Leveraging Generative AI to Modernize Applications

Scott Smith
NCDEQ-DIT

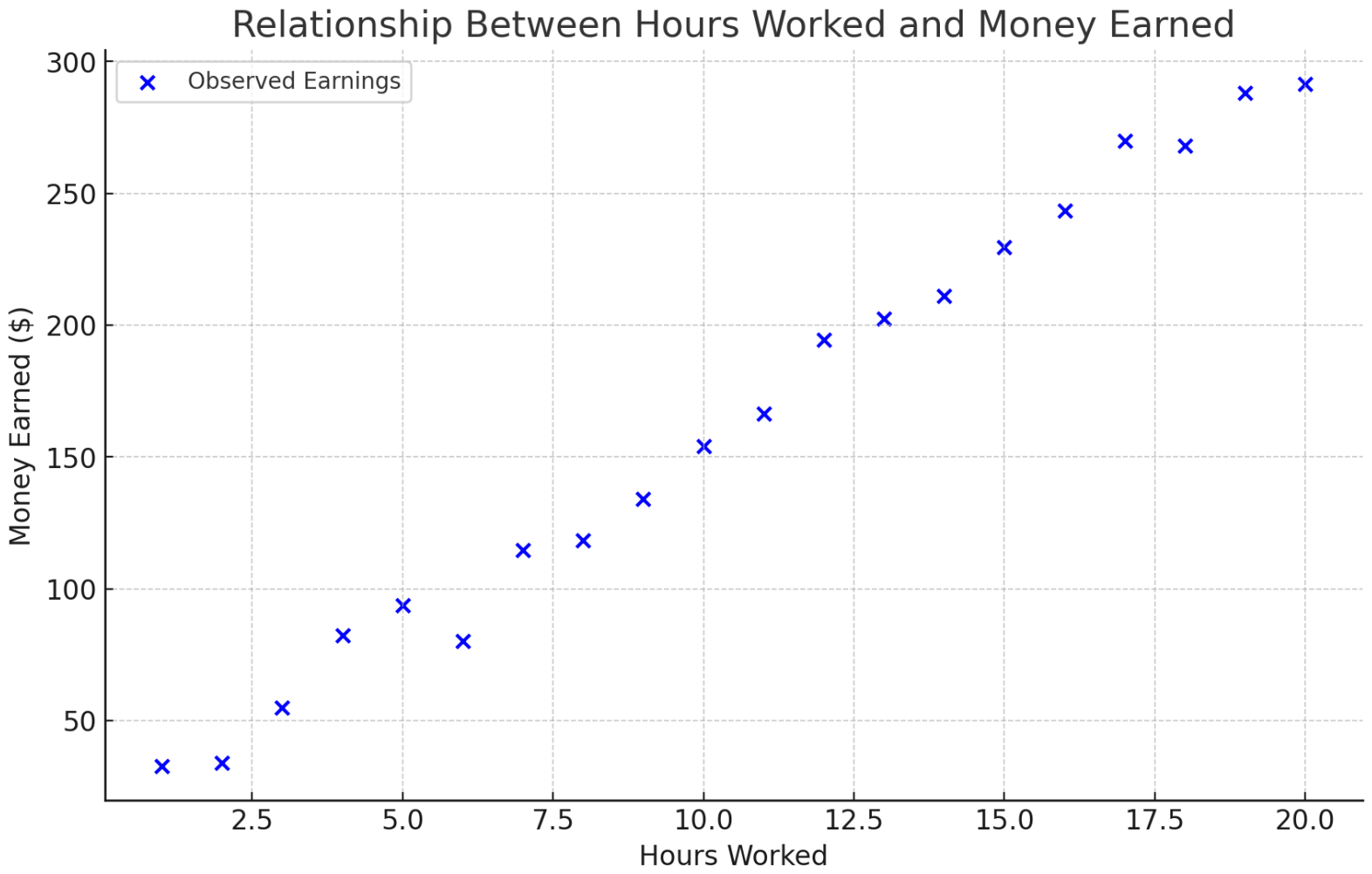
