

Bioinformatics

Individual Major Proposal

St Olaf. College

Center for Integrative Studies

Elise Hachfeld

Class of 2026

Martha Zillig

Faculty Advisor

I. Major Description

Bioinformatics is an interdisciplinary field that integrates components of biology, computer science, statistics and data science to answer biological questions by analyzing and interpreting complex datasets with computational methods. Bioinformaticians use data to unearth patterns and/or build models to predict the behavior of biological systems.

Biological specialties that often incorporate bioinformatics include genomics, proteomics and evolutionary biology. Applications include analyzing DNA sequencing data, predicting the structure of proteins, constructing phylogenetic trees and modeling population changes. Analysis of images from microscopes or medical scans, called bioimage informatics, and signals like audio, seismic waves, or EEG are also subspecialties in bioinformatics.

Scientists working with biological data must be well versed in the ethical considerations of human and animal subjects research, and be able to protect the integrity of their research.

The proposed major emphasizes skills in programming, data analysis, and mathematical modeling, and their application to foundational biology.

II. Central Questions

- A. What kind of biological questions can be addressed with computational methods?
- B. How can computational tools be used to analyze and interpret biological data?
 - 1. What programs and programming languages are used?
 - 2. What statistical methods are used for analyzing biological data?
- C. What are the ethical considerations in the use of bioinformatics in science and medicine?
- D. How can bioinformaticians responsibly and accurately communicate findings?
- E. What careers are a bioinformatics major appropriate for?

III. Course Requirements

A. Foundation Courses

BIO 150: Biodiversity Foundations

Completed

- Introduces the use of the R programming language for analyzing biological data.
- Molecular Evolutionary Genetics Analysis (MEGA) software used for sequence alignment and building phylogenetic trees.
- Covers evolution, phylogenetic trees, and population genetics.

MSCS 164: Data Science 1

Completed

- Introduces the basics of data science using the R programming language.
- Covers data cleaning and manipulation with the tidyverse and visualization with ggplot.

SDS 272: Statistics 1

Completed

- Introduces the concept of hypothesis testing.
- Covers t-tests, randomization tests, linear regression, and ANOVA.
- Covers interpretation of p-values, R-squared, confidence intervals, t-statistics, and F-statistics.

IM 392: Senior Project

- See Section IV. Senior Integrative Project.

B. Genetics Core

BIO 233: Intermediate Genetics

Completed

- Covers inheritance, genetic variation, structure, and gene expression, genetic disorders, epigenetics, and penetrance.
- Discusses genetics on the molecular, cellular and organismal levels.

BIO 292: Gene Editing

Completed

- Discusses gene editing technology such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and CRISPR-Cas9.
- Elaborates on methods of genome sequencing.

C. Data Science & Computer Science Core

CSCI 200B: Intro to Machine Learning

In Progress

- Covers different machine learning algorithms, including their strengths and limitations.
- Discusses the real-world applications of machine learning and the ethical implications.

SDS 264: Data Science 2

Completed

- Covers functional programming, iteration, code quality, web scraping, APIs, text analysis, web design with R, and database management with SQL.
- Introduces version control using Git and Github.

SDS 341: Algorithms for Decision Making

Still Need

- Explores the application of machine learning to real-world data.
- Covers linear regression, nearest neighbor models, k-means clustering, shrinkage methods, decision trees and forests, boosting bagging, support vector machines and hierarchical clustering.

D. Mathematics & Statistics Core

MABIO 130: Exploring Biomathematics (0.25)

Completed

- Professors introduce a new technique each week for using mathematics to solve biology research problems.
- Several lectures had a bioinformatics focus, including one that used R.

SDS 272: Statistics 2

Completed

- Covers simple linear regression, multiple linear regression, and logistic regression.

SDS 284: Biostatistics: Design and Analysis

Completed

- Focuses on study design, methods of analyzing health-related data and communicating findings.

SDS 316: Advanced Statistical Modeling

Still Need

- Covers logistic and poisson regression, longitudinal data analysis and multilevel modeling.

E. Ethics Core

PHIL 250: Biomedical Ethics

Still Need

- Covers common ethical issues encountered in medicine, including euthanasia, abortion, and culturally-conscious medical care.

CSCI 263: Ethical Issues in Software Design

Still Need

- Addresses ethical issues involved in creating software, including algorithmic bias, privacy, and liability.

