UDACITY

NANODEGREE PROGRAM SYLLABUS

AI for Healthcare

Overview

Play a critical role in enhancing clinical decision-making with machine learning to build the treatments of the future. Learn to build, evaluate, and integrate predictive models that have the power to transform patient outcomes. Begin by classifying and segmenting 2D and 3D medical images to augment diagnosis and then move on to modeling patient outcomes with electronic health records to optimize clinical trial testing decisions. Finally, build an algorithm that uses data collected from wearable devices to estimate the wearer's pulse rate in the presence of motion.

A graduate of this program will be able to:

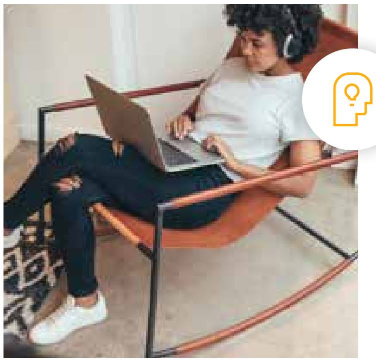
- Recommend appropriate imaging modalities for common clinical applications of 2D medical imaging
- Perform exploratory data analysis (EDA) on 2D medical imaging data to inform model training and explain model performance
- Establish the appropriate 'ground truth' methodologies for training algorithms to label medical images
- Extract images from a DICOM dataset
- Train common CNN architectures to classify 2D medical images

IN COLLABORATION WITH

AT&T lyft Google

Estimated Time:
4 Months at
10 hours / week

Prerequisites:
Intermediate Python, and Experience with Machine Learning



Flexible Learning:
Self-paced, so you can learn on the schedule that works best for you.

Need Help?
udacity.com/advisor
Discuss this program with an enrollment advisor.

- Translate outputs of medical imaging models for use by a clinician
- Plan necessary validations to prepare a medical imaging model for regulatory approval
- Detect major clinical abnormalities in a DICOM dataset
- Train machine learning models for classification tasks using real-world 3D medical imaging data
- Integrate models into a clinician's workflow and troubleshoot deployments
- Build machine learning models in a manner that is compliant with U.S. healthcare data security and privacy standards
- Use the Tensorflow Dataset API to scalably extract, transform, and load datasets that are aggregated at the line, encounter, and longitudinal (patient) data levels
- Analyze EHR datasets to check for common issues (data leakage, statistical properties, missing values, high cardinality) by performing exploratory data analysis with Tensorflow Data Analysis and Validation library
- Create categorical features from Key Industry Code Sets (ICD, CPT, NDC) and reduce dimensionality for high cardinality features
- Use Tensorflow feature columns on both continuous and categorical input features to create derived features (bucketing, cross-features, embeddings)
- Use Shapley values to select features for a model and identify the marginal contribution for each selected feature
- Analyze and determine biases for a model for key demographic groups
- Use the Tensorflow Probability library to train a model that provides uncertainty range predictions in order to allow for risk adjustment/prioritization and triaging of predictions
- Preprocess data (eliminate "noise") collected by IMU, PPG, and ECG sensors based on mechanical, physiology and environmental effects on the signal.
- Create an activity classification algorithm using signal processing and machine learning techniques
- Detect QRS complexes using one-dimensional time series processing techniques
- Evaluate algorithm performance without ground truth labels
- Generate a pulse rate algorithm that combines information from the PPG and IMU sensor streams

Course 1: Applying AI to EHR Data

With the transition to electronic health records (EHR) over the last decade, the amount of EHR data has increased exponentially, providing an incredible opportunity to unlock this data with AI to benefit the healthcare system. Learn the fundamental skills of working with EHR data in order to build and evaluate compliant, interpretable machine learning models that account for bias and uncertainty using cutting-edge libraries and tools including Tensorflow Probability, Aequitas, and Shapley. Understand the implications of key data privacy and security standards in healthcare. Apply industry code sets (ICD10-CM, CPT, HCPCS, NDC), transform datasets at different EHR data levels, and use Tensorflow to engineer features.

Course Project
Patient Selection for Diabetes Drug Testing

EHR data is becoming a key source of real-world evidence (RWE) for the pharmaceutical industry and regulators **to make decisions** on clinical trials. In this project, you will act as a data scientist for an exciting unicorn healthcare startup that has created a groundbreaking diabetes drug that is ready for clinical trial testing. Your task will be to build a regression model to predict the estimated hospitalization time for a patient in order to help select/filter patients for your study. First, you will perform exploratory data analysis in order to identify the dataset level and perform feature selection. Next, you will build necessary categorical and numerical feature transformations with Tensorflow. Lastly, you will build a model and apply various analysis frameworks, including Tensorflow Probability and Aequitas, to evaluate model bias and uncertainty.

	LEARNING OUTCOMES	
LESSON ONE	**EHR Data Security and Analysis**	• Understand U.S. healthcare data security and privacy best practices (e.g. HIPAA, HITECH) and how they affect utilizing protected health information (PHI) data and building models • Analyze EHR datasets to check for common issues (data leakage, statistical properties, missing values, high cardinality) by performing exploratory data analysis
LESSON TWO	**EHR Code Sets**	• Understand the usage and structure of key industry code sets (ICD, CPT, NDC). • Group and categorize data within EHR datasets using code sets.

LESSON THREE	**EHR Transformations & Feature Engineering**	• Use the Tensorflow Dataset API to scalably extract, transform, and load datasets • Build datasets aggregated at the line, encounter, and longitudinal(patient) data levels • Create derived features (bucketing, cross-features, embeddings) utilizing Tensorflow feature columns on both continuous and categorical input features
LESSON FOUR	**Building, Evaluating, and Interpreting Models**	• Analyze and determine biases for a model for key demographic groups by evaluating performance metrics across groups by using the Aequitas framework. • Train a model that provides an uncertainty range with the Tensorflow Probability library • Use Shapley values to select features for a model and identify the marginal contribution for each selected feature

Course 2: Applying AI to 2D Medical Imaging Data

2D imaging, such as X-ray, is widely used when making critical decisions about patient care and accessible by most healthcare centers around the world. With the advent of deep learning for non-medical imaging data over the past half decade, the world has quickly turned its attention to how AI could be specifically applied to medical imaging to improve clinical decision-making and to optimize workflows. Learn the fundamental skills needed to work with 2D medical imaging data and how to use AI to derive clinically-relevant insights from data gathered via different types of 2D medical imaging such as x-ray, mammography, and digital pathology. Extract 2D images from DICOM files and apply the appropriate tools to perform exploratory data analysis on them. Build different AI models for different clinical scenarios that involve 2D images and learn how to position AI tools for regulatory approval.

Course Project
Pneumonia Detection from Chest X-Rays

Chest X-ray exams are one of the most frequent and cost-effective types of medical imaging examinations. Deriving clinical diagnoses from chest X-rays can be challenging, however, even by skilled radiologists. When it comes to pneumonia, chest X-rays are the best available method for point-of-care diagnosis. More than 1 million adults are hospitalized with pneumonia and around 50,000 die from the disease every year in the US alone. The high prevalence of pneumonia makes it a good candidate for the development of a deep learning application for two reasons: 1) Data availability in a high enough quantity for training deep learning models for image classification 2) Opportunity for clinical aid by providing higher accuracy image reads of a difficult-to-diagnose disease and/or reduce clinical burnout by performing automated reads of very common scans. In this project, you will analyze data from the NIH Chest X-ray dataset and train a CNN to classify a given chest X-ray for the presence or absence of pneumonia. First, you'll curate training and testing sets that are appropriate for the clinical question at hand from a large collection of medical images. Then, you will create a pipeline to extract images from DICOM files that can be fed into the CNN for model training. Lastly, you'll write an FDA 501(k) validation plan that formally describes your model, the data that it was trained on, and a validation plan that meets FDA criteria in order to obtain clearance of the software being used as a medical device.

	LEARNING OUTCOMES	
LESSON ONE	**Introduction to AI for 2D Medical Imaging**	• Explain what AI for 2D medical imaging is and why it is relevant.
LESSON TWO	**Clinical Foundations of 2D Medical Imaging**	• Learn about different 2D medical imaging modalities and their clinical applications • Understand how different types of machine learning algorithms can be applied to 2D medical imaging • Learn how to statistically assess an algorithm's performance • Understand the key stakeholders in the 2D medical imaging space .
LESSON THREE	**2D Medical Imaging Exploratory Data Analysi**	• Learn what the DICOM standard it is and why it exists • Use Python tools to explore images extracted from DICOM files • Apply Python tools to explore DICOM header data • Prepare a DICOM dataset for machine learning • Explore a dataset in preparation for machine learning
LESSON FOUR	**Classification Models of 2D Medical Images**	• Understand architectures of different machine learning and deep learning models, and the differences between them • Split a dataset for training and testing an algorithm • Learn how to define a gold standard • Apply common image pre-processing and augmentation techniques to data • Fine-tune an existing CNN architecture for transfer learning with 2D medical imaging applications • Evaluate a model's performance and optimize its parameters
LESSON FIVE	**Translating AI Algorithms for Clinical Settings with the FDA**	• Learn about the FDA's risk categorization for medical devices and how to define an Intended Use statement • Identify and describe algorithmic limitations for the FDA • Translate algorithm performance statistics into clinically meaningful information that can trusted by professionals • Learn how to create an FDA validation plan

Course 3: Applying AI to 3D Medical Imaging Data

3D medical imaging exams such as CT and MRI serve as critical decision-making tools in the clinician's everyday diagnostic armamentarium. These modalities provide a detailed view of the patient's anatomy and potential diseases, and are a challenging though highly promising data type for AI applications. Learn the fundamental skills needed to work with 3D medical imaging datasets and frame insights derived from the data in a clinically relevant context. Understand how these images are acquired, stored in clinical archives, and subsequently read and analyzed. Discover how clinicians use 3D medical images in practice and where AI holds most potential in their work with these images. Design and apply machine learning algorithms to solve the challenging problems in 3D medical imaging and how to integrate the algorithms into the clinical workflow.

Course Project
Hippocampal Volume Quantification in Alzheimer's Progression

Hippocampus is one of the major structures of the human brain with functions that are primarily connected to learning and memory. The volume of the hippocampus may change over time, with age, or as a result of disease. In order to measure hippocampal volume, a 3D imaging technique with good soft tissue contrast is required. MRI provides such imaging characteristics, but manual volume measurement still requires careful and time consuming delineation of the hippocampal boundary. In this project, you will go through the steps that will have you create an algorithm that will help clinicians assess hippocampal volume in an automated way and integrate this algorithm into a clinician's working environment. First, you'll prepare a hippocampal image dataset to train the U-net based segmentation model, and capture performance on the test data. Then, you will connect the machine learning execution code into a clinical network, create code that will generate reports based on the algorithm output, and inspect results in a medical image viewer. Lastly, you'll write up a validation plan that would help collect clinical evidence of the algorithm performance, similar to that required by regulatory authorities.

	LEARNING OUTCOMES	
LESSON ONE	**Introduction to AI for 3D Medical Imaging**	• Explain what AI for 3D medical imaging is and why it is relevant

LESSON TWO	**3D Medical Imaging - Clinical Fundamentals**	• Identify medical imaging modalities that generate 3D images • List clinical specialties who use 3D images to influence clinical decision making • Describe use cases for 3D medical images • Explain the principles of clinical decision making • Articulate the basic principles of CT and MR scanner operation • Perform some of the common 3D medical image analysis tasks such as windowing, MPR and 3D reconstruction
LESSON THREE	**3D Medical Imaging Exploratory Data Analysis**	• Describe and use DICOM and NIFTI representations of 3D medical imaging data • Explain specifics of spatial and dimensional encoding of 3D medical images • Use Python-based software packages to load and inspect 3D medical imaging volumes • Use Python-based software packages to explore datasets of 3D medical images and prepare it for machine learning pipelines • Visualize 3D medical images using open software packages
LESSON FOUR	**3D Medical Imaging - Deep Learning Methods**	• Distinguish between classification and segmentation problems as they apply to 3D imaging • Apply 2D, 2.5D and 3D convolutions to a medical imaging volume • Apply U-net algorithm to train an automatic segmentation model of a real-world CT dataset using PyTorch • Interpret results of training, measure efficiency using Dice and Jaccard performance metrics
LESSON FIVE	**Deploying AI Algorithms in the Real World**	• Identify the components of a clinical medical imaging network and integration points as well as DICOM protocol for medical image exchange • Define the requirements for integration of AI algorithms • Use tools for modeling of clinical environments so that it is possible to emulate and troubleshoot real-world AI deployments • Describe regulatory requirements such as FDA medical device framework and HIPAA required for operating AI for clinical care • Provide input into regulatory process, as a data scientist

Course 4: Applying AI to Wearable Device Data

Wearable devices are an emerging source of physical health data. With continuous, unobtrusive monitoring they hold the promise to add richness to a patient's health information in remarkable ways. Understand the functional mechanisms of three sensors (IMU, PPG, and ECG) that are common to most wearable devices and the foundational signal processing knowledge critical for success in this domain. Attribute physiology and environmental context's effect on the sensor signal. Build algorithms that process the data collected by multiple sensor streams from wearable devices to surface insights about the wearer's health.

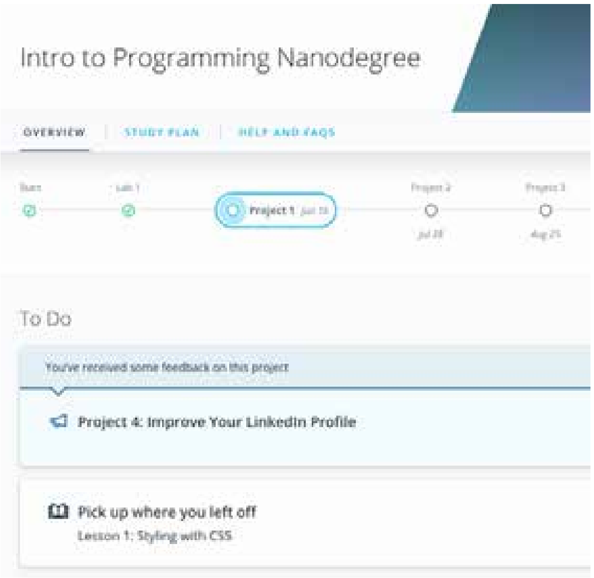
Course Project
Motion Compensated Pulse Rate Estimation

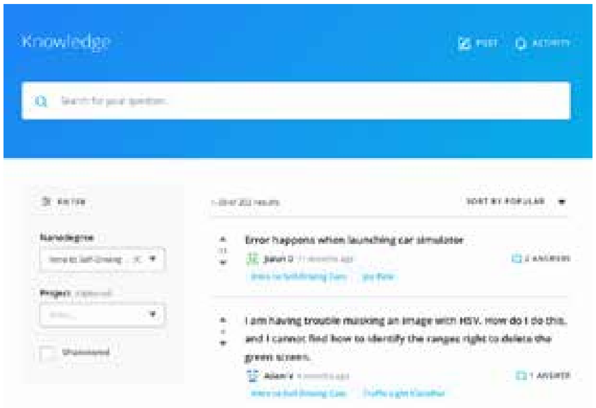
Wearable devices have multiple sensors all collecting information about the same person at the same time. Combining these data streams allows us to accomplish many tasks that would be impossible from a single sensor. In this project, you will build an algorithm which combines information from two of the sensors that are covered in this course -- the IMU and PPG sensors -- that can estimate the wearer's pulse rate in the presence of motion. First, you'll create and evaluate an activity classification algorithm by building signal processing features and a random forest model. Then, you will build a pulse rate algorithm that uses the activity classifier and frequency domain techniques, and also produces an associated confidence metric that estimates the accuracy of the pulse rate estimate. Lastly, you will evaluate algorithm performance and iterate on design until the desired accuracy is achieved.

	LEARNING OUTCOMES	
LESSON ONE	**Intro to Digital Sampling & Signal Processing**	• Describe how to digitally sample analog signals • Apply signal processing techniques (eg. filtering, resampling, interpolation) to time series signals. • Apply frequency domain techniques (eg. FFT, STFT, spectrogram) to time series signals • Use matplotlib's plotting functionality to visualize signals

LESSON TWO	**Introduction to Sensors**	• Describe how sensors convert a physical phenomenon into an electrical one. • Understand the signal and noise characteristics of the IMU and PPG signals
LESSON THREE	**Activity Classification**	• Perform exploratory data analysis to understand class imbalance and subject imbalance • Gain an intuitive understanding signal characteristics and potential feature performance • Write code to implement features from literature • Recognize the danger overfitting of technique (esp. on small datasets), not simply of model parameters or hyperparameters
LESSON FOUR	**ECG Signal Processing**	• Understand the electrophysiology of the heart at a basic level • Understand the signal and noise characteristics of the ECG • Understand how atrial fibrillation manifests in the ECG • Build a QRS complex detection algorithm • Build an arrhythmia detection algorithm from a wearable ECG signal • Understand how models can be cascaded together to achieve higher-order functionality

Our Classroom Experience





REAL-WORLD PROJECTS

Build your skills through industry-relevant projects. Get personalized feedback from our network of 900+ project reviewers. Our simple interface makes it easy to submit your projects as often as you need and receive unlimited feedback on your work.

KNOWLEDGE

Find answers to your questions with Knowledge, our proprietary wiki. Search questions asked by other students and discover in real-time how to solve the challenges that you encounter.

STUDENT HUB

Leverage the power of community through a simple, yet powerful chat interface built within the classroom. Use Student Hub to connect with your technical mentor and fellow students in your Nanodegree program.

WORKSPACES

See your code in action. Check the output and quality of your code by running them on workspaces that are a part of our classroom.

QUIZZES

Check your understanding of concepts learned in the program by answering simple and auto-graded quizzes. Easily go back to the lessons to brush up on concepts anytime you get an answer wrong.

CUSTOM STUDY PLANS

Work with a mentor to create a custom study plan to suit your personal needs. Use this plan to keep track of your progress toward your goal.

PROGRESS TRACKER

Stay on track to complete your Nanodegree program with useful milestone reminders.

Learn with the Best

Nikhil Bikhchandani

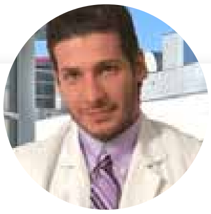
DATA SCIENTIST
AT VERILY LIFE SCIENCES

David Silver leads the School of Autonomous Systems at Udacity. Before Udacity, David was a research engineer on the autonomous vehicle team at Ford. He has an MBA from Stanford, and a BSE in Computer Science from Princeton.

Emily Lindemer

DIRECTOR
AT WELLFRAME

Stephen is a Content Developer at Udacity and has worked on the C++ and Self-Driving Car Engineer Nanodegree programs. He started teaching and coding while completing a Ph.D. in mathematics, and has been passionate about engineering education ever since.



Mazen Zawaideh

RADIOLOGY CONSULTANT AT MICROSOFT RESEARCH - PROJECT INNEREYE | CHIEF RADIOLOGY RESIDENT - UNIVERSITY OF WASHINGTON

Ermin Kreponic is a skilled Java & C++ developer who has taught dozens of online courses in multiple coding languages. Ermin currently works as a cyber-security training architect and is a strong proponent of open-source technologies.

Ivan Tarapov

SENIOR PROGRAM MANAGER - PROJECT INNEREYE, MICROSOFT AI & RESEARCH

Andreas Haja is an engineer, educator, and autonomous vehicle enthusiast. Andreas now works as an engineering professor in Germany. Previously, he developed computer vision algorithms and autonomous vehicle prototypes using C++.

Learn with the Best

Michael Dandrea

PRINCIPAL DATA SCIENTIST | GENENTECH (MEMBER OF THE ROCHE GROUP)

Michael is on the Pharma Development Informatics team at Genentech, where he works on improving clinical trials and developing safer, personalized treatments with clinical and EHR data. Previously, he was a Lead Data Scientist at Change Healthcare. eton.

All Our Nanodegree Programs Include:

EXPERIENCED PROJECT REVIEWERS

REVIEWER SERVICES

- Personalized feedback & line by line code reviews
- 1600+ Reviewers with a 4.85/5 average rating
- 3 hour average project review turnaround time
- Unlimited submissions and feedback loops
- Practical tips and industry best practices
- Additional suggested resources to improve

TECHNICAL MENTOR SUPPORT

MENTORSHIP SERVICES

- Questions answered quickly by our team of technical mentors
- 1000+ Mentors with a 4.7/5 average rating
- Support for all your technical questions

PERSONAL CAREER SERVICES

CAREER SUPPORT

- Resume support
- Github portfolio review
- LinkedIn profile optimization

Frequently Asked Questions

PROGRAM OVERVIEW

WHY SHOULD I ENROLL?