November 15, 2024

NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

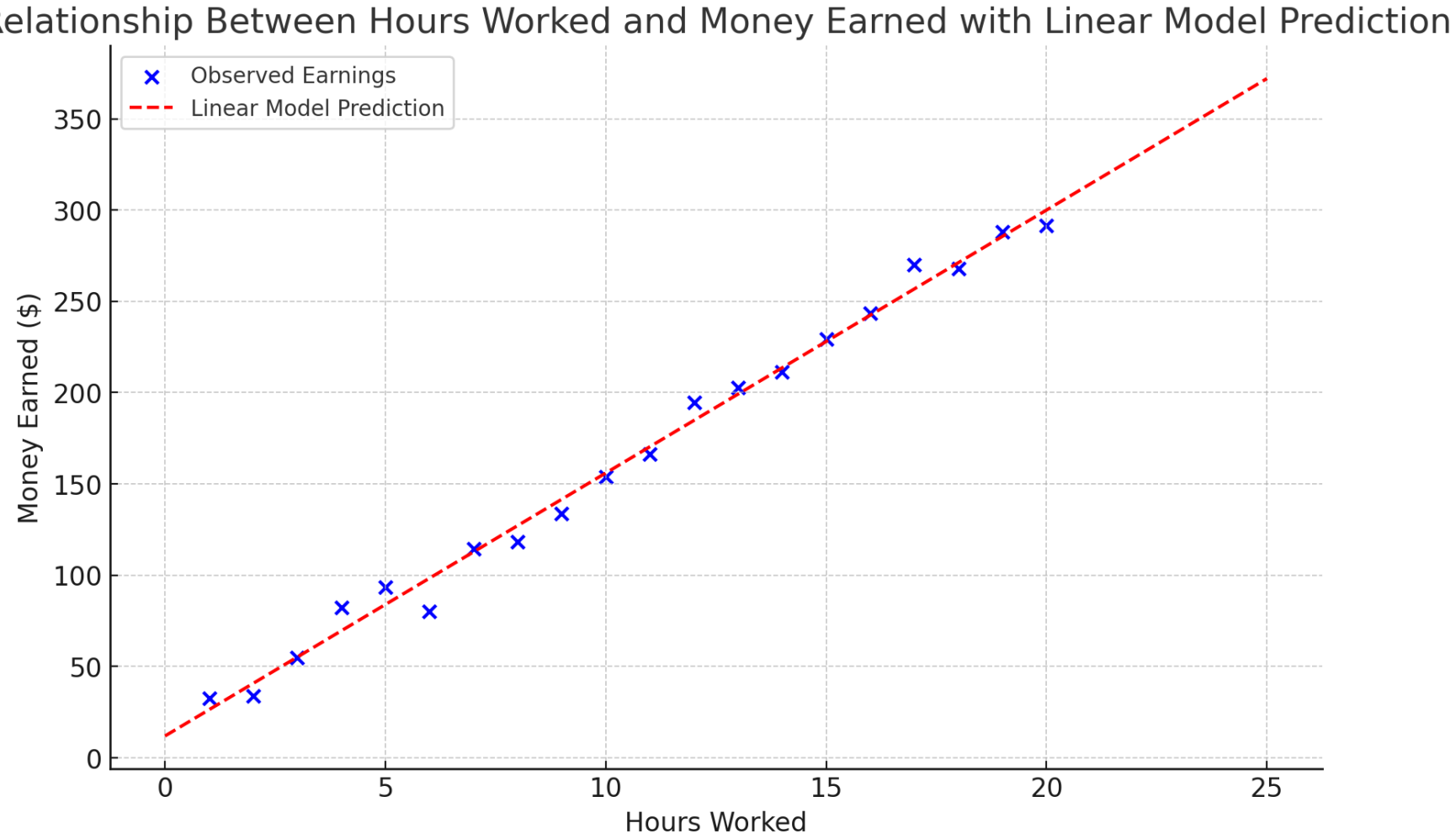
What is a model?