HIST 299: The Misinformation Age

Completed

- Provides an overview of how misinformation is created and propagated, both in history and in modern times.
- Covers the psychological basis of misinformation—why people believe and spread it.

SOAN 267: Medical Anthropology

Still Need

- Covers how different cultures experience and understand medicine and healing.

F. Optional Courses

BIO 391: Computational Neuroscience (topic)

BIO 391: Bioinformatics (topic)

G. Experiences

CURI: Research with Professor Martha Zillig

Completed

- Worked with Professor Zillig to gather animal species data from camera trap images using Timelapse 2 software.

BIO 297: Research with Professor Jean Porterfield

Completed

- Created an image processing program in Python to count the number of lipid droplets within a *Tetrahymena thermophila* cell.

BIO 297: Research with Professor Steven Freedberg

Completed

- Worked on optimizing Professor Freedberg's population simulation of altruism in Python.

IM 394: Academic Internship

Still Need

- Complete an academic internship related to my individual major during either summer 2025 or interim 2026.
- Consider doing an internship at Children's MN Minneapolis with the Genetics department.

Year 1			
Course #	Course Title	Ole Core req.	Major (yes/no)
Fall			
MATH 126	Calculus II	QCR	Yes
NORW 111	Beginning Norwegian I	<i>None</i>	No
WRIT 120	Wellness/Disability in the Nordic Region	WRR	No
Interim			
	Beginning Norwegian II	<i>None</i>	No
Spring			
BIO 150	Biodiversity Foundations	<i>NTS</i>	Yes
NORW 231	Intermediate Norwegian I	WLC	No
HIST 233	The Misinformation Age	GHS	Yes
FYS 120	The Pulse of Careers in Healthcare	FYS	No
MABIO 130	Exploring Biomathematics (0.25)	<i>None</i>	Yes

Year 2			
Course #	Course Title	Ole Core req.	Major (yes/no)
Fall			
BIO 233	Intermediate Genetics	<i>None</i>	Yes
STAT 172	Statistics I	<i>QCR</i>	Yes
REL 122	The Hebrew Bible	RFV	No
Interim			
BIO 292	Gene Editing	<i>None</i>	Yes
Spring			
BIO 227	Cell Biology	<i>None</i>	No
MSCS 164	Data Science I	<i>QCR</i>	Yes
NORW 244	Sami Traditions in Transition	<i>GHS, PAR</i>	No
BIO 297	Independent Research	<i>None</i>	Yes

Year 3			
Course #	Course Title	Ole Core req.	Major (yes/no)
Fall			
SDS 264	Data Science II	None	Yes
SDS 272	Statistics II	None	Yes
PSYCH 238	Biopsychology	None	No
BIO 297	Independent Research	None	Yes
Interim			
SDS 284	Biostatistics	None	Yes
Spring			
CSCI 200	Intro to Machine Learning	None	Yes
PSYCH 235	Sensation/Perception	None	No
NORW 205	Nordic Nature	CRE, WAC	No
PHIL 233	Kierkegaard	CTD, GHS	No

Year 4			
Course #	Course Title	Ole Core req.	Major (yes/no)
Fall			
PHIL 250	Biomedical Ethics	ERC	Yes
SDS 341	Algorithms for Decision Making	None	Yes
CSCI 263	Ethical Issues in Software Design	ERC, OEP	Yes
SOAN 267	Medical Anthropology		Yes
KINES 105	Inclusive Fitness (0.25)	ACB	No
Interim			
IM 394	Academic Internship	None	Yes
Spring			
SDS 316	Advanced Statistical Modeling	None	Yes
NEURO 239	Cellular & Molecular Neuroscience	None	No
	TBD		
IM 392	Senior Project II	None	Yes

IV. Rationale

Academic Rationale

While St. Olaf offers majors in biology, statistics & data science, and computer science that provide a comprehensive overview of their respective fields, a major in bioinformatics requires a specialized curriculum that integrates all these areas. The biology major provides a strong foundation for the biological basis of the questions asked of bioinformaticians, however it does not address the computational and statistical aspects of analyzing large sets of data. The statistics and data science major covers statistical theory and data analysis, particularly using R, but does not dive deep into its biological applications. Similarly, the computer science major covers algorithms using Python and C++ but does not focus on the challenges specific to large-scale biological data.

Bioinformatics necessitates a curriculum that emphasizes the integration of these disciplines. Combining coursework from biology, statistics & data science, and computer science with an emphasis on subtopics like genetics, computation, and biostatistics allows for an interdisciplinary approach.