Artificial Intelligence has revolutionized many industries in the past decade, and healthcare is no exception. In fact, the amount of data in **healthcare has grown 20x in the past 7 years**, causing an expected surge in the Healthcare AI market from **$2.1 to $36.1 billion by 2025** at an annual growth rate of 50.4%. AI in Healthcare is transforming the way patient care is delivered, and is impacting all aspects of the medical industry, including early detection, more accurate diagnosis, advanced treatment, health monitoring, robotics, training, research and much more.

By leveraging the power of AI, providers can deploy more precise, efficient, and impactful interventions at exactly the right moment in a patient's care. In light of the worldwide COVID-19 pandemic, there has never been a better time to understand the possibilities of artificial intelligence within the healthcare industry and learn how you can make an impact to better the world's healthcare infrastructure.

WHAT JOBS WILL THIS PROGRAM PREPARE ME FOR?

This program will help you apply your Data Science and Machine Learning expertise in roles including Physician Data Scientist; Healthcare Data Scientist; Healthcare Data Scientist, Machine Learning; Healthcare Machine Learning Engineer, Research Scientist, Machine Learning, and more roles in the healthcare and health tech industries that necessitate knowledge of AI and machine learning techniques.

HOW DO I KNOW IF THIS PROGRAM IS RIGHT FOR ME?

If you are interested in applying your data science and machine learning experience in the healthcare industry, then this program is right for you.

Additional job titles and backgrounds that could be helpful include Data Scientist, Machine Learning Engineer, AI Specialist, Deep Learning Research Engineer, and AI Scientist. This program is also a good fit for Researchers, Scientists, and Engineers who want to make an impact in the medical field.

ENROLLMENT AND ADMISSION

DO I NEED TO APPLY? WHAT ARE THE ADMISSION CRITERIA?

There is no application. This Nanodegree program accepts everyone, regardless of experience and specific background.

FAQs Continued

WHAT ARE THE PREREQUISITES FOR ENROLLMENT?

To be best prepared to succeed in this program, students should be able to:

Intermediate Python:

- Read, understand, and write code in Python, including language constructs such as functions and classes.
- Read code using vectorized operations with the NumPy library.

Machine Learning:

- Build a machine learning model for a supervised learning problem and understand basic methods to represent categorical and numerical features as inputs for this model
- Perform simple machine learning tasks, such as classification and regression, from a set of features
- Apply basic knowledge of Python data and machine learning frameworks (Pandas, Numpy, Tensorflow, Pytorch) to manipulate and clean data for consumption by different estimators/algorithms (e.g. CNNs, RNNs, tree-based models).

IF I DO NOT MEET THE REQUIREMENTS TO ENROLL, WHAT SHOULD I DO?

To best prepare for this program, we recommend the **AI Programming with Python Nanodegree program** and the **Deep Learning Nanodegree program** or **Intro to Machine Learning with PyTorch Nanodegree program**.

TUITION AND TERM OF PROGRAM

HOW IS THIS NANODEGREE PROGRAM STRUCTURED?

The AI for Healthcare Nanodegree program is comprised of content and curriculum to support four projects. Once you subscribe to a Nanodegree program, you will have access to the content and services for the length of time specified by your subscription. We estimate that students can complete the program in four months, working 15 hours per week.

Each project will be reviewed by the Udacity reviewer network. Feedback will be provided and if you do not pass the project, you will be asked to resubmit the project until it passes.

HOW LONG IS THIS NANODEGREE PROGRAM?

Access to this Nanodegree program runs for the length of time specified in the payment card on the Nanodegree program overview page. If you do not graduate within that time period, you will continue learning with month to month payments. See the **Terms of Use** for other policies around the terms of access to our Nanodegree programs.

FAQs Continued

SOFTWARE AND HARDWARE

WHAT SOFTWARE AND VERSIONS WILL I NEED IN THIS PROGRAM?

For this Nanodegree program, you will need a desktop or laptop computer running recent versions of Windows, Mac OS X, or Linux and an unmetered broadband Internet connection. For an ideal learning experience, a computer with Mac or Linux OS is recommended.

You will use Python, PyTorch, Tensorflow, and Aequitas in this Nanodegree program.