- A mathematical representation of a real-world process.
- When you supply some input to the model, the model can process it and relate it to some output.

What is a Model?



What is a Model?



NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

Generative AI is just a model

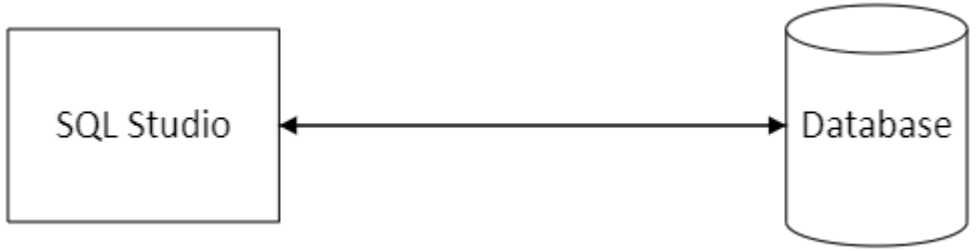
(albeit a very complicated one!)

Use Case Number 1

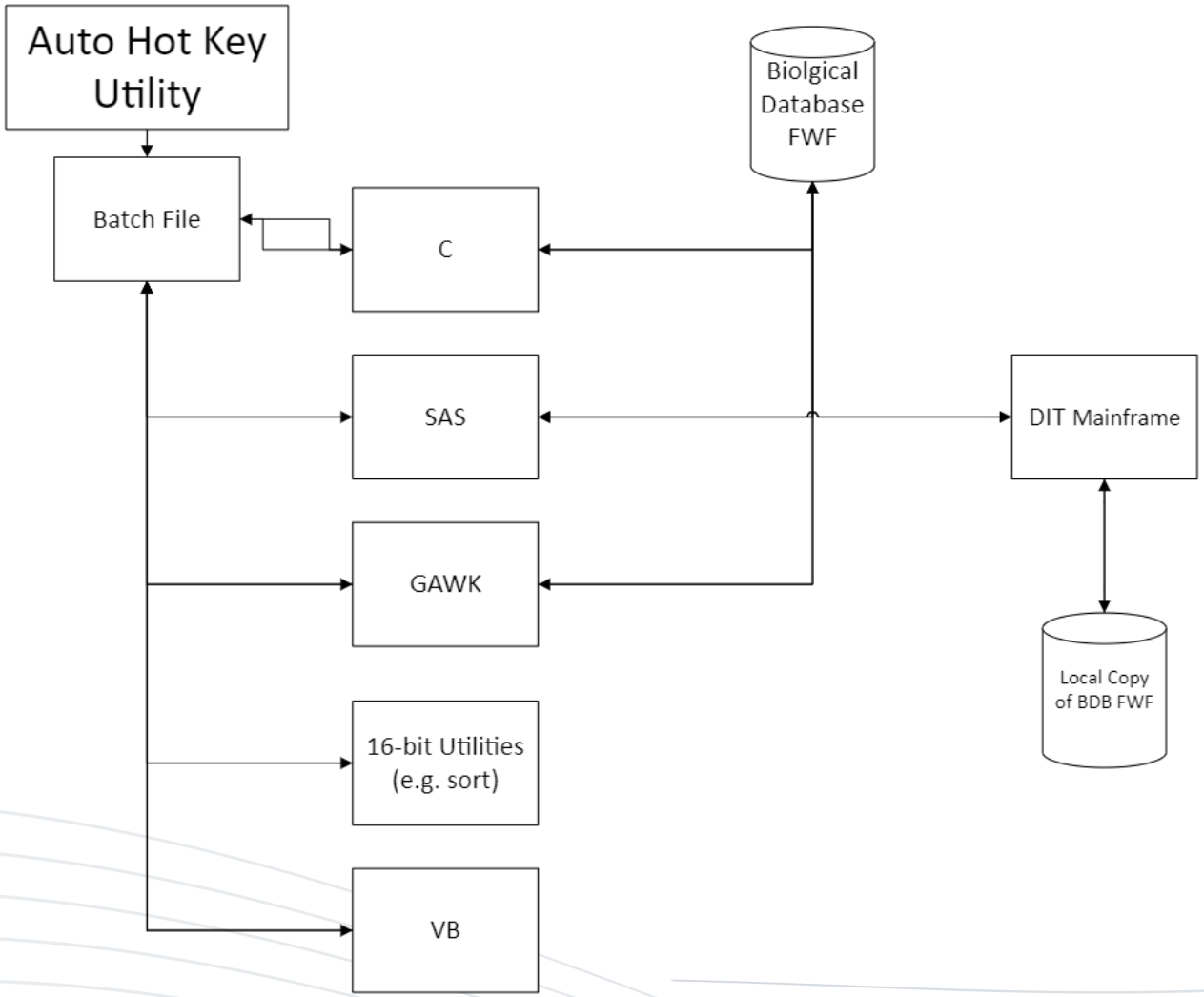
Decoding the Past: Using Generative AI to Navigate Legacy Code

NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

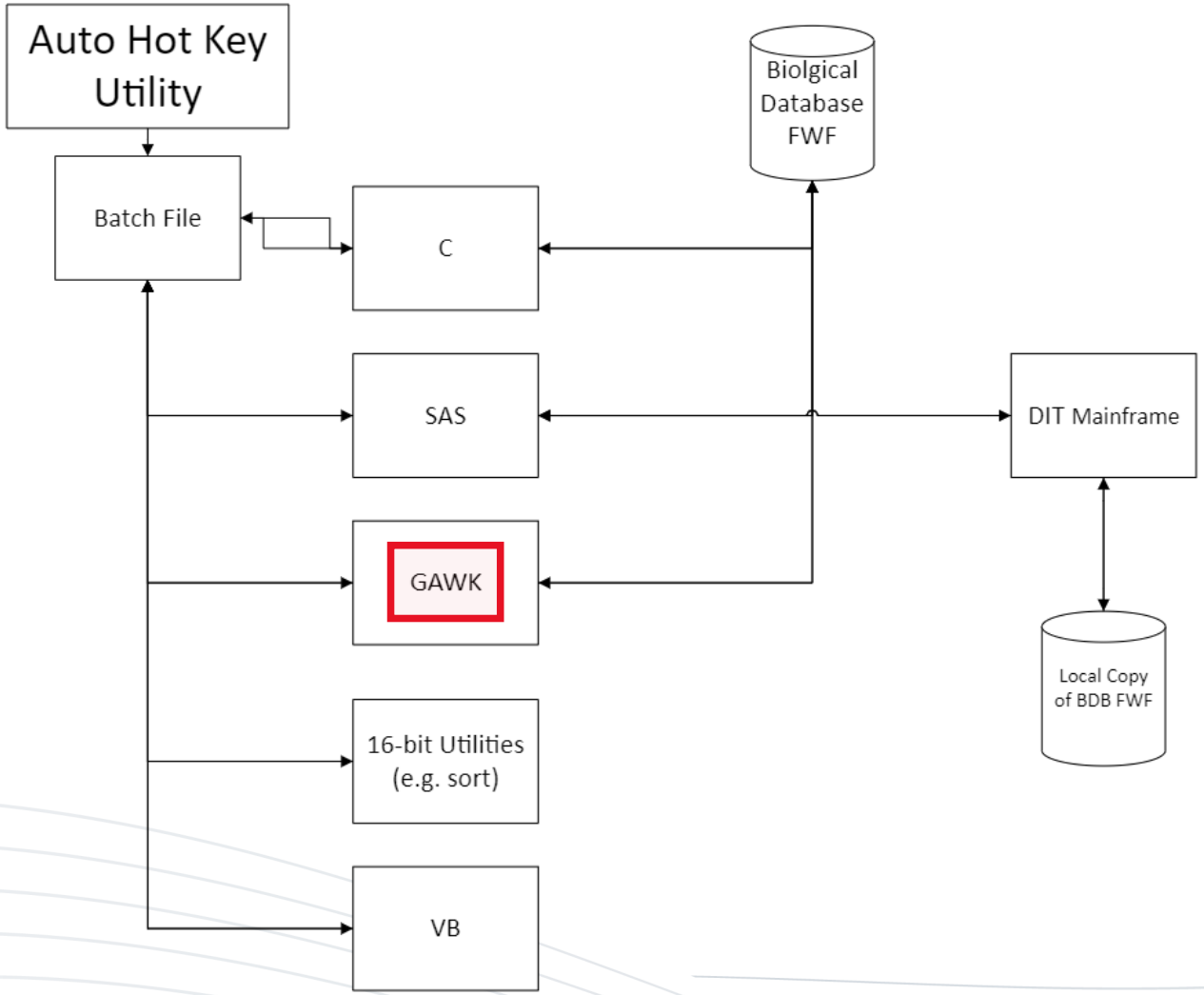
Use Case No. 1



Use Case No. 1


Auto Hot Key Utility
Batch File
Biolgical Database FWF
C
SAS
GAWK
16-bit Utilities (e.g. sort)
VB
DIT Mainframe
Local Copy of BDB FWF

Use Case No. 1


Auto Hot Key Utility
Batch File
Biolgical Database FWF
C
SAS
GAWK
DIT Mainframe
Local Copy of BDB FWF
16-bit Utilities (e.g. sort)
VB

AWK is a powerful text-processing language commonly used in Unix/Linux environments to manipulate and analyze structured text data

I love AWK because it allows me to build the wrong solution to problems so quickly :)

-Dave on Stack Overflow

If we wanted to hire an AWK expert, we would need to recruit from a retirement home, not a university!

-Anonymous

Use Case No. 1

```
BEGIN{k=1;i=1;nr=1;fc=1; hpset="\x1b(8U\x1b\x26\x6c\x31\x4f\x1b\x26\x6c\x31\x53\x1b\x26\x6c\x38\x44";
hp16="\x1b\x28\x73\x33\x34\x48\x1b\x26\x6c\x32\x43";#hp12="\x1b\x28\x73\x31\x37\x48\x1b\x26\x6c\x34\x43\x1b(s7B";}hp
12="\x1b(6N\x1b(s0p16.67h8.5v0s0b0T\x1b\x26\x6c\x35\x43";
}{if(NR==1)fn=FILENAME;}{if(fc==1){if(fn==FILENAME){spc=substr($0,27,10);if(spc ~ /^[0-