Since bioinformatics has the potential to have an impact on public health and political policy, bioinformaticians must address the ethical and societal issues surrounding this research. My proposed curriculum addresses data privacy, the ethical concerns about gene editing, biased algorithms and responsible human subject research. These topics are especially important in a time where medical and scientific misinformation is pervasive and undermines public trust in the validity of science. By integrating these topics, I aim to educate myself not only the technical responsibilities of a bioinformatician, but also the social ones.

Personal Rationale

I have always been fascinated by life, both human and non-human. My backyard in Apple Valley, Minnesota, lovingly called “the swamp,” was the place I spent most of my childhood. I know practically every tree and rock by heart from spending countless hours exploring, always returning with the mud and burs to prove it. My family and I enjoyed hiking together and frequently visited Frontenac, Nerstrand Big Woods, William O’Brien,

Gooseberry Falls, Tettegouche and Temperance River State Parks. Through this exploration, I not only gained a respect for the natural world, but I also gained a curiosity that has stayed with me all these years.

As a teenager I was diagnosed with a rare neurological condition that changed my understanding of how biology impacts our lives. Since my mobility was limited for much of high school, I learned to explore my love for biology from a different, technological, perspective. During this time, I volunteered for Trace Labs where I learned about both open-source intelligence (OSINT), and for an online group dedicated to helping adoptees find their biological parents using genetic genealogy.

Living with a disability has made me acutely aware of the impact that research can have in addressing real-world issues. Bioinformatics is often at the forefront of rare disease research—helping to find the genetic basis for disease or design new drugs. By combining my interests in genetics, neuroscience, and computational tools, I hope to contribute to the efforts in improving the identification and treatment of rare diseases.

V. Senior Integrative Project

Idea #1: Genomic Analysis

Synopsis: An analysis and report of my full genome with annotation using the bioconductor in R or biopython.

Ideas:

- Sequence alignment (would need to get the raw FASTQ file)
- Variant calling (BAM > VCF) of SNPs - [SNPedia](#), CNVs and Indels
- Acquire whole genome data of my parents for comparison.

Steps:

1. Acquire genomic data (have the preliminary 11x coverage BAM file)
2. Check data quality and process into suitable format.
3. Exploratory data analysis.
4. Visualize results.

Idea #2: Machine Learning for Image Processing

Synopsis: Continue the project on counting lipid droplets with image processing, but implement a machine learning algorithm. Another student worked on this for a while but stopped.

VI. Web Portfolio

I will publish my web portfolio as a R and Github page so it is available regardless of St. Olaf's web hosting. The web portfolio will include the information contained in this portfolio, the sources that I am using to complete my senior project, and the final senior project itself. Additionally, it will include any materials I have found useful in learning about bioinformatics and addressing the central questions.

Structurally, this will be organized in an about section, where I discuss the rationale, major description and central questions, a separate section for course requirements, and at least one other for my senior project. I will either link to my web portfolio on my current github website (<https://elisefeld.github.io/>), or I will restructure this website to be my portfolio.

VII. Consultations

Anne Berry

- She suggested considering how my major would be named and whether to include something about genetics in the name. Additionally, she advised me that I don't need to provide too much detail about my senior project yet. (1/15/25)
- We discussed the possibility of getting funding for my senior project, maybe Steen/Magnus the Good. I will follow up with the Piper Center. (2/23/25)

Audrey Gunn

- We discussed the library databases that could be useful for my senior project. She recommended Web of Science, PubMed, and Cinahl for Project #1 and [this guide](#) for CSCI 263 for Project #2. Additionally, philosophy databases might be useful for answering the ethics core questions, and LISTA could be useful for data privacy. (01/16/25)

Dana Zimmerman

- In my monthly meeting with Dana, we discussed what it would look like to switch majors at this point in my year, since I was

currently taking required classes for the biology major that would not count towards my individual major. (10/21/24)

Jean Porterfield

- Professor Porterfield provided the initial consultation on my independent major. We discussed whether the Biology major still suited my academic interests, and whether an individual major could accommodate them. (09/6/24)
- She suggested doing a senior project that is genetics related to include a 300 level genetics component and round out the coursework. (1/28/25)

Martha Zillig

- We discussed what kind of presentation I would do for my senior project. It could be interesting to do a website with visualizations of my genome. Additionally, we discussed whether I would pursue more ancestry or health focused analysis—the latter having the issue of privacy. (1/12/25)