE case and the likelihood of sperm retrieval and subsequent live birth. The results of the

study showed the highest likelihood for sperm retrieval occurred in cases that lasted for 2 hours or less, 88.6% reported SRR. Cases that lasted longer than 2 hours reported an average SRR of less than half of that, averaging 30.7%. The mix of ASD pathologies was not a statistically significant factor in the SRR. The study concluded that there was “no time after which MicroTESE became universally unsuccessful.” (Ramasamy et al. 2011) The study noted the first 120 cases performed by the surgeon were excluded based on the findings by Ishikawa et al. about MicroTESE learning curves.

This data indicates the learning curve for MicroTESE compared to other sperm retrieval techniques is more demanding for the novice surgeon. Unlike percutaneous or epididymal aspiration methods, microdissection requires objective analysis of the seminiferous tubules for differentiating features prior to sampling. Furthermore, the more obvious the features are to the surgeon, the more quickly he or she is able to harvest sperm from the testis. However, even in testes that do not have obvious visual indications of sperm presence (cases that require more than 2 hours operative time), sperm retrieval success is possible in at least 30% of patients. The common limiting factor in both cases is the surgeon’s ability to visually identify sperm presence in testes where sperm exists.

G. Proposal

The proposed solution to the problem of tubule identification is a programmatic optical model to identify sperm locations consistently thereby reducing the surgeon learning curve and increasing sperm yield.

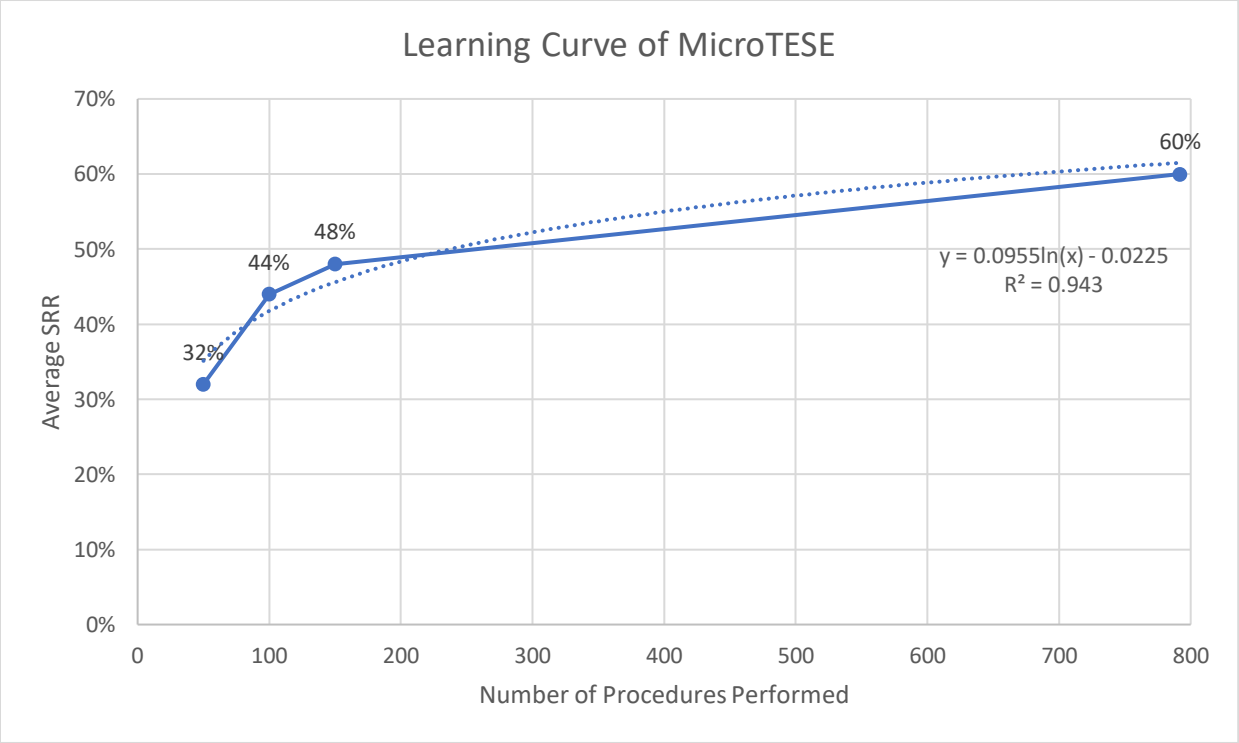


Figure 6 - Learning curve of MicroTESE modeled as a Logarithmic curve.

II. Technical Background

A. Traditional Computer Vision

The first step to developing a system to computationally identify objects based on specific optical characteristics was established by P.V.C. Hough in 1962. Hough's algorithm identified a method to find straight lines between coplanar points in an image. In 1972, Duda and Hart optimized this algorithm for computational efficiency and to apply it to fit generalized curves in an image (Duda and Hart 1972). This led to the development of the Hough Transform, a method to identify arbitrary shapes based on parameterizing the border of the shape (Ballard 1981). This method allowed the hard coding of specific shape information for on-line object detection; however, it was not robust enough to identify the same object from different perspectives or with scalar/affine transformations applied. The next leap forward came in 2004 with the introduction of the Scale Invariant Feature Transform (SIFT) algorithm (Low 2004). SIFT created a way for image objects to be identified regardless of affine transformations based on their footprint in the gaussian scale-space. The object is identified by local extrema computed for a 3x3 window in the current and adjacent scales of the Gaussian filter function. These local extrema are the key points used for object detection. If the footprint of the object (the signature of the key points) is known, the object should be identifiable. An alternative to SIFT was proposed in 2006 by Leonardis et al. labeled Speeded Up Robust Features (SURF) (Leonardis et al. 2006). This algorithm used the Hessian matrix instead of the Laplacian (the sum of the Hessian matrix eigenvalues) to generate key points. The difference yields a greater robustness of features with less overall dimensions for a clustering algorithm to sift through (64 for SURF vs 128 for SIFT).

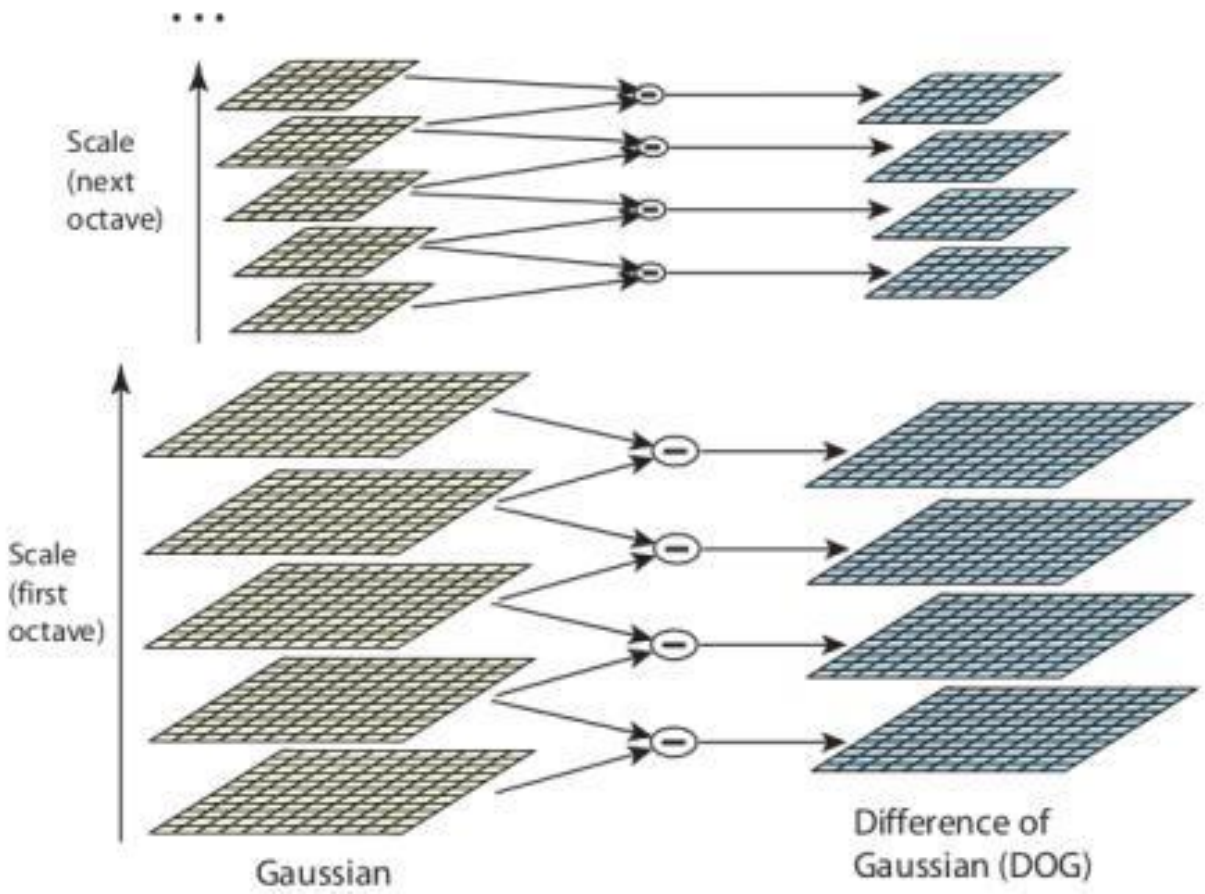


Figure 7 Sift Algorithm - Difference of Gaussian calculation is used to identify significant key points in the image

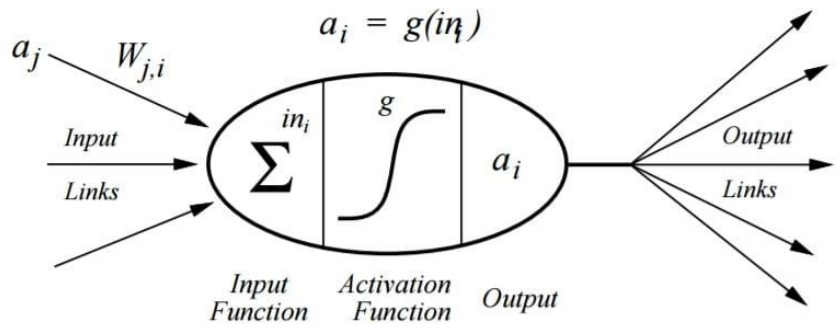
B. The Perceptron

One of the basic elements in the field of machine learning is the perceptron. First published in 1958, the perceptron was designed as an electromechanical neuron with 3 parts to simulate the process of memory and organization in the brain. The input function of the perceptron brought in a signal (or many signals) to be processed and summed the inputs into a single signal to pass to the next part. The activation function was originally a simple step function designed to block empty signals (zeroes) and pass live signals (ones) but can be any waveform. Figure 3 shows a perceptron with a sigmoid activation function. The final part is the output: signals that are passed through the activation function can be linearly modified before being passed out. The perceptron was imagined as a versatile gate that could be programmed in any useful way, much like a neuron (Rosenblatt 1958). The combining of these individual units led to the multilevel perceptron, and eventually would go on to become the model for the individual node in the modern neural network.

C. Backpropagating Errors and the Rise of Convolutional Neural Networks

To identify an object using the hardcoded algorithms of traditional computer vision, the target object had to be known and parameterized accurately. If the object complexity was too high or the imaging perspective obscured important details about the object, it simply could not be identified. This problem had been solved for millennia in the biological domain, but firmly evaded the grasp of mechanical computation. In his 1986 paper, Geoffrey Hinton et al. proposed a method of back-propagation of errors to learn patterns between an input and output set (Rumelhart, Hinton, and Williams 1986). This method was inspired by the signal cascade between interconnected neurons and the learned efficiency of interconnections resulting from action repetition. The input signal is first passed through, and manipulated by, uniformly weighted nodes towards an output node. Since

the output is known, the difference between the input and output signal can be calculated. The difference, or error, at the output node is traversed backwards through the network, effecting every node by a magnitude proportional to the error. Using this method on a net of perceptrons until the error reaches zero trains the system to convert the known input to a known output. The backpropagation algorithm paired with the perceptron appropriately modeled the neural pathways related to human perception of images. However, an additional component was required to model the ocular photoreceptors to turn a complex image into signals that could be manipulated and analyzed by the neural network.



$$a_i = g(\sum_j W_{j,i} a_j)$$

Figure 8 A basic perceptron, modeled here with a sigmoid activation function

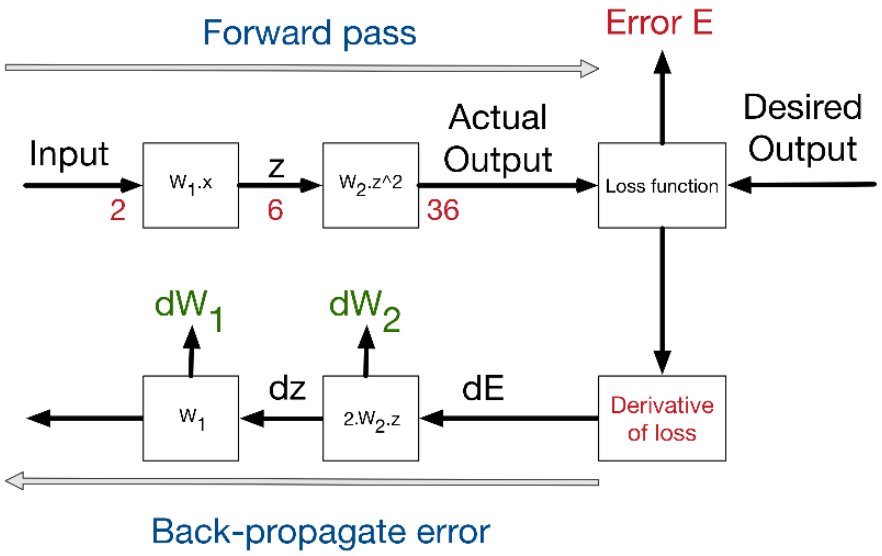


Figure 9 Backpropagation of error as proposed by Hinton et al.