9]/){for(jj=length(substr($0,38));jj>0;jj--)if(substr($0,37+jj,1)==" ");else
break;sp[spc]=substr($0,38,jj);}next;}else{fn=FILENAME;fc++;}}}{if(fn==FILENAME);else{if(i){i=0;pt(n,l,w,k,id,f);c[id]=k;print
hpset "CONTENTS OF TRANSACTION FILE: " FILENAME " "AKA;}id=substr($0,2,1);
if(svid==id);else {svid=id;if(id ~ /[3459]/){print  hp16 f[id];latin( );}else print  hp16 f[id]"\n";}z="";for(y=1;y<c[id];y++)z=z
substr($0,b[id,y],w[id,y]) "|";if(id ~ /[3459]/ && spc != substr($0,12,10)){latin( );}printf("%s%03d %s\n",hp12,nr++ %
1000,z);next}}
{if($1==id){ n[id,k]=$2;l[id,k]=length($2);b[id,k]=$3;w[id,k++]=$4;}else{if(NR==1);else {pt(n,l,w,k,id,f);c[id]=k;}
k=1;id=$1;n[id,k]=$2;l[id,k]=length($2);b[id,k]=$3;w[id,k++]=$4;f[id]="";}}function pt(n, l, w, k, id, f,     j, p, q, r, s, t, e, u){\e="
";
p=0;for(j=1;j<k;j++)if(l[id,j]>p)p=l[id,j];for(j=1;j<=p;j++){for(q=1;q<k;q++){t=2*w[id,q];r=int(t/2);t=t-r;u=" ";#if(j==p)u="|";else u="
";if(l[id,q]<1+p-j)s=s substr(e,1,r+t+1)u;else s=s substr(e,1,r)substr(n[id,q],j-(p-l[id,q]),1)substr(e,1,t)u;}if(f[id]=="")f[id]="       "
s;else f[id]=f[id] "\n       " s;s="";}}function latin( ){spc=substr($0,12,10);print hp16 "                                    " sp[spc]"\n\n";}
```

