A. Suplementary Materials

A.1. Fair Use

We strictly followed the criteria of Fair Use by The U.S. Copyright Office[1], which also applies to YouTube platform. Section 107 of the Copyright Act provides the statutory framework for determining whether something is a fair use and identifies certain types of uses—such as criticism, comment, news reporting, teaching, scholarship, and research—as examples of activities that may qualify as fair use. Section 107 calls for consideration of the following four factors in evaluating a question of fair use:

- (1) **Purpose and character of the use, including whether the use is of a commercial nature or is for nonprofit educational purposes:** Courts look at how the party claiming fair use is using the copyrighted work, and are more likely to find that nonprofit educational and noncommercial uses are fair. Additionally, “transformative” uses are more likely to be considered fair. Transformative uses are those that add something new, with a further purpose or different character, and do not substitute for the original use of the work.

- (2) **Nature of the copyrighted work:** This factor analyzes the degree to which the work that was used relates to copyright’s purpose of encouraging creative expression. Thus, using a more creative or imaginative work (such as a novel, movie, or song) is less likely to support a claim of a fair use than using a factual work (such as a technical article or news item). In addition, use of an unpublished work is less likely to be considered fair.

- (3) **Amount and substantiality of the portion used in relation to the copyrighted work as a whole:** Under this factor, courts look at both the quantity and quality of the copyrighted material that was used. That said, some courts have found use of an entire work to be fair under certain circumstances. And in other contexts, using even a small amount of a copyrighted work was determined not to be fair because the selection was an important part—or the “heart”—of the work.

- (4) **Effect of the use upon the potential market for or value of the copyrighted work:** Here, courts review whether, and to what extent, the unlicensed use harms the existing or future market for the copyright owner’s original work.

According to the law, we assert our defense under the Fair Use doctrine with the help of Fair Use explanation[2] by copyrightalliance.org and ELRC Report on legal issues in web crawling [3] by Pawel Kamocki as follows:

- (1) Obviously we crawled the data and published only for non-commercial and research purposes.

- (1) We did not directly use videos crawled from YouTube. Instead, we transformed them into audio files with a predefined sampling rate. Additionally, we divided lengthy audio files, approximately one hour in duration, into shorter segments lasting between 10 to 30 seconds. These segments were then randomly shuffled, making it impossible for users to piece them together to comprehend the entirety of the originally crawled videos. Therefore, our work is transformative and we do not substitute the original use of the crawled videos.

- (2) Our medical conversations are factual (nonfiction) and hence qualified as fair.

- (2) Videos on YouTube platform are universally accessible around the world, therefore we satisfy the criteria for the copyrighted work’s publication status.

- (3) There is no quantitative test to evaluate whether a given use is fair. The randomly shuffled 10-30 second segments we have created do not provide the complete context and meaning of each video, thus making them incapable of representing the ”heart” of the copyrighted work.

- (4) We don’t utilize our publicly available data to compete with the copyright owners’ business. Furthermore, our 10-30 second segments have no impact on the viewership count on YouTube. As a result, our efforts do not undermine the potential market being pursued by the copyright owners.

Besides our work, several similar works exist that involve the extraction of YouTube videos and their conversion into audio files for research and non-commercial intentions, such as GigaSpeech[4] (China & USA), VoxCeleb[5] (UK), VoxLingua107[6] (UK).

[1] https://www.copyright.gov/fair-use/

[2] https://copyrightalliance.org/faqs/what-is-fair-use/

[3] http://www.elra.info/media/filer_public/2021/02/12/elrc-legal-analysis-webcrawling_report-v11.pdf

[4] https://github.com/SpeechColab/GigaSpeech

[5] https://www.robots.ox.ac.uk/ vgg/data/voxceleb/

[6] https://bark.phon.ioc.ee/voxlingua107/

	VietMed-Train	*VietMed-Dev*	*VietMed-Test*
Dur. [hours]	5	5	6
#Speakers	13	21	27
#Words	70k	69k	76k
#Rec. cond.	2	4	6
#Accents	3	4	5
#Roles	3	4	6

Table 1: Data statistics of *VietMed-L*, retrieved from file ”Metadata” in the dataset

A.2. Extra Data Statistics for Labeled Medical Data *VietMed-L*

Table 1 shows the statistics of 3 train-dev-test subsets in *VietMed-L*. We split these 3 subsets in a way that made *VietMed-Train* the least generalizability by having the least number of speakers, recording conditions, accents and roles, while prioritizing *VietMed-Dev* and *VietMed-Test* more generalizability. Note that no speaker overlap occured in the 3 subsets.

A.3. Extra Data Statistics for Unlabeled Medical Data *VietMed-U*

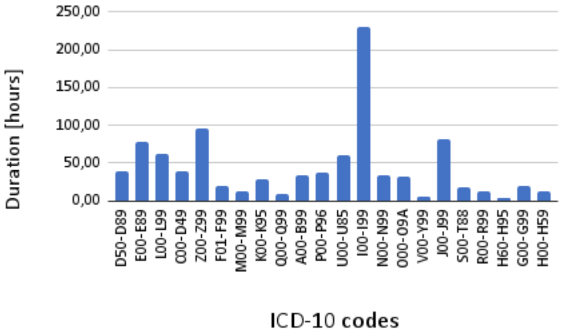

Figure 1: Distribution of ICD-10 code in *VietMed-U*

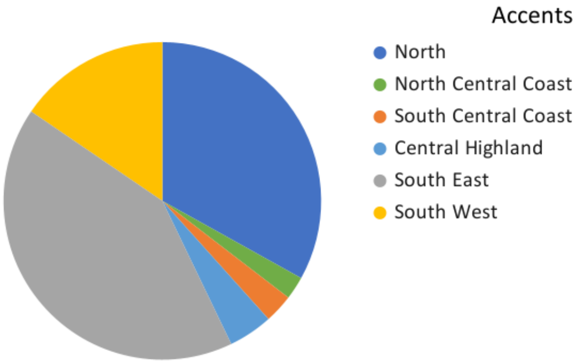

Figure 2: Distribution of accents in *VietMed-U*

Figure 1 shows the distribution of ICD-10 code and Figure 2 shows the distribution of accents in *VietMed-U*. We collected *VietMed-U* in a manner similar to *VietMed-L*, assuring a comparable generalizability as in *VietMed-L*.