Transfer performance between English and Sepedi, Sesotho, and Setswana on news headlines classification shows dissatisfactory results.

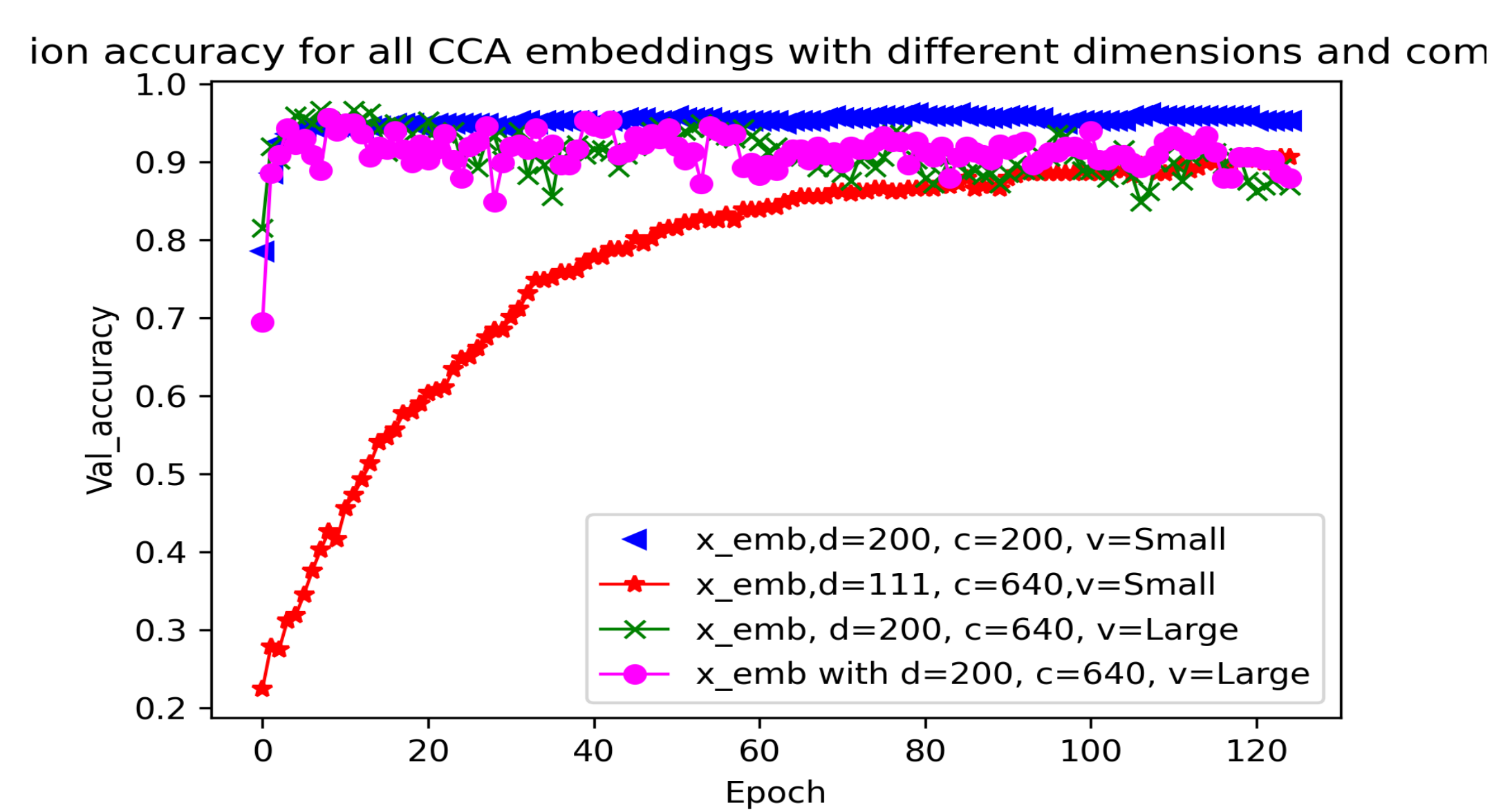## Cross-lingual models and application for South African low resourced languages.

INTRO

Cross-lingual(CL) models have been in active research over the past few years for mitigating issues in resource-deprived languages. South African languages, amongst many others, have not witnessed such benefits. For this we aim to survey and critique available CL models , conduct experiments, and perform zero-shot news headlines classification

METHODS

1. Surveyed CL models
2. Conducted experiments on two of the models
3. Evaluated models on classification task

RESULTS

DISCUSSION

- It is challenging to train monolingual embeddings
- Due to the above difficulty, extracting useful embeddings is exacerbated.
- More work on CL models is necessary

Data Science for Social Impact

VecMap Emb

CCA Emb

Thapelo Sindane, Vukosi Marivate, Abiodun Modupe
sindane.thapelo@tuks.co.za
Data Science for Social Impact Research Group
Department of Computer Science
University of Pretoria
https://dsfsi.github.io/

Department of Computer Science

Faculty of Engineering, Built Environment and Information Technology

Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie / Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA