APPENDIX

A. Performance of Original Script and English

Table III shows the performance on original script (Hindi/Malayalam) and Table IV shows the performance on English.

Test Language →	**hi**	**ml**
mBERT_{en}	54.34	54.07
mBERT_{tr}	60.11	66.20
mBERT-Joint-TS	61.35	66.24
XLM-R_{en}	60.82	78.11
XLM-R_{tr}	59.72	71.40
XLM-R-Joint-TS	62.23	76.46†

TABLE III
EVALUATION (WEIGHTED F1) ON HINDI AND MALAYALAM ORIGINAL SCRIPT SHOWING THAT OUR MODEL PRESERVES OR OUTPERFORMS THEIR PERFORMANCE ON THE BASELINES EXCEPT FOR MALAYALAM ON XLM-R. †: SEE DISCUSSION.

Train Language → **Model**	**en** (Baseline)	**hi** (Joint-TS	**ml**† Model)
mBERT	81.06	82.35	71.25
XLM-R	83.47	84.00	74.16

TABLE IV
EVALUATION (WEIGHTED F1) ON IMDB ENGLISH TEST DATA SHOWING THAT OUR MODEL PRESERVES THE PERFORMANCE ON THE SOURCE LANGUAGE FOR HINDI BUT NOT FOR MALAYALAM. THE RESULTS IN HI AND ML ARE WITH OUR MODIFICATIONS. †: SEE DISCUSSION.

B. Model Runtime

As an addendum to the performance evaluation shown in Table II, we also provide the runtime of our Joint-TS model and key baselines in Table V. The mBERT_{en} and XLM-R_{en} model are trained only on English data without any augmentation while $\text{mBERT}_{\text{en}+tr+tl}$, $\text{XLM-R}_{\text{en}+tr+tl}$, and the Joint-TS models are trained using translated and transliterated target in addition to English data. The additional latency is caused due to the augmented data (two times more data). Our Joint-TS model also consists of unsupervised optimization for alignment, in addition to the augmented data. An interesting observation is that the Teacher-Student model based on mBERT converges faster than its $\text{en}+tr+tl$ counterpart and XLM-R, while also having a comparable performance as shown in Table II.

C. Ethical Considerations

The tweets extraction procedure followed the Twitter Terms of Service and did not violate privacy policies of individual users. Also, the datasets we share include only Tweet IDs in the public domain. Data statement that includes annotator guidelines for the labeling jobs and other dataset information will be provided with the implementation. From a broader impact perspective, our code is open-source and allows NLP technology to be accessible to information systems for emergency services and social scientists in studying a large population in India who use transliterated text for communication in everyday life.

Model →	mBERT_{en}	XLM-R_{en}	$\text{mBERT}_{en+tr+tl}$	$\text{XLM-R}_{en+tr+tl}$	mBERT-Joint-TS	XLM-R-Joint-TS
Runtime →	00:35:55	00:55:53	02:38:39	02:40:48	01:30:48	04:39:16

TABLE V
RUN TIME ON A SINGLE K80 GPU IN HH:MM:SS FOR TRAINING A MALAYALAM MODEL ON CRISIS DATA.