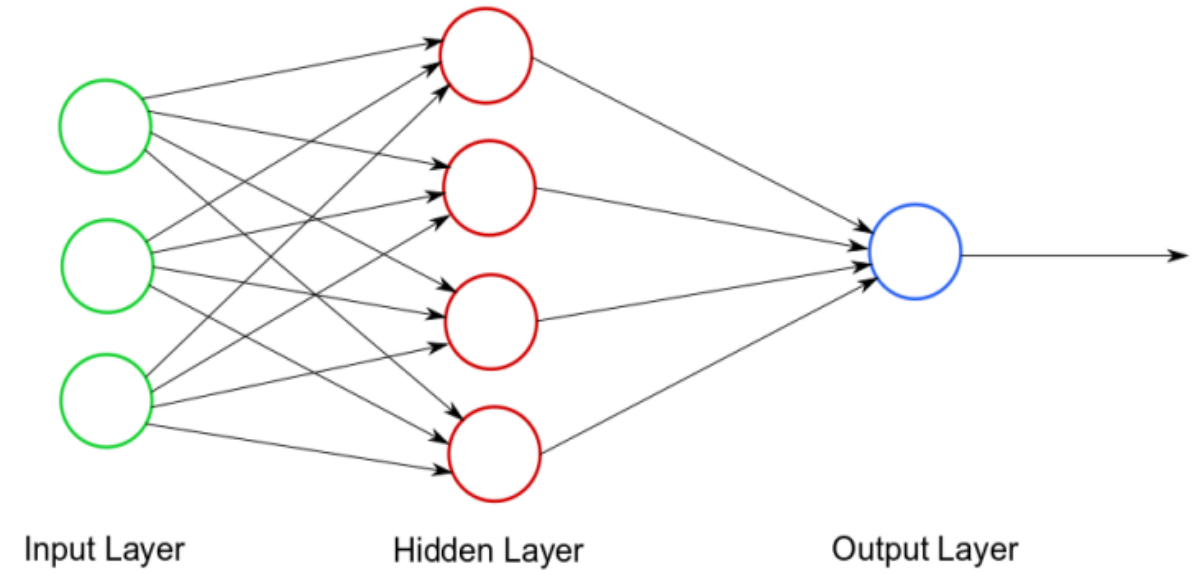
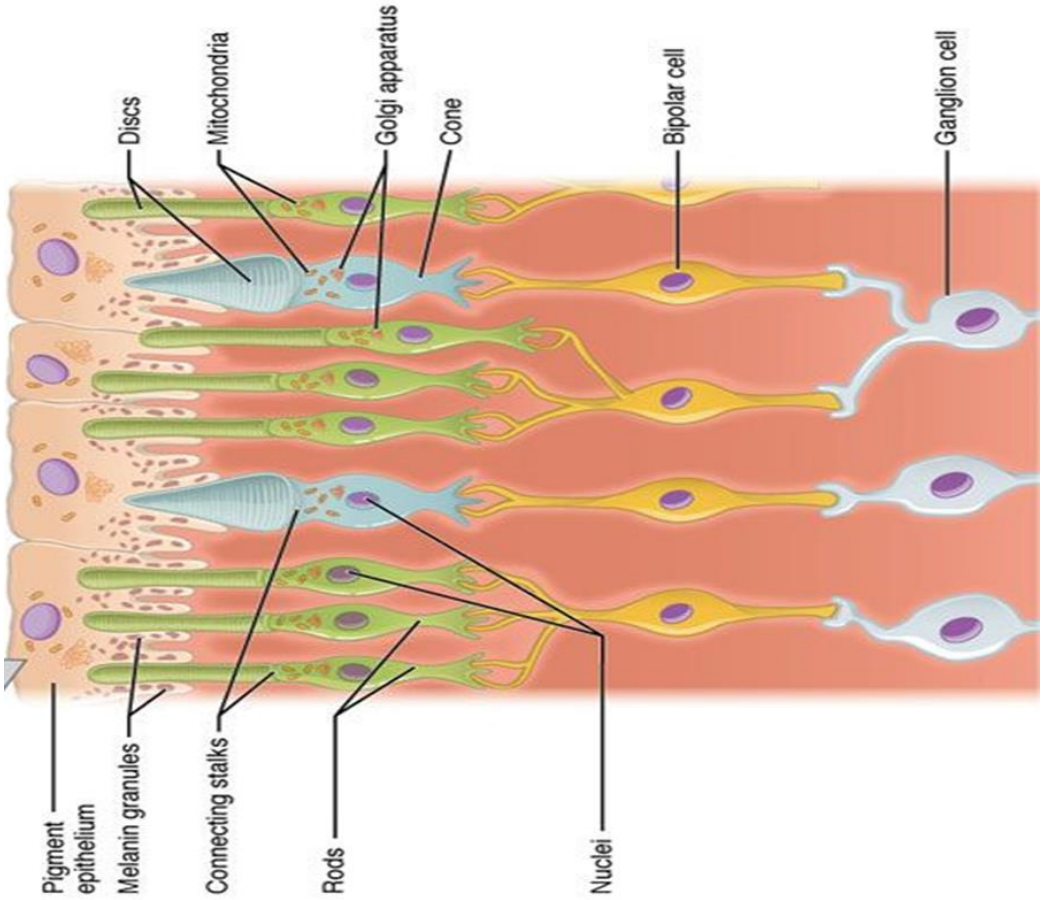


Figure 10 (top) A simple schematic of a neural network with 1 hidden layer (bottom) a schematic of the human retinal system



In 2012, Hinton's student Alex Krizhevsky published a paper that implemented the idea of multi-layer convolutions paired with a back propagating neural network to identify features of objects in images, and another network layer to classify images into classes (Krizhevsky, Sutskever, and Hinton 2012). The advantages of this algorithm are the automation of the feature detector and the separate object classification that can adjust internal weights agnostic to the feature detector. This allows for more abstract representations to be utilized by the neural network to analyze an image more accurately. This characteristic to identify abstract representations can evolve into overtraining, where a feature extractor and classifier pair identify labels not related to the classification task. An example of this is demonstrated by Song et al. (2019) in an analysis of a binary gender facial recognition classifier. The classifier learned to differentiate faces based on race without race being a stated evaluation metric (Song and Shmatikov 2019). In many situations this ability is considered a disadvantage of non-tailored feature extraction, however, in applications where specific metrics of measurement cannot be defined (the ideal size of a sperm laden seminiferous tubule for example) abstractions arising from relaxed training goals may provide an accuracy advantage.

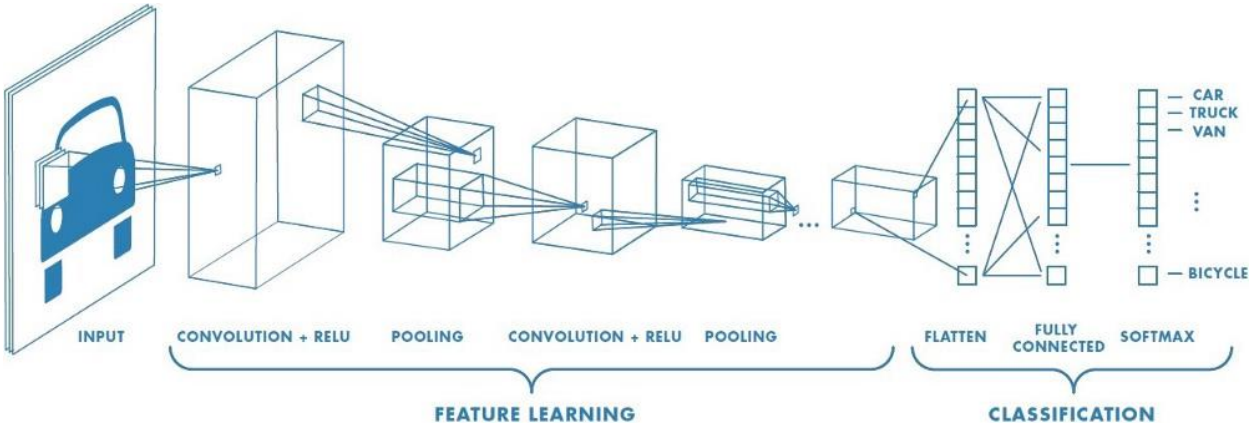


Figure 11 A schematic of AlexNET, the first convolutional neural network

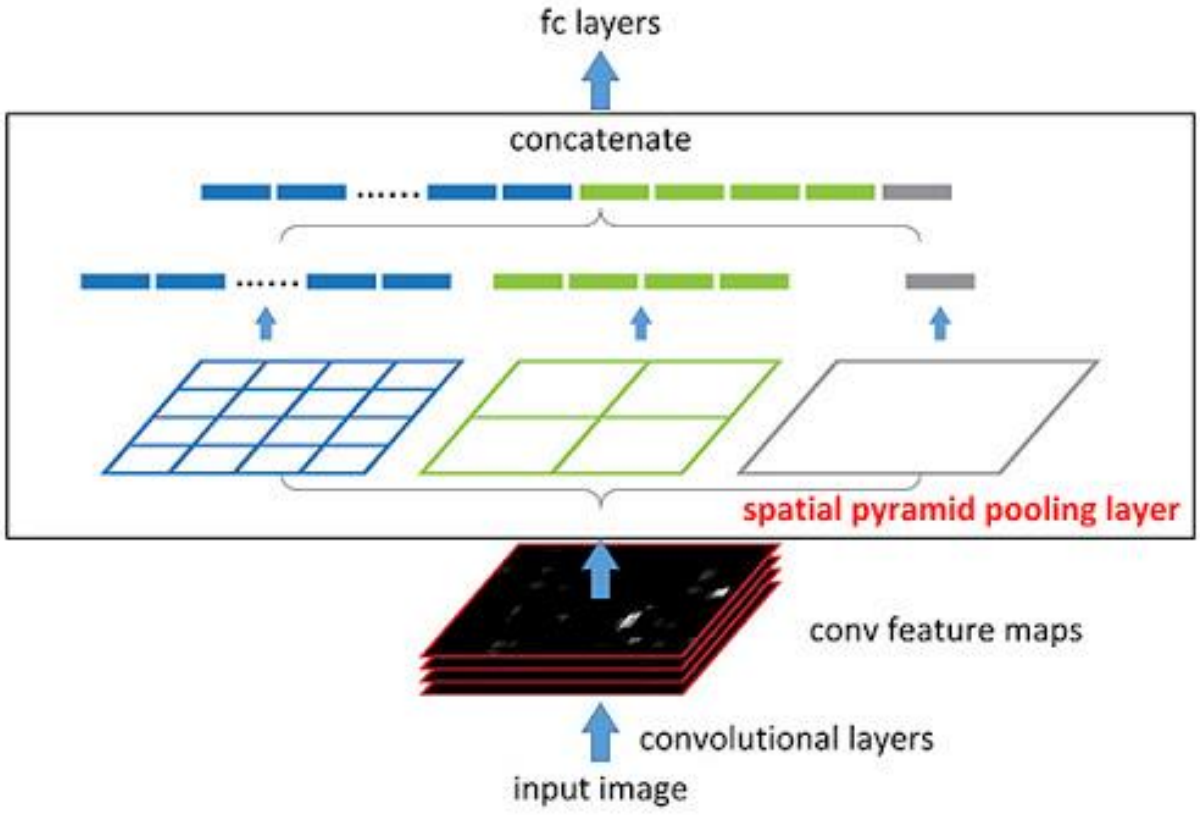


Figure 12 Spatial Pyramid Pooling

D. Feature Pyramids – PSPNet

To learn small scale features, a traditional convolutional structure is powerful. However, to learn scale invariant features with context, the neural network must be able to link relevant features from the smallest convolutional window with those from the largest window. Investigation in this method of feature identification was popularized in 1980 in a book by Tianmoto and Klinger. The concept was improved upon by Rosenfeld and Sher in their 1998 paper (Rosenfeld and Sher 1998). They described a method to identify objects (blobs) within a background using different levels of features starting from very fine (5x5 pixels) to very coarse (32x32 pixels). Each level provides a perspective that other levels may miss, but by compositing the different levels, new insight can be identified in an image that would have been otherwise missed. This concept is known as spatial pyramid pooling (SPP), and has been a key component of modern computer vision architectures (He et al. 2014). The application of multi-scalar feature detection is already present in the convolutional architecture of Alexnet, however SPP works to strengthen the connections between each feature scale by accounting for all scales at once before passing to the final fully connected layer as shown in figure 12 (He et al. 2014).

E. Region Based Convolutional Neural Networks – R-CNN to Faster R-CNN

The feature pyramid method has some key advantages for object segmentation, chief among which is the integration of bulk features with fine detail to generate a more complete picture of the specific image region. A region based convolutional neural network (R-CNN) uses a selective search process to identify 2000 potential regions of interest (known as “proposal regions”) (Girshick et al. 2014). The output of this network is then evaluated by a deep neural network for content. The identified characteristics are collected by higher dimensional network layers to identify the object

and localize it in the image using a bounding box. This method can be time consuming as each proposal region goes through the deep neural network layers.

This R-CNN was superseded by Fast R-CNN, which inserted a simple convolutional stage before the selective search step to generate a lower dimensional feature map that allowed for a simpler search for ROI (Girshick 2015). Fast R-CNN gave way to Faster R-CNN which replaced the selective search process altogether with an adjacent neural network coined the region proposal network (RPN). This less computationally intensive neural network is tailored to identify the presence or absence of an object against a background. This network passes back many proposal regions that are pooled to create a much more definite region for the following CNN to classify (Ren et al. 2017).

F. Mask RCNN

The most recent evolution in the R-CNN line of networks is Mask R-CNN. Published in 2017, Mask R-CNN improves on the individual object segmentation capabilities of Faster RCNN by incorporating an “ROIAlign” layer tangential to the main convolutional neural network. This layer works parallel to the classification and bounding box layers to generate an individual pixelwise binary mask for each ROI. The weight of this mask is integrated with the weight of the classification to prioritize objects in the foreground versus background. As the mask is generated on a pixelwise basis, the resulting segmentation can be as fine as the input image, scaling in definition with the image resolution.