Use Case No. 1

6500 Lines

And this was just one script!

NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

Disclaimer:

We use a version of ChatGPT that does not save any of our inputs and does not train on our data.

We can delete our data at any time.

OpenAI

Demonstration

NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

Use Case Number 2

Bridging Expertise Gaps: Empowering Cross-Disciplinary Collaboration with Generative AI

NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

NCDMF Ageing Program

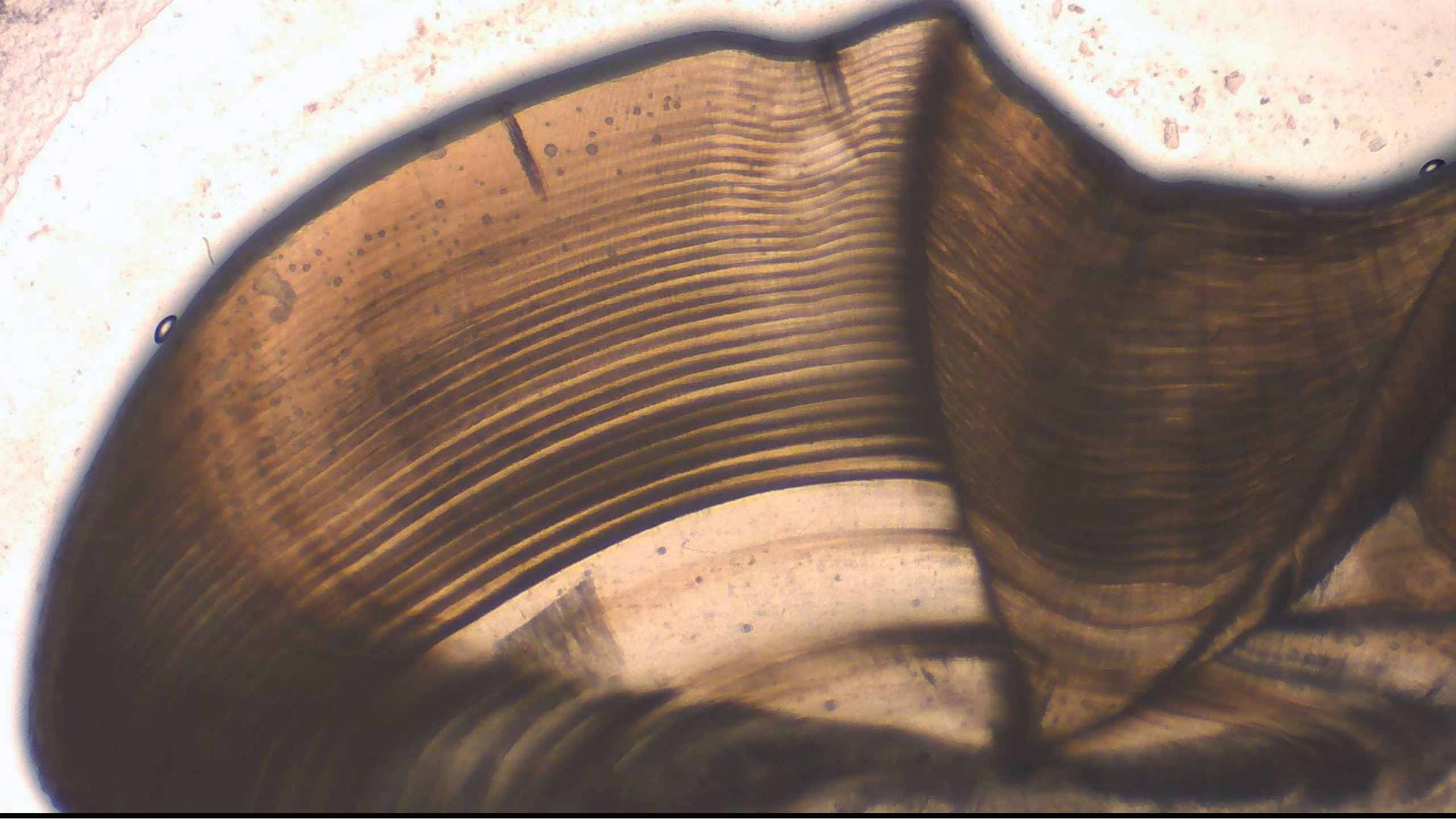
- The program harvests ageing structures from fishes throughout the year to determine fish ages for modelling
- Otoliths are the most common structure

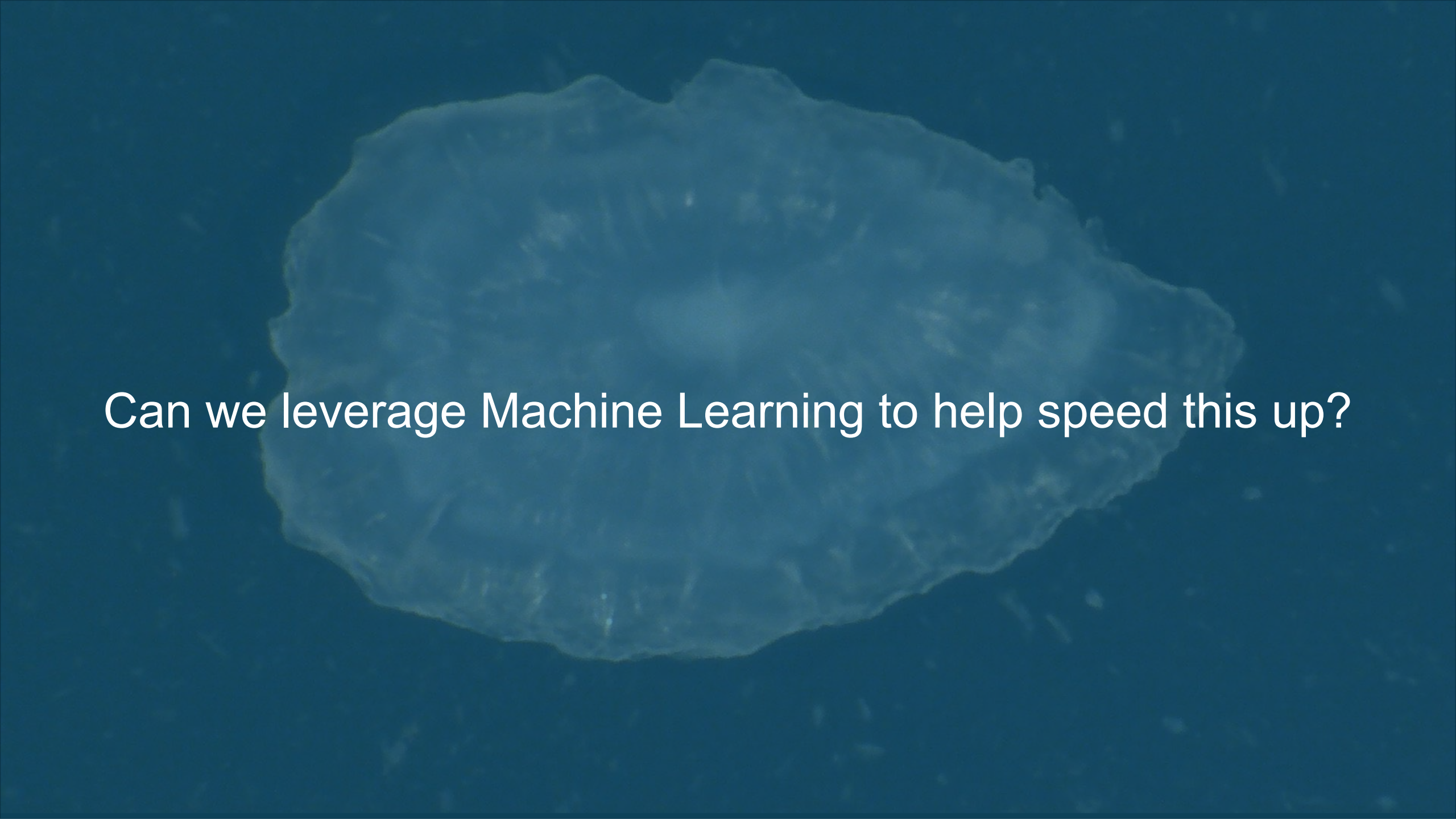
NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

NCDMF Ageing Program

- 12,209 age samples collected in 2023
- 9,596 age samples were read
- This is a manual process!

NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY




Can we leverage Machine Learning to help speed this up?

Humans:

~30 Seconds – 1 Minute per otolith

Convolutional Neural Network:

~0.01 Seconds per otolith

Images processed by a human in one minute:

$$\text{Images per minute} = \frac{60 \text{ seconds}}{30 \text{ seconds per image}} = 2 \text{ images per minute}$$

4,798 Minutes Total

x3 People

Images processed by a CNN in one minute:

$$\text{Images per minute} = \frac{60 \text{ seconds}}{0.01 \text{ seconds per image}} = 6,000 \text{ images per minute}$$

1.59 Minutes Total

NCDIT NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

Use of GenAI

- Walks you through code creation for a CNN
- Helps you understand the stats, functions, and tests
- Gives tailored guidance on errors, optimizations, etc.

```
# Load the image
image = cv2.imread('C:\\Users\\Scott\\Documents\\Otolith Project\\img\\1511232-1.jpg')

# Convert the image to grayscale
gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

# Apply Gaussian blur
blurred = cv2.GaussianBlur(gray, (5, 5), 0)

# Apply adaptive threshold
thresh = cv2.adaptiveThreshold(blurred, 255, cv2.ADAPTIVE_THRESH_GAUSSIAN_C, cv2.THRESH_

# Erosion and dilation to remove noise
kernel = cv2.getStructuringElement(cv2.MORPH_ELLIPSE, (5,5))
thresh = cv2.erode(thresh, kernel, iterations=1)
thresh = cv2.dilate(thresh, kernel, iterations=1)

# Find contours
contours, hierarchy = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPL

# Iterate over the contours and save the cropped otoliths
count = 0
for i, contour in enumerate(contours):
    area = cv2.contourArea(contour)

    # Filtering based on area
    if area > 500:  # you can adjust this threshold as needed
        x, y, w, h = cv2.boundingRect(contour)
        cropped_otolith = image[y:y+h, x:x+w]
        cv2.imwrite(f'cropped_otolith_{count+1}.jpg', cropped_otolith)
        count += 1
```

NCDIT | NORTH CAROLINA DEPARTMENT OF INFORMATION TECHNOLOGY

Thank You

Mike Ware
Chief Information Officer
NCDEQ-DIT
mike.ware@nc.gov

Scott Smith
Biological Database Administrator
NCDEQ-DIT
scott.smith@deq.nc.gov

Semir Badic
Applications Systems Analyst
NCDEQ-DIT
semir.badic@deq.nc.gov