**Name:** Hayden Housen

**Year in Course:** 2nd (Junior)

**Mentor:** Dhirai Joshi, IBM

**Title:** Lecture2Notes: An automatic multimodal machine learning system to summarize lecture videos

**Abstract:**

Note-taking is a universal activity among students. Students take notes during lectures to force the active interpretation of the information they are learning. This research focuses on applying extractive and abstractive summarization techniques to transcripts of the visual and auditory content of lectures to create detailed notes. This automated multimodal approach will decrease the time required to create notes, increase quiz scores and content knowledge, and enable faster learning through enhanced previewing. The project is broken into four main components: the slide classifier, the summarization models, the end-to-end-process, and finally the website that enables users to process their videos. The slide classifier is an EfficientNet, although other architectures were tested, trained on tens of thousands of frames from lecture videos. Google's Tesseract project is used to perform OCR on the identified slides and Mozilla's implementation of Baidu's DeepSpeech is used to transcribe the audio.  The process of combining these transcripts is novel and can be accomplished in several ways. For the summarization stage, state-of-the-art models are used, including BART, PreSumm, and novel models specifically for this project, which are collectively called "TransformerExtSum." Extractive and abstractive approaches are combined to summarize the long-form content extracted from the lectures. While the end-to-end process yields promising results, key areas of weakness include the speech-to-text algorithm failing to identify certain words and the summarization models producing sub-par summarizes. These areas provide opportunities for further research.