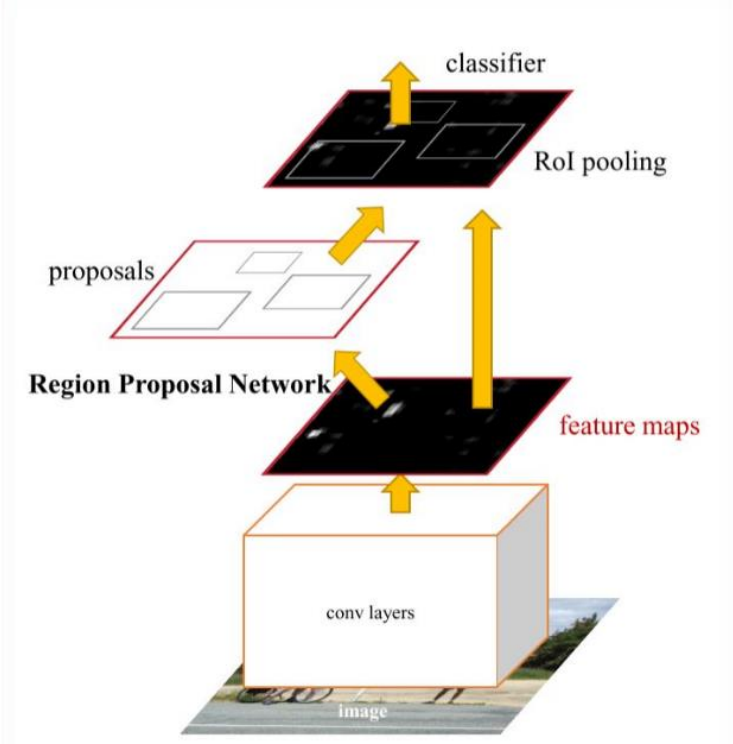


Figure 13 Faster RCNN network schematic

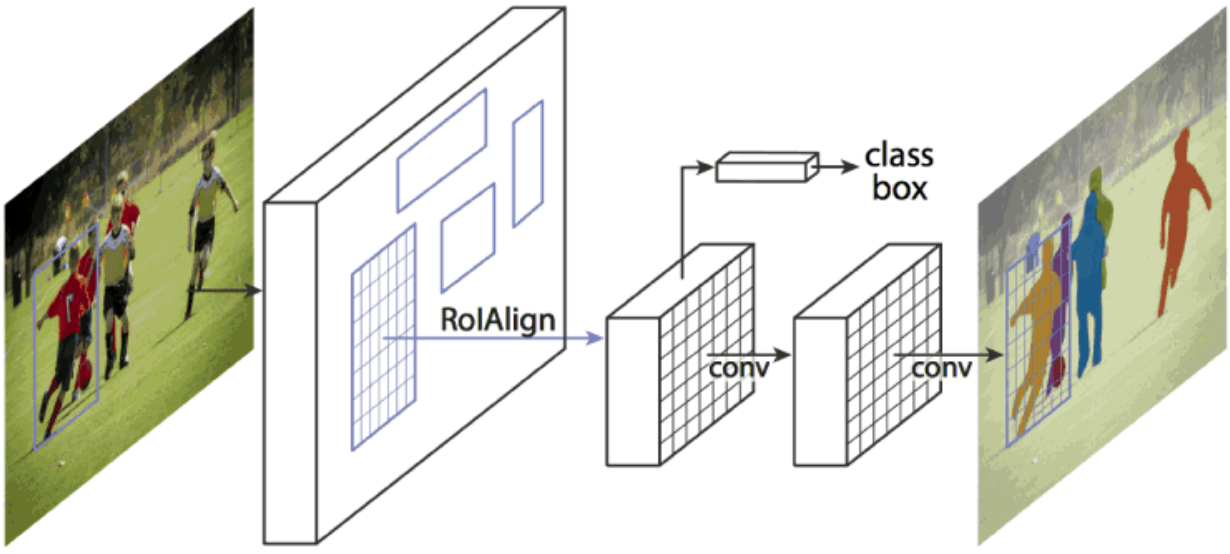


Figure 14 Mask R-CNN ROIAlign branch

III. Section Summary: Proposed Solution

A. Identifying the problem

MicroTESE requires a thorough examination of all seminiferous tubules in the testis. Granular analysis of each section of testis requires a longer operative time. The learning curve for this procedure is steep and requires hundreds of procedures for each surgeon to achieve proficiency in identifying sampling locations for sperm in the testis.

B. Proposed Solution

A neural network can be trained from seminiferous tubule images taken during a MicroTESE procedure and the corresponding sperm counts from each biopsy site. Once trained, the neural network can analyze images from the surgical microscope during biopsy and generate a probability map of sperm locations overlaid with the operative image.

C. Experimental Methodology

To train this neural network, an animal model of azoospermia due to spermatogenic dysfunction must be created. The images and sperm counts obtained while performing MicroTESE on the first group of treated animals was used to train the neural network. The trained neural network was used to identify sampling locations on a second group of study animals. The performance of the neural network probability map for sperm locations was evaluated using the sperm counts obtained from the biopsy sites.

IV. Clinical Methods

A. Introduction

To acquire training data for the neural network, the experimental setup was designed to mimic the MicroTESE procedure as closely as possible. The variation between experimental animals was limited by strictly defining age and weight. Variations in lighting and extraneous environmental conditions were mitigated by performing all procedures in the same dissection room.

B. Species Selection

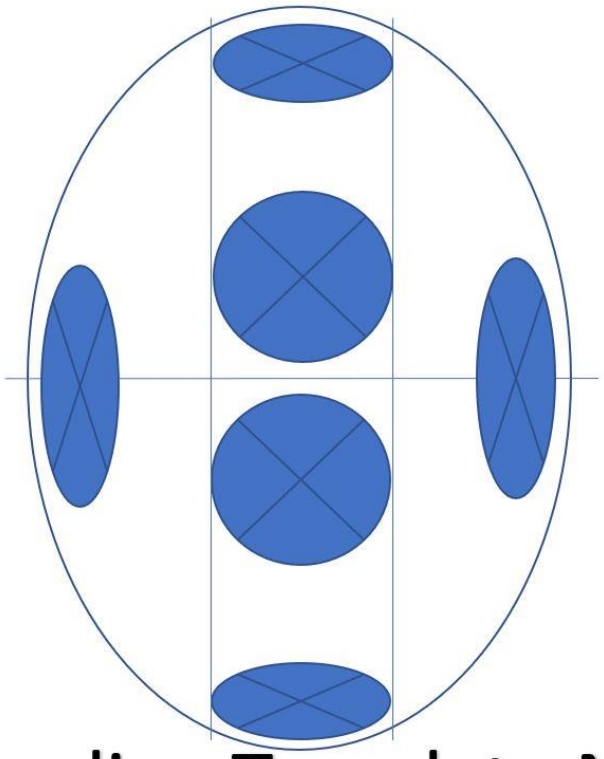
To generate a dataset from which to train a neural network, an animal model for the human testis was required. A healthy 20 cm long Sprague-Dawley rat's testes measure 2.5 cm in length on average compared to an average of 4 cm length for the average human male (approximately 168 cm tall). The body to testis ratio is 0.125 and 0.023 for the rat and human respectively. The large testes of the rodent compared to the body size allows for more samples to be taken per animal, resulting in fewer animals necessary for significant data. Furthermore, the sizes of seminiferous tubules in the rat and the human testes are similar in the wild type. Aside from a lower number of tubules in the rodent model, it approximates its human counterpart in tubule diameter, averaging 208 µm, and total length of functioning tubule. The human azoospermia due to spermatogenic dysfunction (ASD) case presents as a non-uniform atrophy of the seminiferous tubules. To model this, two distinct treatment vectors were considered.

C. Surgical Setup

An Amscope 7x – 45x trinocular stereo microscope with a camera mounting channel was used for the experiment. An additional 2x magnification lens was attached to the microscope. An Amscope MU1803 4K digital microscope camera was chosen to record biopsies.

D. Biopsy Procedure

Each animal was anesthetized using a Ketamine/Xylazine solution and the testes were shaved prior to testicular biopsy. The testes were under a surgical microscope at 35x – 50x depending on testis size. Six samples were taken from each testis and suspended in PBS solution in a 96 microwell plate (Fisher Scientific). Sampling was done using a template map to ensure uniform sample distribution. One sample was taken from each hemisphere, from each pole, and two were taken from the equatorial region of the testis. The procedure was recorded using a 4K surgical microscope mounted imaging system. Post biopsy, animals were sacrificed using a guillotine while under anesthesia.



Sampling Template Map

Label	Sperm Count Range (sperm/mL)	Sperm Count Range (sperm/sample)
Sperm	60K – 3M	2+
No Sperm	0 – 30K	0-1

Table 2 Tissue sperm count stratification

E. Biopsy Tissue Processing

Tissue samples were dissected in a 12 well culture dish using two 18-gauge hypodermic needles modified to have blunted flat blade ends instead of points to allow for manipulation of the seminiferous tubules with reduced risk of tubule puncture or tearing. Sperm was extracted from individual seminiferous tubules using a sliding compression along the tubule to squeeze out the cells contained inside. The extracted cells were suspended in PBS solution (Vial A) and spun down at 3000 G for 15 minutes. The remaining seminiferous tubule tissue was resuspended in PBS solution (Vial B) and spun for the same force and time. Supernatant was removed and discarded from Vial A. The cell pellet remaining in Vial A was resuspended with the supernatant harvested from Vial B. This processed sample was transferred into a fresh 96 well plate for storage.

F. Sperm Cell Counting

A 5 µL aliquot of the processed sample was suspended within a cell counter device and was observed under phase contrast microscopy at 150X magnification using an Olympus microscope with an oil immersed objective lens. Visible sperm were counted on the measuring grid when a high enough quantity of sperm was available. For lower sperm quantity, a full field count was taken and extrapolated for sample sperm density.

G. Testosterone

Testosterone is secreted from Leydig cells in the testis. It is regulated by Luteinizing hormone (LH) and Follicle Stimulating hormone (FSH) produced by the anterior pituitary. LH stimulates the Lyedig cells to produce testosterone. FSH activates the cellular activity of the Sertoli cells to facilitate the development of germ cells into spermatozoa (Walker 2010). FSH and LH are released in response to Gonadotropin Releasing Hormone (GnRH) from the hypothalamus (Crosnoe et al. 2013). Excessive exogenous testosterone can downregulate both FSH/LH as well as GnRH production.

Testosterone supplementation can be used to induce the inhibition of FSH, LH, and GnRH, resulting in a disruption of the spermatogenic process (Walker 2010).

Exogenous testosterone mediated spermatogenic dysfunction in the rat model presents as a decrease in testis volume and a significant decrease in sperm count. Early studies on the use of subcutaneous testosterone implants as a male contraceptive tested such vectors on Sprague-Dawley rats. A range of eluting implant sizes were tested and the ideal size (corresponding to dosage of testosterone) was 4.0 cm length, 1.98 mm id. Rats in this test group presented with a testis weight reduction of 50% when compared to the control. Epididymal sperm count showed a 100X decrease in sperm count when compared to the control. The number of impregnated females was correspondingly low, and only 10 live births were recorded in this study group compared to 221 in the control group (Hales 1984).

Bhul et al. investigated the effects of exogenous testosterone administered via eluting implant on seminiferous tubule morphology. Their work suggested the size of the tubules would decrease by around 50% while spermatogenesis was arrested at the spermatocyte stage. (Bhul et al. 1982)

H. Testosterone Experiment

Adapted from the protocol described by Bhul et al. 1982, this treatment utilized testosterone to induce azoospermia in the experimental rat group. Dow corning silastic tubing (#602-305) was used to create four-centimeter implants with an id of 1.98 millimeters and an od of 3.18 millimeters. Crystalized testosterone was packed into each implant using a vacuum packing method. Glass beads acted as caps at the ends of the implant to ensure implant integrity when sealing the implant with silicone sealant. Each implant was immersed in ethanol for 12 hours in a slow-moving plate rocker to ensure the implants were correctly sealed and did not allow ethanol to permeate the membrane. Implants failing the integrity test were discarded, the remaining implants were dried and stored for

use. A second set of empty implants were created as a placebo using the silastic tubing, glass beads, and sealant, but no hormonal contents.

The experiment was conducted on 2 test groups of rats with treatment starting while all animals were 6 weeks of age and between 325 – 375 grams bodyweight. The youngest rats of adequate size were chosen to begin treatment as close to the onset of puberty (weeks 3-5) while ensuring the flank of the animal was large enough for the implant. The first group contained 8 rats, of which 4 were in the experimental-testosterone group and 4 were in the non-hormonal placebo group. Labeled as treatment group A, the rats were surgically implanted with their corresponding hormonal or placebo implants into the subcutaneous tissue layer in the rear flank area under Ketamine-Xylazine anesthesia and Buprenorphine analgesia. Wounds were closed with staples and appropriate animal handling protocol was followed to ensure animal safety.

After an initial four-week incubation period, a pair of rats were selected from each group and sacrificed at regular one-week intervals. One control and one experimental rat were sacrificed on the same day using the same technique to limit study variance. Of the initial 8 rats in this block, one experimental rat was lost due to cage-mate aggression.

Treatment group B mirrored the initial experiment with eight animals, however, were divided as two control and six experimental. The initial cycle showed that the ideal biopsy window with appropriate induced-azoospermia was at the eight-week mark, therefore all eight rats in this group were sacrificed over a two-day period eight weeks from implant date.

I. Results of the Testosterone Experiment

Unfortunately, the effects of testosterone implantation on rat sperm production was insufficient. Experimental animals showed a testis size reduction by 50%, however seminiferous tubules were not atrophied as expected from a true azoospermic case. In addition, sperm

production, while significantly reduced, was not altogether inhibited. From 192 samples taken through the duration of the experiment, only one sample had zero sperm present during cell counting. Following the second testosterone study group, busulfan was chosen as the next vector to induce azoospermia in the rat model.

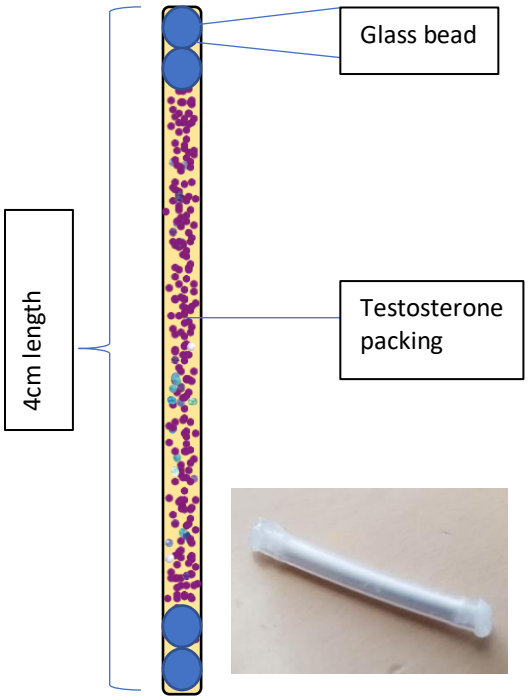


Figure 15 Schematic of a testosterone implant (bottom right) Image of an implant.

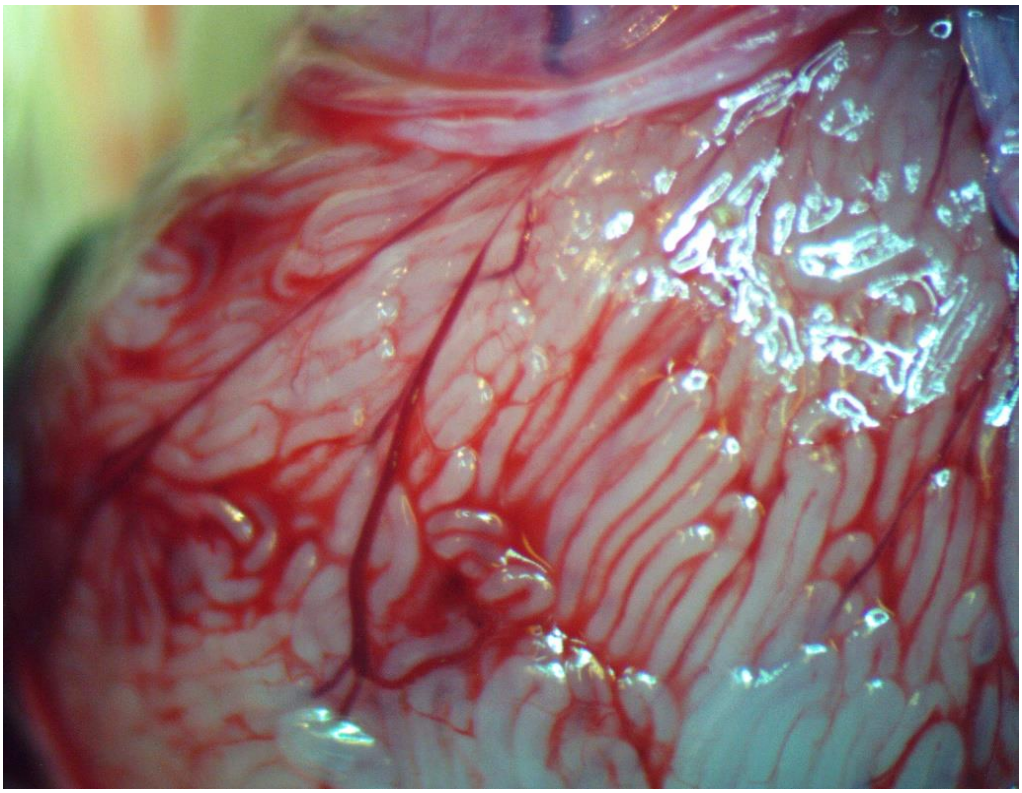


Figure 16 Seminiferous tubules of a testosterone treated rat

J. Busulfan

Busulfan is a chemotherapeutic agent used in the treatment of chronic myeloid leukemia. It is a highly active antineoplastic alkylating agent that binds to one of the DNA strands during cell division. Therefore, cells with high division rates are most susceptible to busulfan. This cytotoxic quality causes infertility in mice with relatively low doses. In the larger rat model, busulfan dosing for spermatogenic dysfunction requires carefully timed administration. To maximize testicular impact and reduce chances of fatality, Panahi et al. recommended administering two 10 mg/kg intra peritoneal injections spaced 21 days apart instead of one larger dose (Panahi et al. 2015).

K. Busulfan Experiment

The Busulfan treatment protocol was based on the study conducted by Panahi et al. 2015. In this experiment, three treatment groups were conducted with rats starting treatment at 4 weeks of age and 85 – 115 grams bodyweight. As no implant was required, the rat age was chosen to most closely coincide with the beginning of puberty in the male rat. The first two groups, treatment group C and group D, both contained eight animals, 2 control and 6 experimental. The rats in group C were injected with 15 mg/kg busulfan on two occasions 14 days apart. Sixty days after the first injection, all rats were sacrificed yielding treatment failure with insufficient induced azoospermia. The protocol was repeated for Block D with two injections 14 days apart, however, it was followed by a third injection after an additional 14 days. Unfortunately, the aggressive dosing measure led to the demise of four of the experimental rats. Treatment group E was comprised of 16 animals divided into 12 experimental and four placebos. The protocol was altered to increase the dosage of the Busulfan to 17.5 mg/kg, but to maintain only two injections 14 days apart. Animal sacrifice was completed 60 days post-first injection therapy with appropriate azoospermic presentation.

L. Results of the Busulfan Experiment

Busulfan proved to be an effective agent to disrupt spermatogenesis in the rat model. From the 96 samples taken from the first trial group, 27 samples showed 0 sperm during cell counting. All treatment animals also had at least one location where sperm was present showing the spermatogenesis was similar to the human case where loci of sperm remain in an otherwise barren testis.

	Testosterone		Busulfan		
	Group A	Group B	Group C	Group D	Group E
Total Animals	8	8	8	8	16
Controls	4	2	2	2	4
Treated	4	6	6	6	12
Dosage	4 cm implants		15 mg/kg x 2 injections		17.5 mg/kg x 2 injections
Incubation time	8 weeks		60 days		

Table 3 Experimental groups

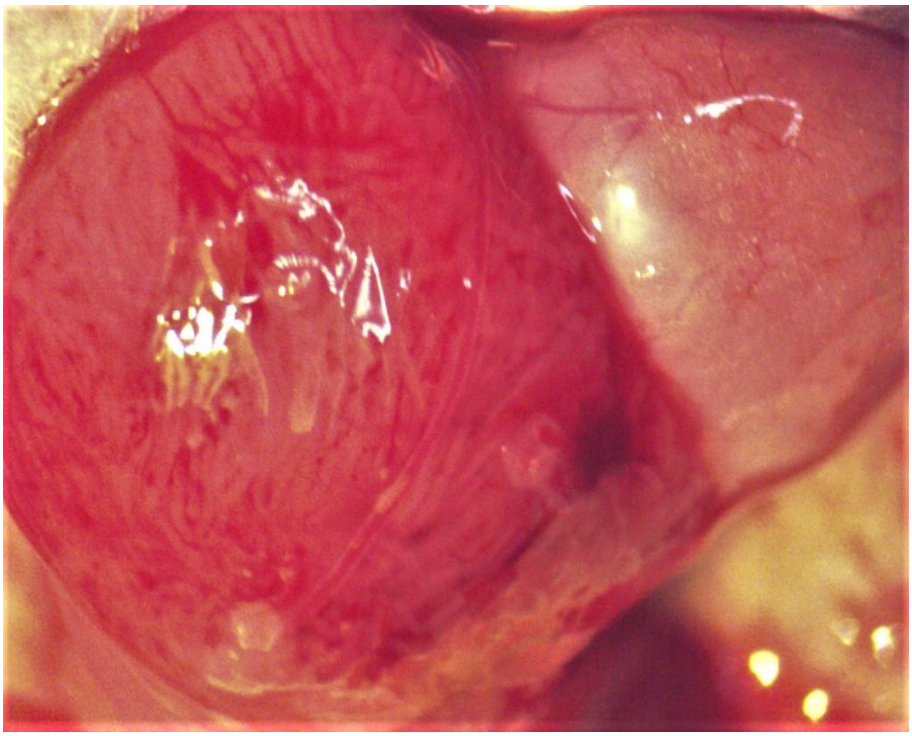


Figure 17 The seminiferous tubules of a busulfan treated rat

V.Computational Methods

A. Neural Network Training Data Creation

One pre-biopsy image was selected for each sample using the surgical video and was uploaded into the VGG Image Annotator (VIA) software for manual markup. The VIA software is a graphical tool used to create a label mask in which pixels are coded to a specific object category. The surgical video was used to identify exact sampling locations on the testis and a polygon was manually created by a human annotator around the sampling region. The polygon was labeled with the sperm count range (low, medium or high) identified during the cell counting stage. The annotated images and annotations were saved as two separate files: Images were exported as JPEG files while annotations were exported as JSON files containing polygon points. A monochromatic image (PNG) mask was generated from these points; the pixel value indicated the category of object - 1 coded for low sperm, 2 for medium, and 3 for high sperm. A “0” value pixel value coded for background regions.

Each image pair was augmented using skew, affine transformation, rotation, cropping, and blurring to increase the dataset size. From an initial count of 189 images for Sperm Hunter V1, a training set containing 5000 images was created.

B. Neural Network Training Procedure

One hundred of the original images were extracted and mirrored to produce the model evaluation set. The remaining training images were further separated into a training and internal validation set with a ratio of 80:20 to allow for enough variation in the internal validation set. A Mask RCNN model was trained in python using this data on a single consumer grade PC containing a quad-core processor and an NVIDIA GTX 1070 with 6 GB GPU ram and 16 GB system ram. Training time was 20-26 hours per model. The GPU’s lack of memory contributed to the extended training

times. The Amazon web service (AWS) cloud was evaluated as a faster training platform. A PSP-Net model was trained in python on an AWS cloud compute instance featuring an 8-core CPU, 1 Nvidia V100 GPU with 16 GB GPU ram, and 61 GB system ram. The larger enterprise class GPU was able to store more images in memory at once and reduced the end-to-end training time to under 3 hours. A more powerful instance containing a 16-core CPU, and 4 Nvidia V100 GPUs was evaluated but did not produce any appreciable improvement in training time. Models were trained in an average of 2.5 hours on both node types. The primary metrics for model optimization were Intersection over Union (IoU) and Mean Square Error (MSE) loss. While the testosterone experiment did not produce complete azoospermia in the animal model, the images produced from the trial did produce a large range for the control case (seminiferous tubules with sperm) and therefore were still useful as training images.

C. Neural Network Evaluation Procedure

The images set aside for evaluation were uploaded to the AWS cloud for inference by the trained models. Each sampling location predicted by the Sperm Hunter model was evaluated as if it was a real biopsy sample. The prediction location was compared to all known sperm sampling locations on the same testis as the evaluation image. Each prediction could be either a true positive, a false positive, or an unknown. If the predicted location overlapped any region that was known to be sperm positive, it was awarded a true positive prediction. The unknown predictions were recorded and discarded from the statistical analysis as the ground truth sperm count at the corresponding locations was unavailable. If the model did not predict a sampling location for evaluation images containing a low sperm count location, the model was awarded a true negative prediction. Vice versa, if the model failed to identify a sperm positive location on an evaluation image with a known sperm positive location, it was penalized with a false negative prediction. An example of the evaluation of a single image is shown in figure 17. The red dashed region indicates a

known sperm negative location; The green dashed region indicates a sperm dense location. The model's prediction is shown as a green highlight on the original image (bottom of figure 17.)

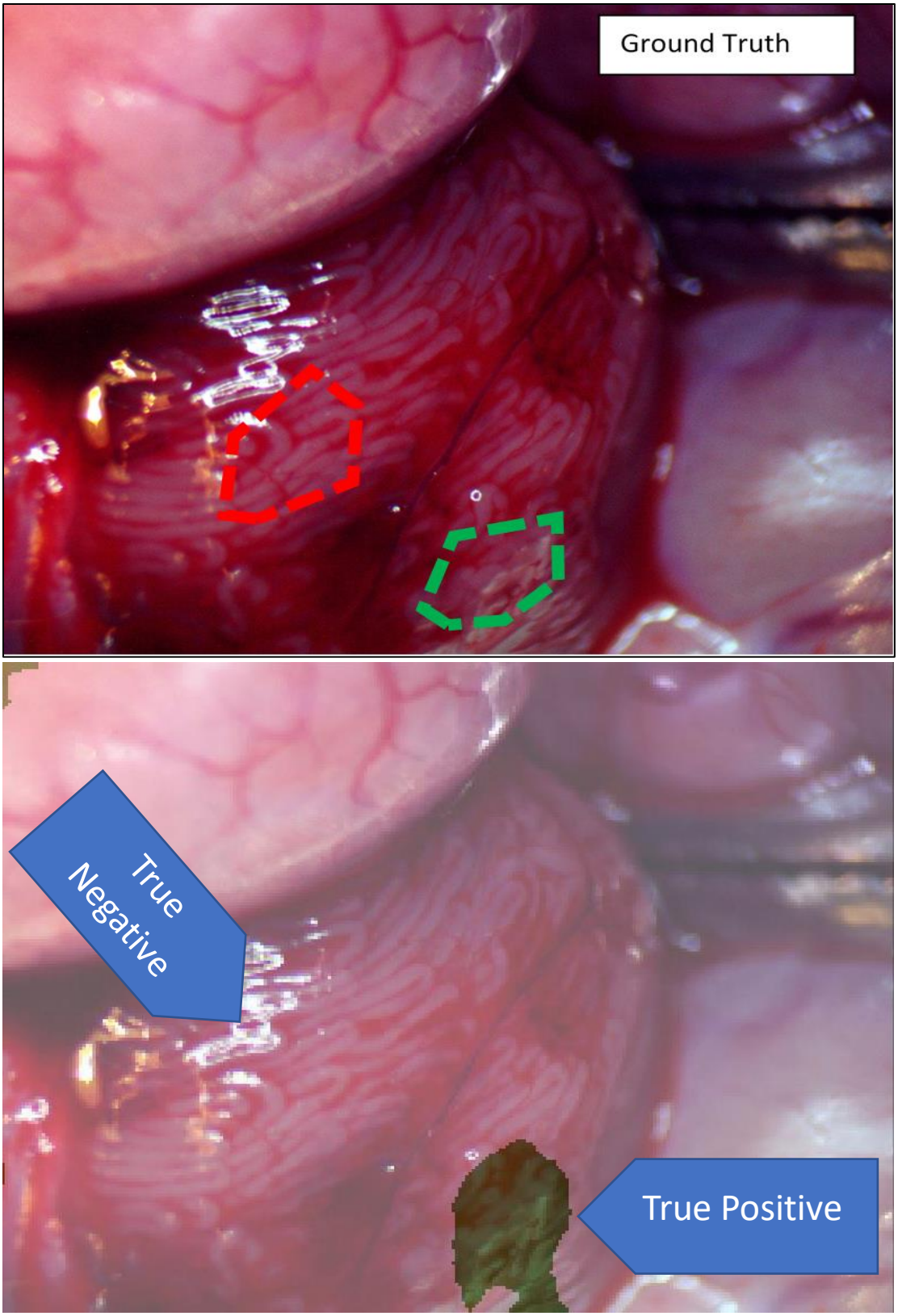


Figure 18. An evaluation of predicted region. The ground truth image (top) indicates the sperm positive (green) and sperm negative (red) regions. The model prediction (bottom) shows the model predicting sperm (green highlight) on top of the known sperm positive region. This image is scored with one true positive, and one true negative prediction.

VI. Model Performance

A. Sperm Hunter Prime

In the final iteration of the Sperm Hunter model the following metrics were recorded from an evaluation image set (n=100) comprised of randomly selected images removed from the training set. The prediction classes and validation metrics are shown in table 4. The receiver-operator plot of Prime is shown in figure 18 including 2 prior iterations of the model (V1.1 and V1.3).

B. Predictions During Live Surgery

To evaluate model performance using real world conditions 24 further animals were treated with busulfan to induce spermatogenic dysfunction. After exposing the testis, an image of the seminiferous tubules was taken with a camera attached to the surgical microscope. Sampling locations were predicted by Sperm Hunter Prime and 3 samples were taken in sequence. Two samples were taken from locations indicated to have a high probability of sperm, and one was taken from a region not indicated. After sperm counting, an optimal dataset was assembled from testes containing at least 1 sample completely devoid of sperm. From these highly affected animals, 60 samples were identified for final model validation. The outcome of pre-sampling prediction as well as post sampling individual image predictions are shown in figure 20. The difference between the live sampling and post sampling was attributed to blood obscuring the seminiferous tubules after the initial sample was taken.

Sperm Hunter Prime Confusion matrix		Model Prediction: Sperm	Model Prediction: NO sperm	
Sperm Found in biopsy	Yes	86	24	**110**
	No	21	55	**76**
		107	**79**	

Table 4 Confusion matrix for Sperm Hunter Prime

Selected Validation Metrics	
Sensitivity	78%
Specificity	72%
Accuracy	76%
Precision	80%
Diagnostics Odds Ratio	9.38
Chi Square Metric	47.006

Table 5Evaluation metrics for Sperm Hunter Prime

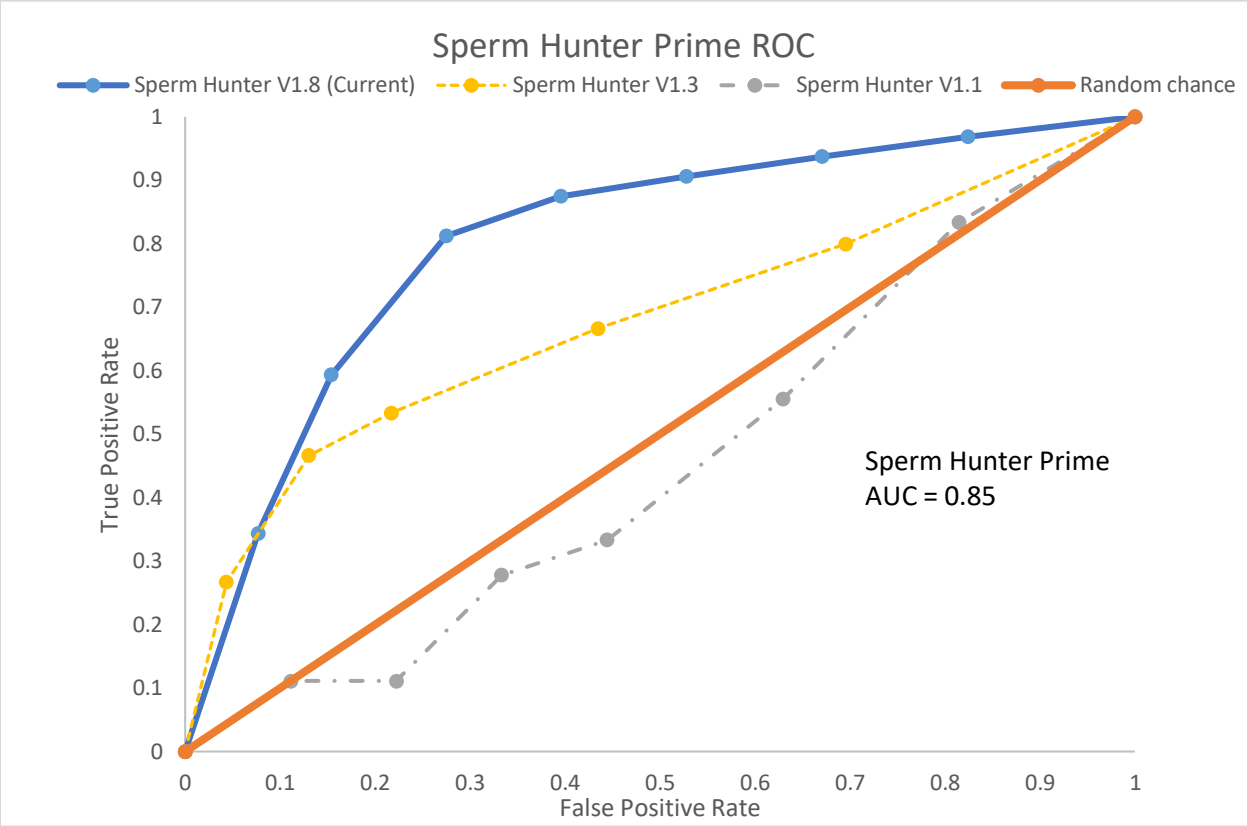


Figure 19 ROC plot of predicted sperm locations on validation images

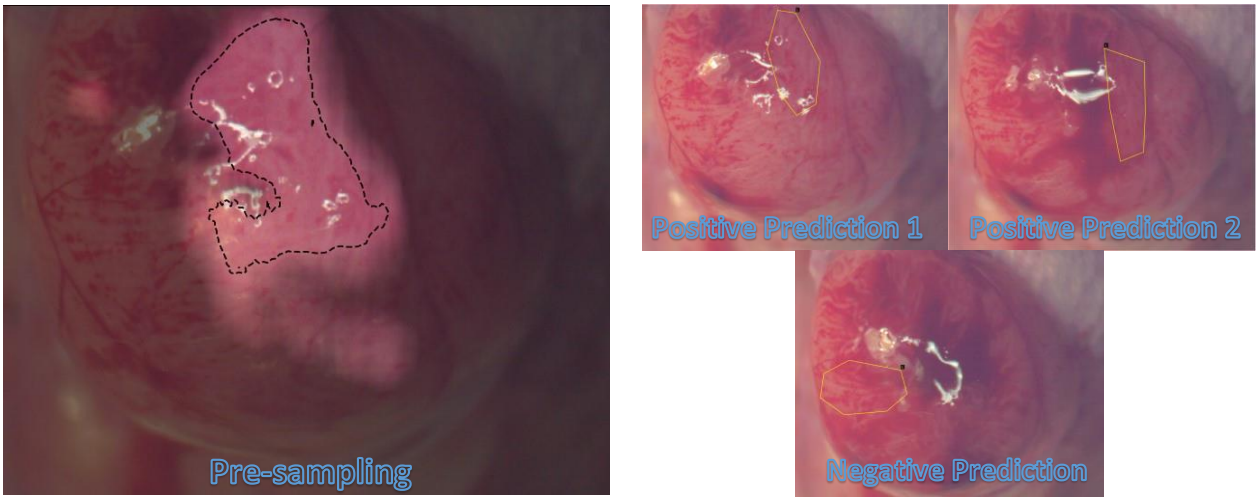


Figure 20 Pre-sampling prediction on the left with the corresponding sampled locations on the right

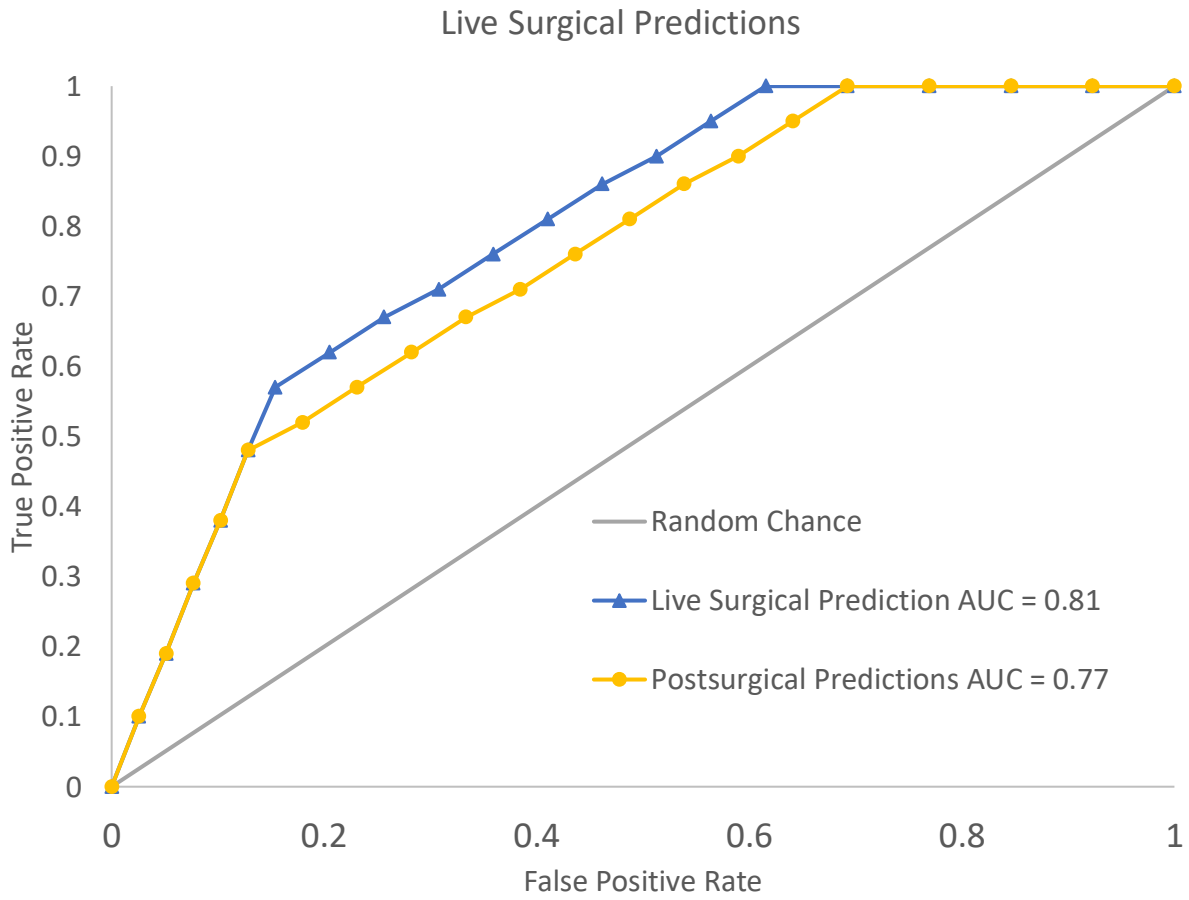


Figure 21 ROC plot showing the performance difference of live predictions and post sampling individual predictions

Live Surgical Predictions Confusion matrix		Model Prediction: Sperm	Model Prediction: NO sperm	
Sperm Found in biopsy	Yes	30	3	**33**
	No	17	10	**27**
		47	**13**	

Table 6 Confusion matrix for live surgical predictions

Selected Validation Metrics	
Sensitivity	91%
Specificity	37%
Accuracy	67%
Precision	64%
Diagnostics Odds Ratio	5.88
Chi Square Metric	6.83

Table 7 Selected validation metrics for Live Surgical Predictions

C. Novice Surgeon Performance Comparison

Seven post graduate year 3 (PGY3) internal medicine residents and 3 PGY4 Urology residents were asked to select 3 sampling locations on each of the 20 highly affected images. Figure 21 shows an example evaluation slide. If no suitable sampling locations remained on the image, evaluators were instructed to change the circle color to red. Times to complete the evaluation were also recorded.

The indicated sample sites were compared with samples taken during biopsy. Each target was considered a positive prediction of sperm presence by the evaluator. If the target overlapped a region known to have sperm, the choice was marked a true positive. If a known sperm positive location was not marked, the selection was considered a type II false negative error. False positive and true negative scores were given if a known sperm negative location was marked or unmarked by the evaluator, respectively. To ensure a fair comparison between evaluators and Sperm Hunter, the evaluation protocol was completed by an experienced user using the Sperm Hunter predictions to choose sampling locations. The receiver-operator curves for both resident groups are shown in figure 22. The primary metric for success was to reduce type II errors, a secondary metric was the time taken to complete the evaluation. Figure 23 shows the time taken by each resident group compared to their number of type II errors. In both figures, Sperm Hunter Prime is shown in red.

D. Expert Surgeon Performance Comparison

Three fellowship trained surgeons and one urology attending surgeon were asked to complete the same evaluation. Figure 24 shows the ROC of these surgeons compared to Sperm Hunter.

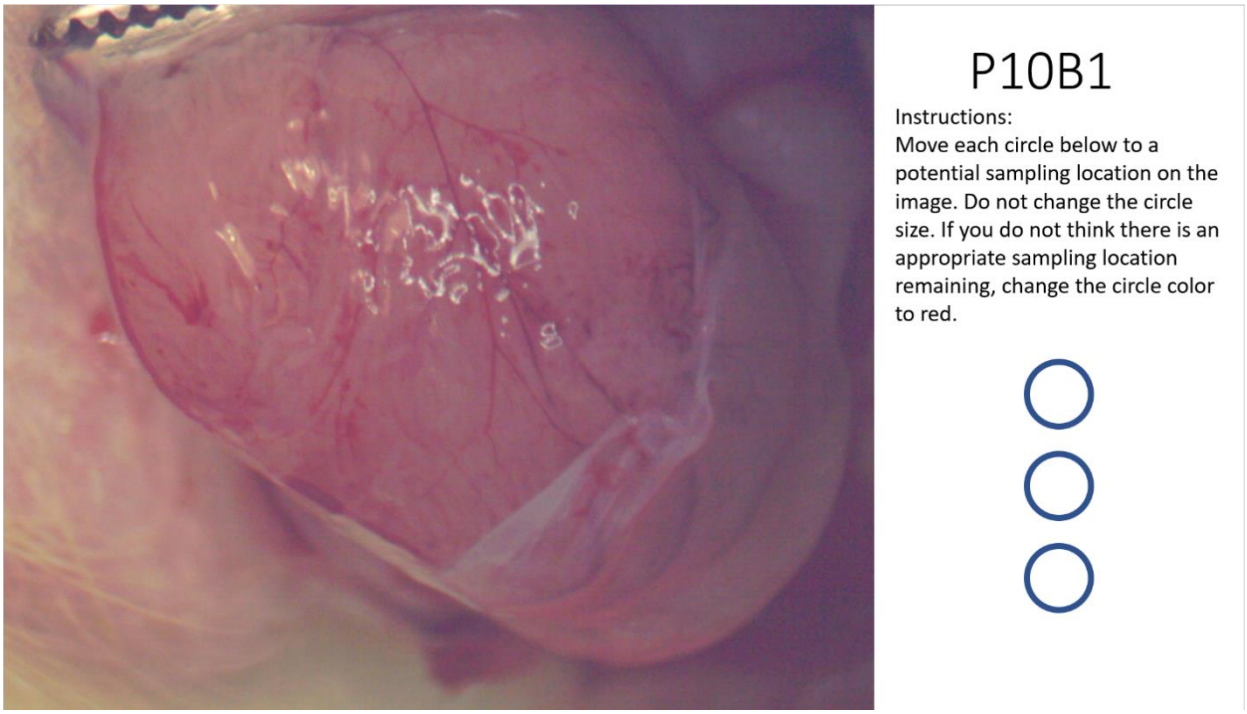


Figure 22 Example evaluation slide completed by all evaluators

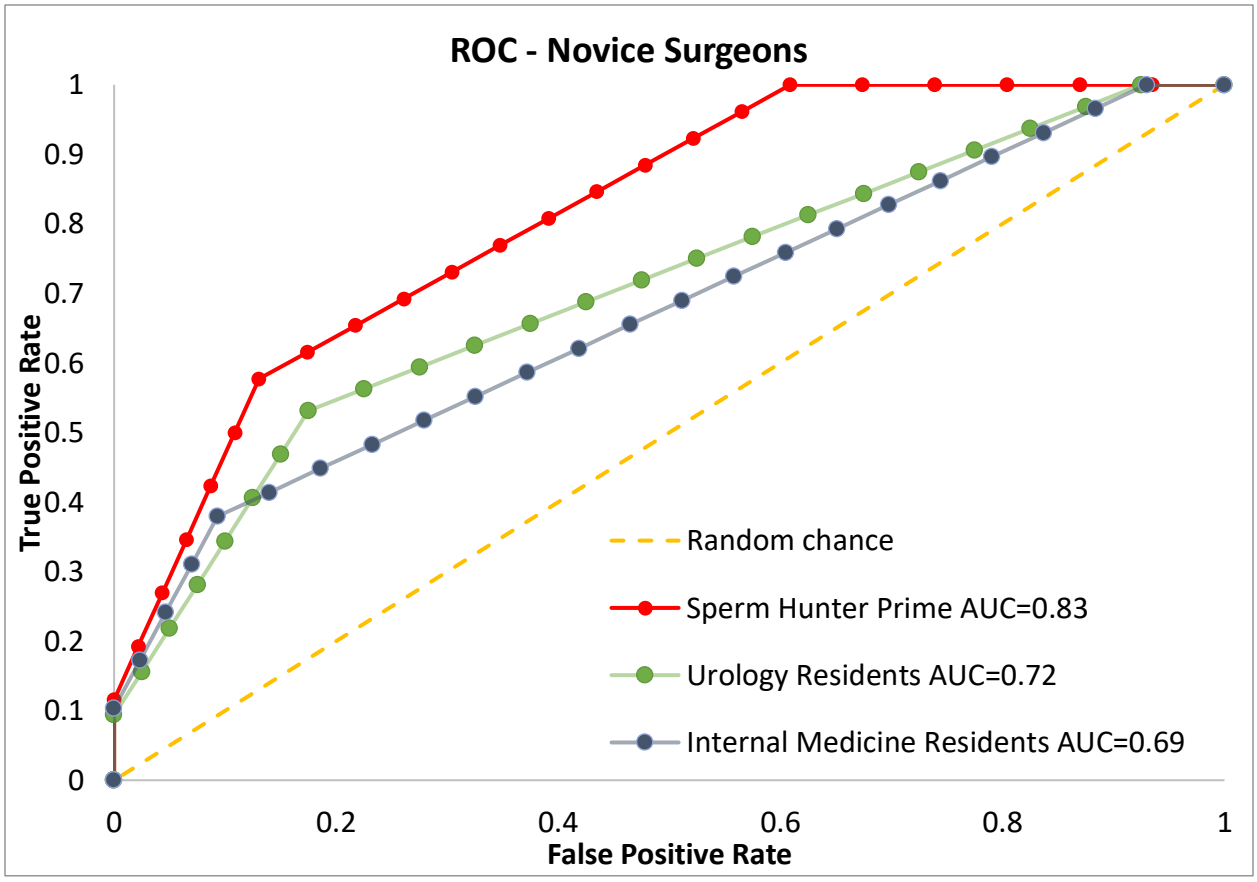


Figure 24 ROC plots of each group of novice evaluators

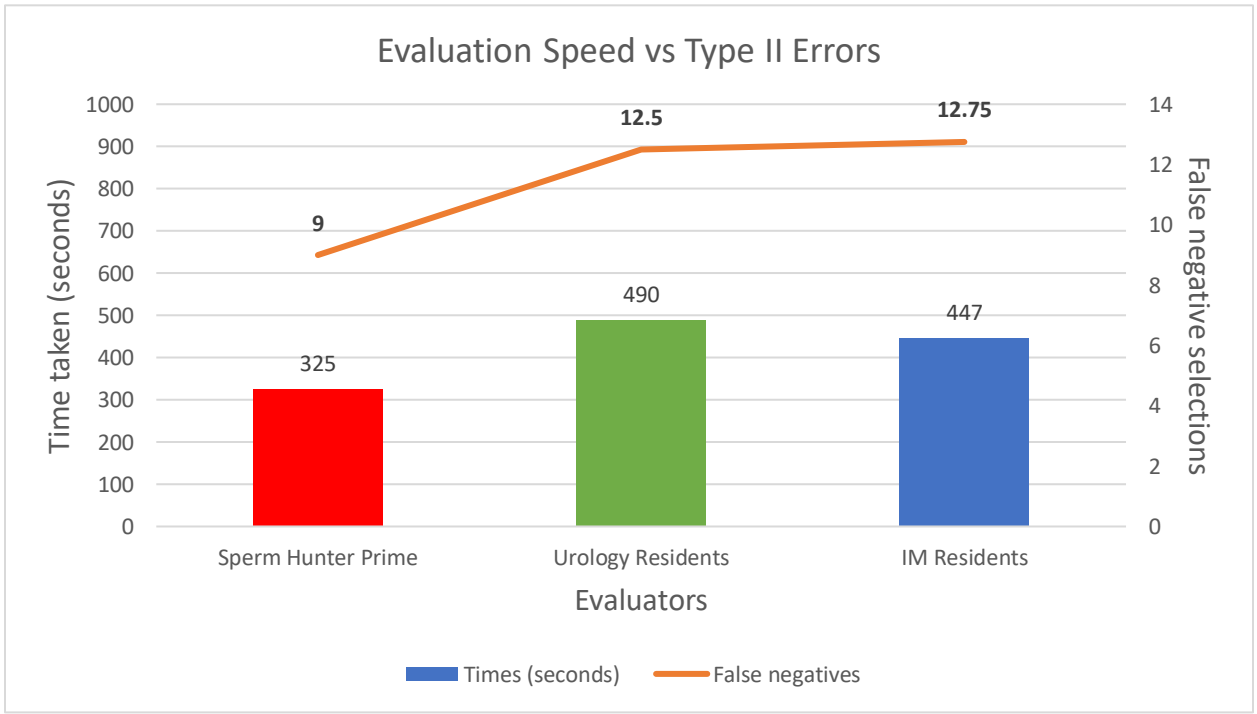


Figure 235 Time taken to complete the evaluation plotted with the number of false negative selections.

E. Naive Evaluator Performance Comparison

The stated goal of this project was to improve sperm retrieval rate and reduce operative time. This was predicated on the use of the system by expert surgeons. To evaluate the impact of the system on novice surgeon performance, an internal medicine PGY3 resident was instructed to select sampling locations on the evaluation form using Sperm Hunter predictions for guidance. The resident's selections with assistance are shown in figure 25 compared to expert surgeons and other internal medicine residents. Figure 26 shows the average time taken to complete the evaluation protocol by each group of evaluators. Sperm Hunter figures are an average of both operators.

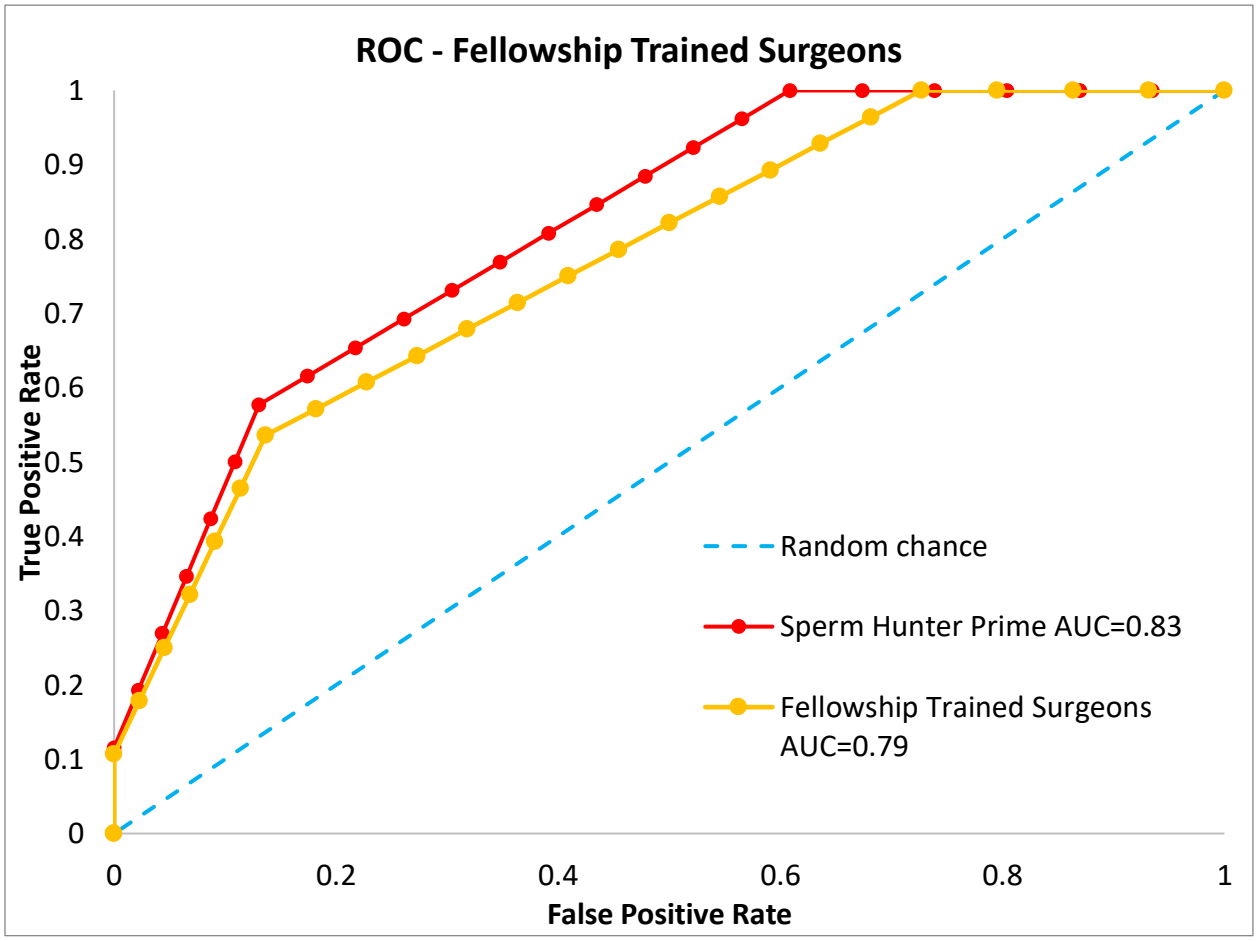


Figure 24 ROC of expert surgeons compared to Sperm Hunter Prime

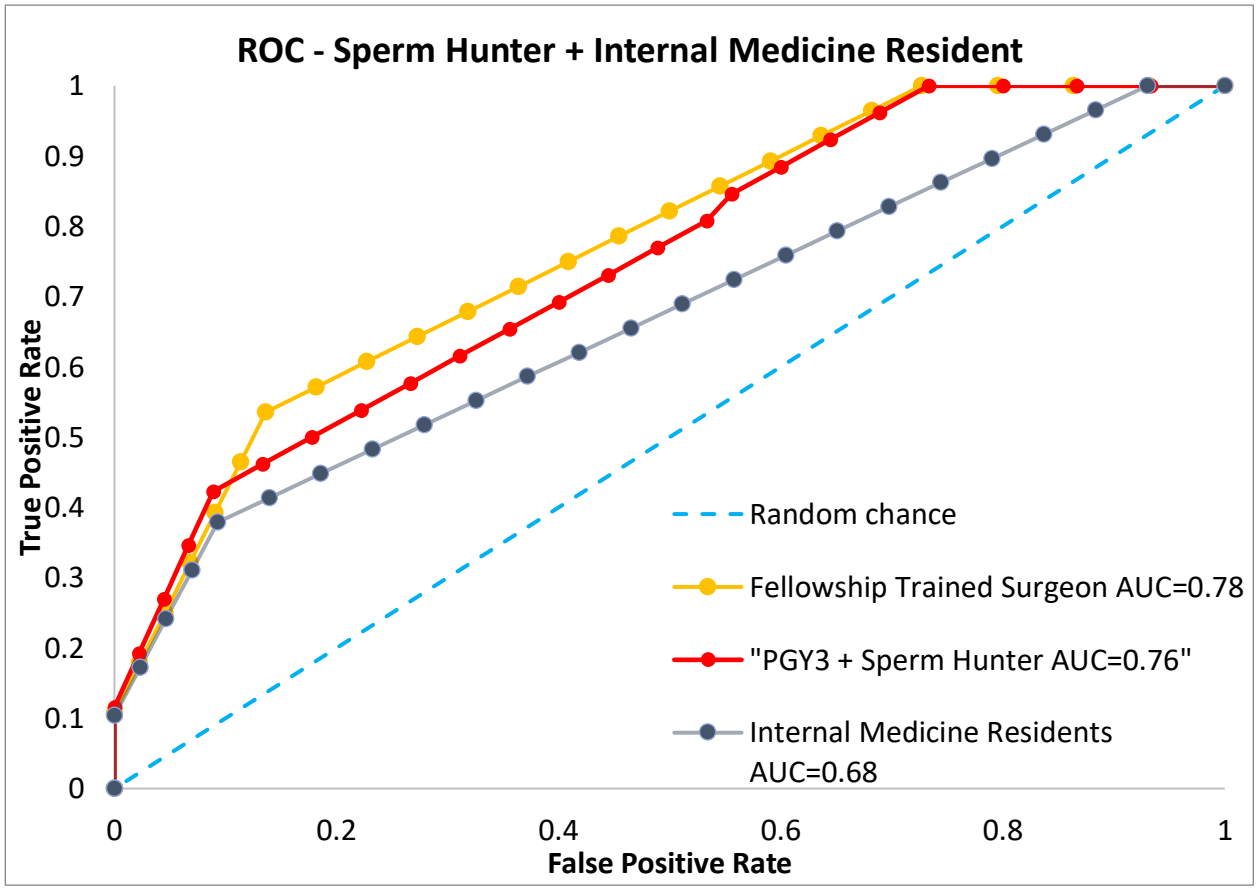


Figure 30 ROC of an internal medicine resident guided by Sperm Hunter predictions compared to expert surgeons and other IM residents

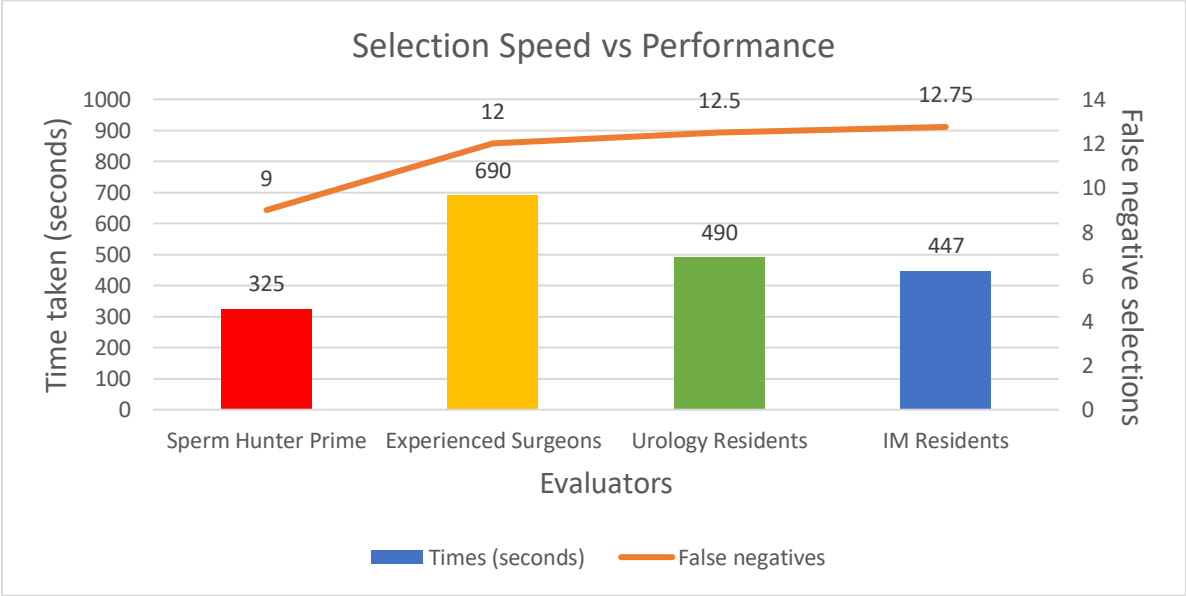


Figure 29 Time taken to complete evaluation plotted with the average number of type II errors for each group.

VII. Discussion

A. Capability of Sperm Hunter Prime

Sperm Hunter Prime was designed to improve sperm retrieval rates by automating sample site selection. Sperm Hunter had a 9% greater sperm retrieval rate than expert surgeons in the evaluation set (75% vs 66%). The average time taken to evaluate all images was 690 seconds for the experts and 325 seconds for Sperm Hunter, a time reduction of 53%. Even when operated by evaluators of different skill levels, Sperm Hunter provided consistent performance due to consistent predictions on the sample images. Overall image prediction time was 79 seconds; the remaining 246 seconds was the human operator moving markers on target.

Experiments predicting sampling locations on every frame of the biopsy showed artifacts arising when a tool was introduced into the surgical field. Light reflections and white markers on the surgical tool caused the most disturbance in the prediction. It was determined that a single prediction image taken immediately prior to sampling and frozen during sampling was the most effective way to communicate high sperm likelihood predictions without distraction. Figure 27 shows the proposed heads-up display layout: a side-by-side freezeframe of the predicted location.

B. Observations of Features

During the analysis of model performance several physiological features were noted as being preferentially sampled by Sperm Hunter. Loops in seminiferous tubules were present in 54% of high sperm predicted locations. Tubules that were higher than the surface of the testis were present in 37.5% of predicted regions. Most prevalent of all were opaque white tubules which were present in 91% of predictions. Blood resulting from biopsy was a major issue during testing. However, in some cases high sperm predictions were indicated by the model in the center of a blood-filled region of the testis. In all of these cases, multiple features were present: tubule loops that were higher than the testis surface, or loops with a slightly higher opacity than surrounding tubules. Propensity for opacity caused many overall

false positive predictions in non-testis tissue. Notably if the epididymis or remnants of the tunica albuginea were visible, both structures were indicated to have sperm. In these cases, the human operator must have the basic knowledge of testicular structures to ignore these false positives.

During the post biopsy testing of Sperm Hunter, it became apparent that after an initial sample was taken from the testis, the sampled region was no longer highlighted to contain sperm. However, other locations that were previously not indicated showed positive sperm predictions. Sperm Hunter was designed to indicate the locations of highest sperm probability in the image given. When the highest probability location was removed, the next highest probability region was indicated as shown in figure 28. This indicates a method of exhaustive sampling may be appropriate where after initial sampling of high sperm probability regions, a second prediction is processed of the same area of the testis to identify new sperm locations. This process would be repeated until no further locations are indicated or enough sampling has been completed per the surgeon's discretion.

C. Limitations of the Rat ASD Model

The rat ASD model proved to be an effective method of training a neural network to identify biological structures. Unfortunately, the size of the rat testis limited the number of samples that could be taken. After 6 samples, the testis was no longer suitable for sperm extraction due to blood obscuring the surgical field and a lack of undisturbed tubules to sample. Testis size also contributed to the large biopsy size relative to testis volume. In the human case, a 10 mg sample accounts for 0.06% of an average 15 g testis. In the rat case, a 10 mg sample accounts for 0.57% of the 1.75 g testis. The reduction of surface area requires a higher level of magnification to isolate the testis resulting in a smaller surface for evaluation by the neural network.

D. Pathway to Human Use

While there are differences between the size and appearance of seminiferous tubules in the human and the rat testis, there is a physiological similarity. Figure 29 shows the rat ASD trained Sperm

Hunter predictions on a human testis. No preprocessing or alterations were made to the image taken from the digital microscope camera. While the model is able to identify locations on the human testis, further research must be done to identify sperm counts at each predicted location to validate model performance. Using the rat ASD model as the baseline and augmenting the model using human testis as further training data may improve model performance while validating the model's use for human sample site prediction.

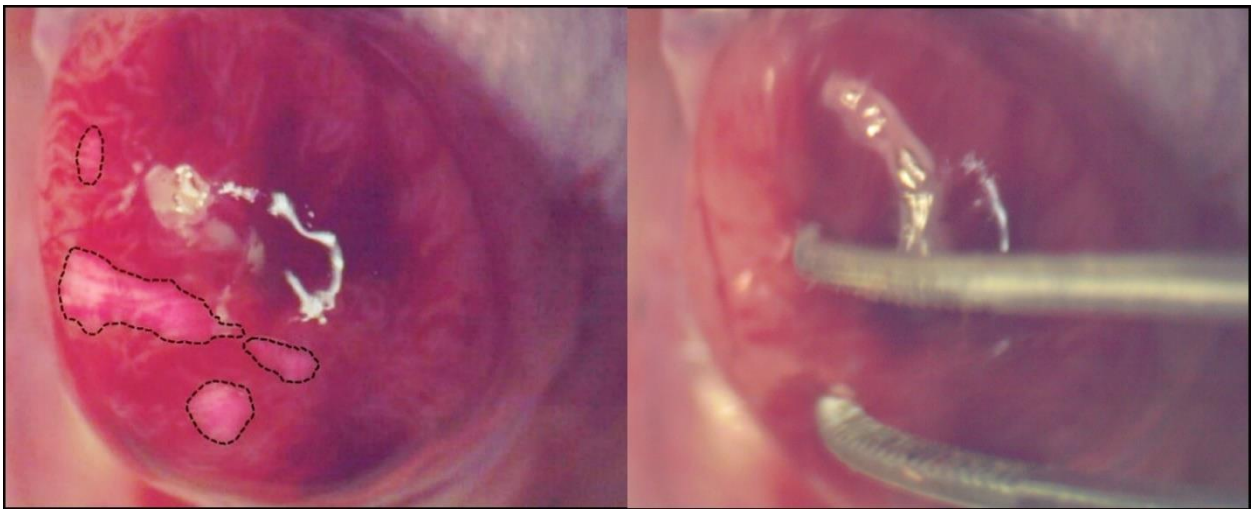


Figure 31 The proposed heads-up-display format. The prediction image is frozen on the left for use as a reference for the live sampling (right)

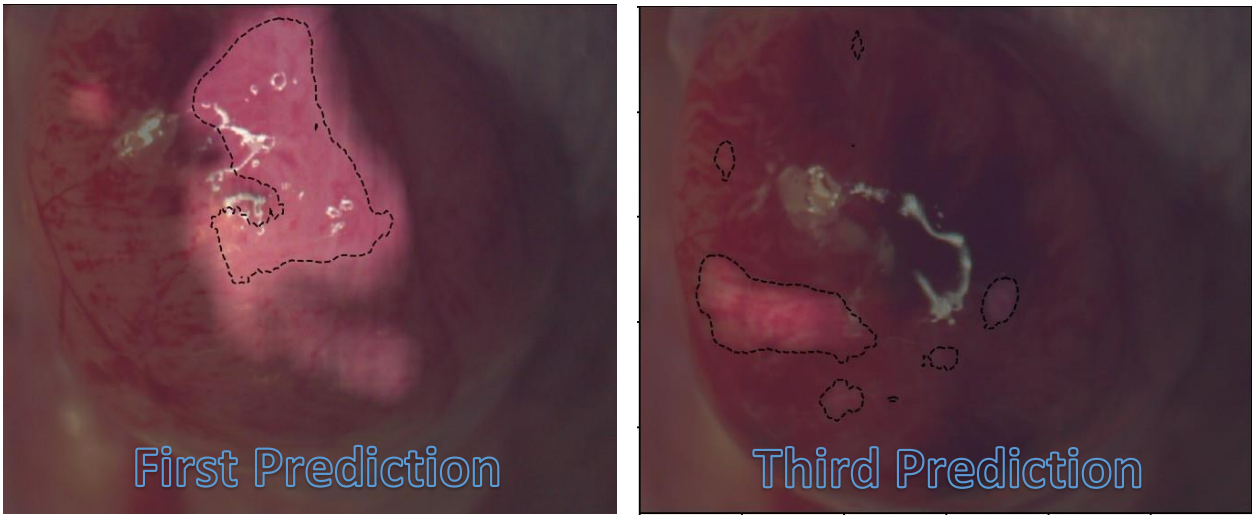


Figure 32 Initial prediction (left) and subsequent prediction (right) showing a previously undetected region of high sperm likelihood

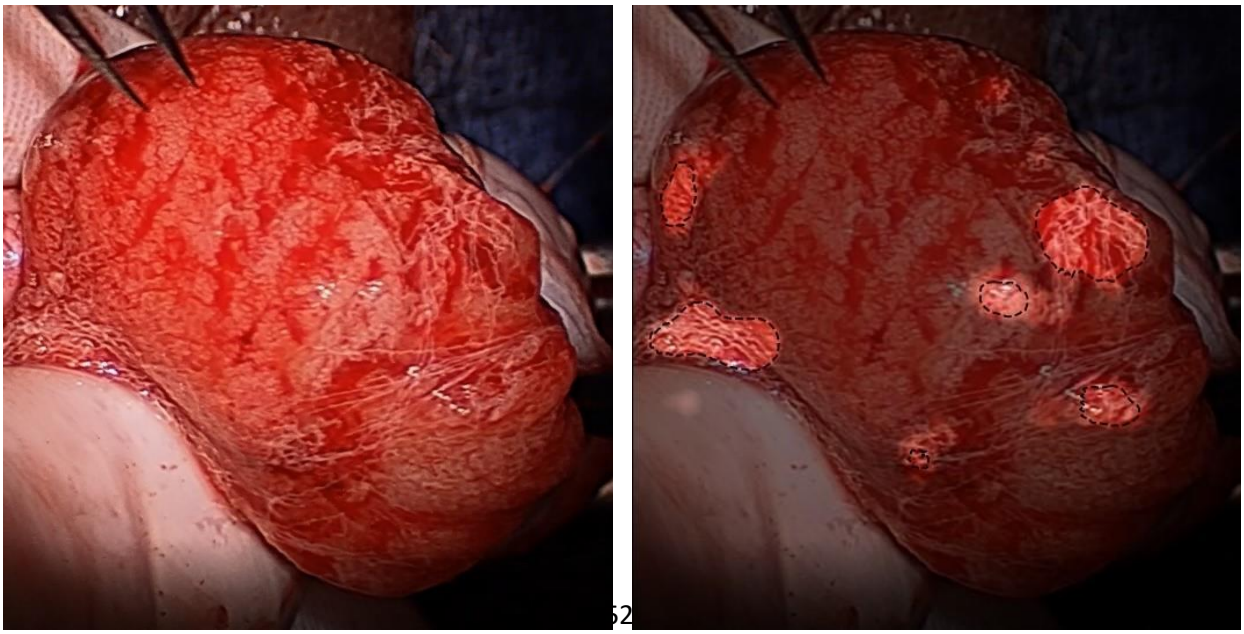


Figure 33 Sperm Hunter predictions on human testis images.

VIII. Conclusion

A. The Problem of Sample Site Selection

For the MicroTESE procedure, sample site selection is the only process controllable by the surgeon during surgery to maximize sperm yield. Differentiation and stratification of seminiferous tubules is expected to be an objective process and is directly dependent on the surgeon's experience with the procedure. This is one reason novice surgeons have a 12% lower sperm retrieval rate than expert surgeons.

Creating an algorithm to look for wider and more opaque tubules neglected characteristics such as shape and general layout of a group of tubules that could provide information to retrieve sperm more efficiently and effectively. To avoid the pitfalls of a directed searching algorithm, a neural network was trained without any initial bias regarding which tubules to sample. The network was given an input of a testis image and an output of sperm count at the sampling locations. This approach allowed flexibility for the neural network to identify significant features that coincided with high sperm yield without constraints.

To train the neural network model, Sprague-Dawley rats were turned into animal analogs of the spermatogenic dysfunction case using the spermatotoxic agent busulfan. The inputs of the neural network were the images taken during biopsy with sampling sites annotated. The output was the sperm count at each sampling location. The neural network developed a set of features and weights that determined the locations of highest sperm likelihood in each input image. This trained network was evaluated using a new group of treated rats and the results of the neural network predictions were evaluated for individual image accuracy and compared against experts.

The neural network, Sperm Hunter, identified 91% of all sperm dense locations in the test image set and outperformed the expert surgeons with a 9% greater sperm retrieval rate when operated by an expert. In the hands of a novice operator, Sperm Hunter improved the operator's performance to match expert surgeons. In both cases, use of the Sperm Hunter predictions reduced time taken to complete the evaluations by over 53%.

B. Why the Model Was Successful

The unguided nature of the neural network training process allowed for the emergence of predictive features without external interference. For the neural network, these features coincided with the maximization of sperm extraction. When interpreting the neural network's predictions for sperm dense locations, three features are easily identifiable to the human operator. Opaque tubules were visible in 91% of predicted locations, loops in seminiferous tubules were present in 54% of predicted locations, and tubules protruding from the surface of the testis were present in 37.5% of predicted regions. The latter two features have not been identified in previous literature and may indicate new metrics for manual sperm dense tubule selection.

C. Applicability of this Approach to Other Domains

The approach used to develop Sperm Hunter can be summarized in 3 steps:

1. Image the initial condition of an object
2. Change/sample the object and record the outcomes of each change.
3. Train a neural network on the initial condition image with the location of the change marked and the associated outcome labeled

This simple process will train the neural network to identify relevant features in the initial object condition that would coincide with the measured outcome. This process is best suited to develop models that predict the regions where specific target metrics are most likely to occur. One possible application for this approach is to optically identify regions that maximize the probability of collecting diseased tissue for diagnostic biopsies. This approach is not limited to visual wavelength optical data. Adding imaging modalities will provide dimensionality to the input data and may allow for more complex feature representations to be learned by the neural network. More study must be done on this topic to identify the limits of feature identification by neural networks. Fortunately, the application domain of this process is diverse, and the technology is readily available for experimentation.

Bibliography

Anger, Jennifer T., Gerald J. Wang, Stephen A. Boorjian, and Marc Goldstein. 2004. “Sperm Cryopreservation and in Vitro Fertilization/Intracytoplasmic Sperm Injection in Men with Congenital Bilateral Absence of the Vas Deferens: A Success Story.” *Fertility and Sterility* 82 (5): 1452–54. https://doi.org/10.1016/j.fertnstert.2004.05.079.

Ashraf, Mohamed C., Sankalp Singh, Dharma Raj, Sujatha Ramakrishnan, and Sandro C. Esteves. 2013. “Micro-Dissection Testicular Sperm Extraction as an Alternative for Sperm Acquisition in the Most Difficult Cases of Azoospermia: Technique and Preliminary Results in India.” *Journal of Human Reproductive Sciences* 6 (2): 111–23. https://doi.org/10.4103/0974-1208.117175.

Ballard, D. H. 1981. “Generalizing the Hough Transform to Detect Arbitrary Shapes.” *Pattern Recognition* 13 (2): 111–22. https://doi.org/10.1016/0031-3203(81)90009-1.

Berookhim, Boback M., and Peter N. Schlegel. 2014. “Azoospermia Due to Spermatogenic Failure.” *Urologic Clinics of North America* 41 (1): 97–113. https://doi.org/10.1016/j.ucl.2013.08.004.

Bhul, Allen E., James C. Cornette, Kenneth T. Kirton, and Yang-Dar Yuan. 1982. “Hypophysectomized Male Rats Treated with Polydimethylsiloxane Capsules Containing Testosterone : Efftects on Spermatogenesis , Fertility , and Reproductive Tract Concentrations of Androgens.” *Biology of Reproduction* 27 (July): 183–88.

Craft, Ian, F S Rc, F O G Rc, Valerie Bennett, B Sc, Mohamed Taranissi, M O G Rc, Natasha Nicholson, and B Sc. 1995. “Percutaneous Epididymal Sperm Aspiration and Intracytoplasmic Sperm Injection in the Management of Infertility Due to Obstructive Azoospermia.” *Fertility and Sterility* 63 (5): 1038–42. https://doi.org/10.1016/S0015-0282(16)57544-X.

Crosnoe, Lindsey E., Ethan Grober, Dana Ohl, and Edward D. Kim. 2013. “Exogenous Testosterone: A Preventable Cause of Male Infertility.” *Translational Andrology and Urology* 2 (2): 106–13. https://doi.org/10.3978/j.issn.2223-4683.2013.06.01.

Devroey, P., J. Liu, Z. Nagy, A. Goossens, H. Tournaye, M. Camus, A. van Steirteghem, and S. Silber. 1995. “Pregnancies after Testicular Sperm Extraction and Intracytoplasmic Sperm Injection in Non-Obstructive Azoospermia.” *Human Reproduction* 10 (6): 1457–60. https://doi.org/10.1093/HUMREP/10.6.1457.

Duda, Richard O., and Peter E. Hart. 1972. “Use of the Hough Transformation to Detect Lines and Curves in Pictures.” *Communications of the ACM* 15 (1): 11–15. https://doi.org/10.1145/361237.361242.

Esteves, S C. 2011. “Sperm Retrieval Techniques for the Azoospermic Male.” *Iranian Journal of Reproductive Medicine* 9 (5): 570–83. https://doi.org/10.1590/S1677-55382011000500002.

Esteves, Sandro C., Ricardo Miyaoka, José Eduardo Orosz, and Ashok Agarwal. 2013. “An Update on Sperm Retrieval Techniques for Azoospermic Males.” *Clinics* 68 (SUPPL. 1): 99–110. https://doi.org/10.6061/clinics/2013(Sup01)11.

Girshick, Ross. 2015. “Fast R-CNN.” *Proceedings of the IEEE International Conference on Computer Vision* 2015 Inter: 1440–48. https://doi.org/10.1109/ICCV.2015.169.

Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. “Rich Feature Hierarchies for

Accurate Object Detection and Semantic Segmentation." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–87. https://doi.org/10.1109/CVPR.2014.81.

Hales, F. 1984. "Effect of Spermatogenesis by Testosterone in Adult Male Rats : Pregnancy Outcome and Progeny '." *Biol Reprod* 31 (July): 221–30.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2014. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8691 LNCS (PART 3): 346–61. https://doi.org/10.1007/978-3-319-10578-9_23.

Ishikawa, Tomomoto, Ryuichiro Nose, Kohei Yamaguchi, Koji Chiba, and Masato Fujisawa. 2010. "Learning Curves of Microdissection Testicular Sperm Extraction for Nonobstructive Azoospermia." *Fertility and Sterility* 94 (3): 1008–11. https://doi.org/10.1016/j.fertnstert.2009.03.108.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems* 2: 1097–1105.

Leonardis, Aleš, Horst Bischof, Axel Pinz, Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. "Computer Vision – ECCV 2006 SURF: Speeded Up Robust Features." *Computer Vision – ECCV 2006* 3951: 404-417–417. https://doi.org/10.1007/11744023.

Liu, Yufei, Yong Zhu, Ling Di, E. Charles Osterberg, Feng Liu, Lin He, Hongliang Hu, Yiran Huang, Philip S. Li, and Zheng Li. 2014. "Raman Spectroscopy as an Ex Vivo Noninvasive Approach to Distinguish Complete and Incomplete Spermatogenesis within Human Seminiferous Tubules." *Fertility and Sterility* 102 (1): 54-60.e2. https://doi.org/10.1016/j.fertnstert.2014.03.035.

Low, David G. 2004. "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision*, 91–110. https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf.

Matthews, Gerald J., Peter N. Schlegel, and Marc Goldstein. 1995. "Patency Following Microsurgical Vasoepididymostomy and Vasovasostomy: Temporal Considerations." *The Journal of Urology* 154 (6): 2070–73. https://doi.org/10.1016/S0022-5347(01)66697-7.

Mosquera-Lopez, Clara, Sos Agaian, Alejandro Velez-Hoyos, and Ian Thompson. 2015. "Computer-Aided Prostate Cancer Diagnosis from Digitized Histopathology: A Review on Texture-Based Systems." *IEEE Reviews in Biomedical Engineering* 8: 98–113. https://doi.org/10.1109/RBME.2014.2340401.

Panahi, Mohadeseh, Saeideh Keshavarz, Farhad Rahmanifar, Amin Tamadon, Davood Mehrabani, Negar Karimaghai, Masood Sepehrimanesh, and Heydar Aqababa. 2015. "Busulfan Induced Azoospermia: Stereological Evaluation of Testes in Rat." *Veterinary Research Forum : An International Quarterly Journal* 6 (4): 273–78. http://www.ncbi.nlm.nih.gov/pubmed/26973761%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4769331.

Practice, The, and Reproductive Medicine. 2008. "Sperm Retrieval for Obstructive Azoospermia." *Fertility and Sterility* 90 (5 SUPPL.): 213–18. https://doi.org/10.1016/j.fertnstert.2008.08.047.

Ramasamy, Ranjith, Erik S. Fisher, Joseph A. Ricci, Robert A. Leung, and Peter N. Schlegel. 2011. "Duration of Microdissection Testicular Sperm Extraction Procedures: Relationship to Sperm Retrieval Success." *Journal of Urology* 185 (4): 1394–97.

https://doi.org/10.1016/j.juro.2010.11.074.

Ramasamy, Ranjith, Kathleen Lin, Lucinda Veeck Gosden, Zev Rosenwaks, Gianpiero D. Palermo, and Peter N. Schlegel. 2009. “High Serum FSH Levels in Men with Nonobstructive Azoospermia Does Not Affect Success of Microdissection Testicular Sperm Extraction.” *Fertility and Sterility* 92 (2): 590–93. https://doi.org/10.1016/j.fertnstert.2008.07.1703.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6): 1137–49. https://doi.org/10.1109/TPAMI.2016.2577031.

Rosenfeld, Azriel, and Chiao Yung Sher. 1998. “Detecting Image Primitives Using Feature Pyramids.” *Information Sciences* 107 (1–4): 127–47. https://doi.org/10.1016/S0020-0255(97)10021-4.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. “Learning Representations by Back-Propagating Errors.” *Nature* 323 (6088): 533–36. https://doi.org/10.1038/323533a0.

Schlegel, Peter N. 1999. “Testicular Sperm Extraction: Microdissection Improves Sperm Yield with Minimal Tissue Excision.” *Human Reproduction* 14 (1): 131–35. https://doi.org/10.1093/humrep/14.1.131.

Silber, Sherman J., and H. Edward Grotjan. 2004. “Microscopic Vasectomy Reversal 30 Years Later: A Summary of 4010 Cases by the Same Surgeon.” *Journal of Andrology* 25 (6): 845–59. https://doi.org/10.1002/j.1939-4640.2004.tb03150.x.

Song, Congzheng, and Vitaly Shmatikov. 2019. “Overlearning Reveals Sensitive Attributes,” 1–12. http://arxiv.org/abs/1905.11742.

Temple-Smith, P. D., G. J. Southwick, C. A. Yates, A. O. Trounson, and D. M. de Kretser. 1985. “Human Pregnancy by in Vitro Fertilization (IVF) Using Sperm Aspirated from the Epididymis.” *Journal of In Vitro Fertilization and Embryo Transfer* 2 (3): 119–22. https://doi.org/10.1007/BF01131497.

Walker, William H. 2010. “Non-Classical Actions of Testosterone and Spermatogenesis.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1546): 1557–69. https://doi.org/10.1098/rstb.2009.0258.

Wosnitzer, Matthew, Marc Goldstein, and Matthew P Hardy. 2014. “Review of Azoospermia.” *Spermatogenesis* 4 (1): e28218. https://doi.org/10.4161/spmg.28218.

VITA

NAME: Shrikant Dilipkumar Pandya

EDUCATION: B.S., Bioengineering, University of Illinois at Chicago, Chicago Illinois, 2014

PUBLICATIONS: Dobbs, Ryan & Pandya, Shrikant & Abern, Michael. (2017). Robotics and Urological Surgery.

Hall, D., Todorova-Koteva, K., Pandya, S., Bernard, B., Ouyang, B., Walsh, M., ... & Berry-Kravis, E. (2016). Neurological and endocrine phenotypes of fragile X carrier women. *Clinical genetics*, *89*(1), 60-67.

Vittal, P., Pandya, S., Sharp, K., Berry-Kravis, E., Zhou, L., Ouyang, B., ... & Hall, D. A. (2018). ASFMR1 splice variant: A predictor of fragile X-associated tremor/ataxia syndrome. *Neurology Genetics*, *4*(4).

PRESENTATIONS: Pandya*, S., Halgrimson, W., & Pagani, R. (2019). PD34-04 FINDING A SPERM AMONG THE WEEDS: NOVEL NEURAL NETWORK MODEL FOR AUGMENTED SEMINIFEROUS TUBULES CLASSIFICATION IN MICROTESE. *Journal of Urology*, *201*(Supplement 4). https://doi.org/10.1097/01.JU.0000556287.49270.3d

Pandya*, S., Halgrimson, W., Pagani, R., & Niederberger, C. (2020). PD58-07 SPERM HUNTER. *The Journal of Urology*, *203*, e1202. https://doi.org/10.1097/JU.0000000000000968.07

APPENDIX

UIC UNIVERSITY OF ILLINOIS AT CHICAGO

June 4, 2018

Gail S. Prins
Urology
M/C 955

Office of Animal Care and
Institutional Biosafety Committees (MC 672)
Office of the Vice Chancellor for Research
206 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Dear Dr. Prins:

The protocol indicated below was reviewed at a convened ACC meeting in accordance with the Animal Care Policies of the University of Illinois at Chicago on **4/17/2018**. *The protocol was not initiated until final clarifications were reviewed and approved on* **6/4/2018**. *The protocol is approved for a period of 3* years with annual continuation.

Title of Application: Computer Aided Targeting for Microsurgery

ACC Number: 18-052

Initial Approval Period: 6/4/2018 to 4/17/2019

Current Funding: ***Currently protocol NOT matched to specific funding source. Modification will need to be submitted prior to Just in time or acceptance of award to match protocol to external funding source. All animal work proposed in the funding application must be covered by an approved protocol.***

UIC is the only performance site currently approved for this protocol.

This institution has Animal Welfare Assurance Number A3460.01 on file with the Office of Laboratory Animal Welfare (OLAW), NIH. **This letter may only be provided as proof of IACUC approval for those specific funding sources listed above in which all portions of the funding proposal are matched to this ACC protocol.**

In addition, all investigators are responsible for ensuring compliance with all federal and institutional policies and regulations related to use of animals under this protocol and the funding sources listed on this protocol. Please use OLAW's "*What Investigators Need to Know about the Use of Animals*" (http://grants.nih.gov/grants/olaw/InvestigatorsNeed2Know.pdf) as a reference guide. Thank you for complying with the Animal Care Policies and Procedures of UIC.

Sincerely yours,

Timothy J. Koh, PhD
Chair, Animal Care Committee
TJK /ss
cc: BRL, ACC File, Lynn Birch, Shrikant Pandya

Phone (312) 996-1972 • Fax (312) 996-9088 • www.research.uic.edu

UIC UNIVERSITY OF ILLINOIS AT CHICAGO

Office of Animal Care and
Institutional Biosafety Committees (MC 672)
Office of the Vice Chancellor for Research
206 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

February 19, 2021

Gail S. Prins
Urology
M/C 955

Dear Dr. Prins:

The protocol indicated below was reviewed at a convened ACC meeting in accordance with the Animal Care Policies of the University of Illinois at Chicago on **02-16-2021**. *The protocol is approved for a period of **3*** years with annual continuation.

Title of Application: **Computer Aided Targeting for Microsurgery**

ACC Number: **21-013**

Initial Approval Period: **02-16-2021 to 02-16-2022**

Current Funding: ***Currently protocol NOT matched to specific funding source. Modification will need to be submitted prior to Just in time or acceptance of award to match protocol to external funding source. All animal work proposed in the funding application must be covered by an approved protocol.***

UIC is the only performance site currently approved for this protocol.

This institution has Animal Welfare Assurance Number D16-00290 (A3460.01) on file with the Office of Laboratory Animal Welfare (OLAW), NIH. **This letter may only be provided as proof of IACUC approval for those specific funding sources listed above in which all portions of the funding proposal are matched to this ACC protocol.**

In addition, all investigators are responsible for ensuring compliance with all federal and institutional policies and regulations related to use of animals under this protocol and the funding sources listed on this protocol. Please use OLAW's "*What Investigators Need to Know about the Use of Animals*" (http://grants.nih.gov/grants/olaw/InvestigatorsNeed2Know.pdf) as a reference guide. Thank you for complying with the Animal Care Policies and Procedures of UIC.

Sincerely yours,

Amy Lasek

Amy Lasek, PhD
Chair, Animal Care Committee
AL/ss
cc: BRL, ACC File, Shrikant Pandya, Lynn Birch

Phone (312) 996-1972 • Fax (312) 996-9088 • www.research.uic.edu