eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
ZIENTZIA ETA TEKNOLOGIA FAKULTATEA
FACULTAD DE CIENCIA Y TECNOLOGÍA

Classifier chain based multitask model

This appendix of the work refers to a multitask model based on the sklearn ClassifierChain function. As it can only detect if a parameter is involved in a sample, it has no application for arousal, and thus it was not included in the main work as a multitask model. However, this function tries to make a greater connection between the different outputs, not only computing it with the same body, but also basing the results of the latter classifiers on the previous ones, and thus could fit the data better than a regular model.

The same configuration used for the multitask models in text[1], audio and multi-modal were used, but instead of the MultiOutputClassifier, ClassifierChain was used. The base model was the previously fine-tuned multi-layer perceptron. Thus, with this model, illness and valence could be computed at the same time with the three input options.
The other caveat was which label to compute first. As illness detection has a much better efficiency throughout the main work, this was initially chosen to be the main estimator, but the inverse configuration, computing first valence and then illness, was also considered, and actually proved to have better results, in general.

The results for all six experiments (text, audio and multi-modal data with the two different classifier orders) are summarized in Table 1.

		First illness, then valence		First valence, then illness	
		Validation	Test	Validation	Test
Text	Illness	0.70	0.69	0.73	0.56
	2-labeled valence	0.68	0.50	0.65	0.52
Audio	Illness	0.84	0.85	0.86	0.89
	2-labeled valence	0.68	0.61	0.68	0.64
Multi-modal	Illness	0.89	0.84	0.88	0.89
	2-labeled valence	0.74	0.60	0.70	0.62

Table 1: Results for the classifier chain models with different inputs and configurations

The confusion matrices corresponding to the best classification, chosen to be the audio-based and firstly valence-computing one, are shown in Figures 1, 2, 3 and 4.
The results shown in the table suggest, on the one hand, a generally bad interpretation for the text data (which was also true in the main work), and on the other hand, a preference to

[1]As SMOTE could not be applied, no data was removed even if the length was one.

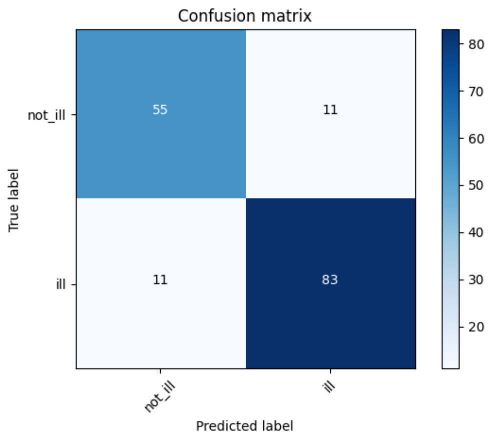

Figure 1: Audio-based classifier chain (valence first): illness detection (validation)

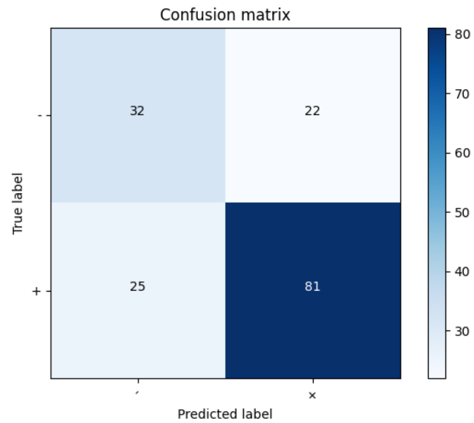

Figure 2: Audio-based classifier chain (valence first): valence classification (validation)

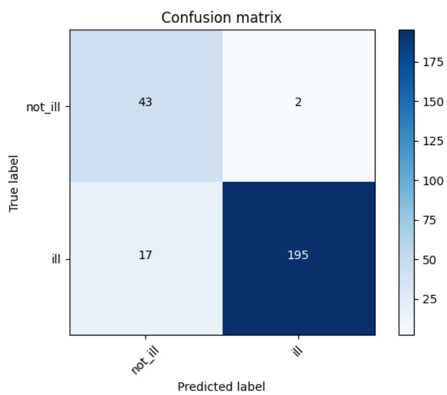

Figure 3: Audio-based classifier chain (valence first): illness detection (test)

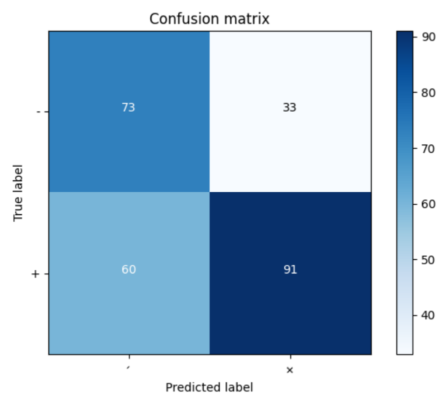

Figure 4: Audio-based classifier chain (valence first): valence classification (test)

train valence first and illness second. This is an interesting conclusion and one that definitely should be taken into account for future works. It can also be seen that audio data shows the best results, even surpassing the valence efficiencies obtained in the main work. Moreover, the model shows great resemblance in validation and test, both in audio and in multi-modal processing, and thus the overfitting and/or the inaccurate data problem seems to be somehow reduced. In general, this model shows great potential and should be considered.