Application of a Novel Voice Recognition-Based Analytical Classification Model in Analysis of Medical Student Medical Decision-Making During Case Presentation

Daniels, T., Perera, I., Moriarty, S., Chotiner, H., Looney, J., Gittings, K., Rawlins II, F., Center for Simulation and Educational Technology, Edward Via College of Osteopathic Medicine, Blacksburg, VA, 24060, USA

VCOM Edward Via College of Osteopathic Medicine VIRGINIA

References and E-copy of Poster

Introduction and Objectives

In January 2023, the Centers for Medicare and Medicaid Services (CMS) initiated a new standard for clinical assessment and reimbursement in clinical practice, structured medical decision making (MDM)[1]. Medical schools strive to ensure students develop a working basis of medical knowledge for quality patient care and effective interdisciplinary communication[2]. However, for medical schools to produce students of the highest caliber we must provide instantaneous, high-quality feedback in these areas. The development of a novel voice recognition-based lexical analysis model, the Daniels-Rawlins-Perera Lexical Analysis Program (DRPLAP) will provide a bridge between students and direct feedback. This program has the potential to concurrently assess thousands of presentations in seconds, as opposed to the multiple days needed by clinicians who grade by hand. Additionally, early exposure to voice recognition software will benefit students as they transition into their clinical years, as voice recognition has become heavily integrated into modern hospital systems[3,4].

This study aims to:

- Validate the DRPLAP model in its application to analyze medical student MDM during clinical case presentations.
- Comparatively analyze the DRPLAP model against a human-based approach.
- Appraise the functionality of the DRPLAP submission process.

Methods

Seventeen second-year students self-recorded their Integrated Clinical Cases (ICC) case presentations using the proprietary mobile submission application. These recordings were deidentified and will be provided to a faculty grader, who will listen to the audio and analyze presentations according to a standard rubric. These deidentified presentations will also be submitted to the DRPLAP **(Figure 1)**. The processing will involve a program-based analysis of the textual responses for appropriate medical lexicons associated with the case, sequence of medical lexicons, and utilization of non-case associated medical language, or language associated with a different case. Additionally, lexical density will be reported on a scale of 0 to 1, with 1 being the best. The results will be paired against the physician's rating for each metric. Each metric will have its own paired t-test analysis.

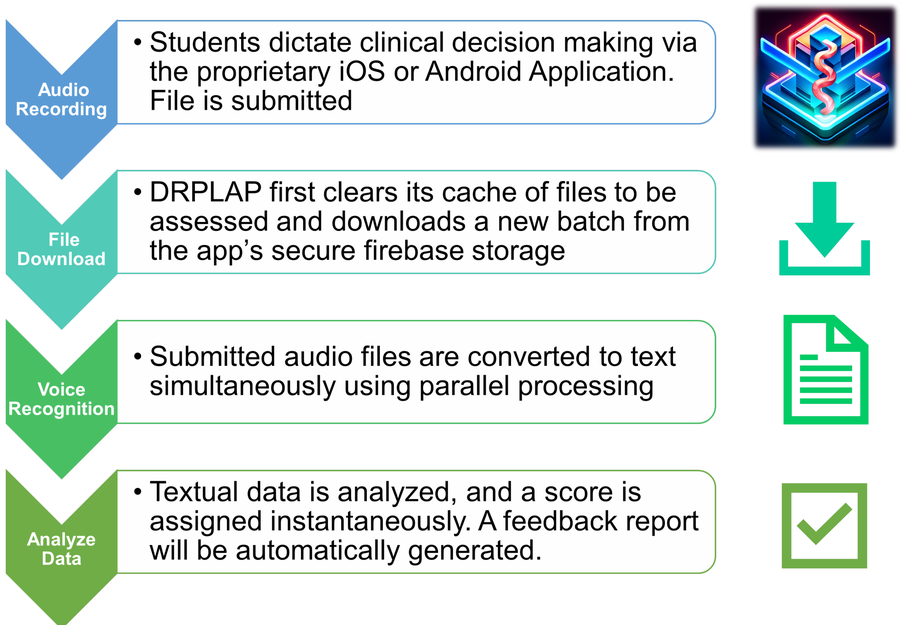

Figure 1: DRPLAP Study Process.

The Model

The DRPLAP Whisper-based model converts audio files to English text by splitting audio into thirty second segments, these segments are fed into an encoder. The encoder portion of the Transformer model reformats the audio segment to a vector representation. This vector representation then gets passed to the decoder part of the model, here the model utilizes context and importance to create weights and affinities as well as positional biases to generate a textual output **(Figure 2)**.

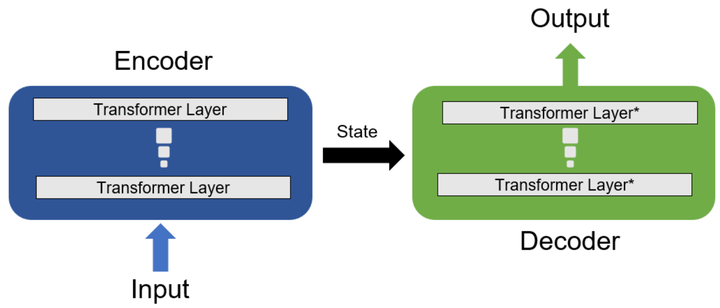

Figure 2: Generalized Transformer Architecture.

De-identified case presentations from previous presentations and publicly available audio files will be used to train a modified Transformer model to specialize in medical terminology. The data will be split into a training and testing set with a 70-30 split. Existing concerns regarding model overfitting will be remedied using a Receiver Operating Characteristic (ROC) curve, which shows the performance of the model at all thresholds and Jaccard Similarity as a benchmark for model performance and hyper parameter optimization. This is a supervised approach as the audio files will have the exact textual translation provided as a key, this will train our model to reach the desired outputs.

Post Model Evaluation and Lexical Analysis

After a final model has been selected, that model will replace the existing generic Transformer Model that is currently used in the DRPLAP app for voice recognition and audio conversion. Functionality testing will be done to ensure successful integration. A new set of participants will be selected via random sampling from recorded case presentations per curricular requirements in Block 4, utilizing approximately 20% of presentations from the class.

Statistical Analysis

Audio recordings will be translated to text and then graded by the lexical analysis algorithm. A physician will also grade the case presentation using the same keyword rubric the algorithm uses. The results will be analyzed using a paired t-test, where we expect to achieve a non-significant p-value. Here, a non-significant p-value means the Translation and Keyword algorithm performs the same as a Physician grader, the gold-standard for this scenario.

Translational Challenges

Whisper, the foundational audio-to-text model integrated into DRPLAP, was trained on 680,000 hours of data from multilingual sources across the internet[5]. However, the generalized use of transcripts and the inclusion of 117,000 hours of non-English data leads to a model that is not specific to our use-case[5]. While the initial implementation of the program performed adequately in medical terminology it struggled in the interpretation of specific medical terminology delivered in varying accents and verbal mannerisms. Finalizing our service with this current model would lead to improper assessment of student lexical density due to aberrant words generated by the model, an unacceptable short-coming of a model designed for evaluation and feedback.

Future Developments

Training the model to become even more specialized for our application should produce precise results. We intend to process these files using the updated model once the training is complete. The long-term goal is to provide a seamless experience to our students, in which students will utilize the mobile submission application to dictate their presentations. These presentations will submit to the DRPLAP, which will almost instantaneously provide feedback to the student.