The use of deep learning solutions to develop a practice tool to support Lámh language for communication partners

Master of Science in Data Analytics – CCT College Dublin

Author: Gabriel Bueno Pimentel Borges

Student Number: SBA21590

Supervisor: Dr. Vladimir Milosavljevic

What is Lámh?

('Hand' in Irish)

Sigafoos *et al.* (2014) and Wilkinson and Hennig (2007)

Augmentative and Alternative Communication (AAC) – Speech-language research for those who cannot communicate through verbal language

Tan *et al.* (2014)

Key Word Signing (KWS) - communication through manual signing alongside the spoken word

Byrne, Pyne and Sheehan (2019)

Lámh is the KWS language practiced in Ireland

Frizelle and Lyons (2022)

Lámh vocabulary is made of 580 different signs originally from the Irish Sign Language

Dolly and Noble (2018)

Communication partners (CP) play an important role by bringing key word signing to the natural setting of the individual

Byrne, Pyne and Sheehan (2019)

Lack of training and practice interferes in the proper application of key word signing

Research question

What solutions could be proposed on a data analytics perspective to assist people who use or must learn Lámh language – or even other KWS languages?

Relevance

"Controlled Practice and Feedback“ – Deep Neural Network model is able to classify Lámh signs

Communication partner (CP) should have the possibility to train the learnt signs more often

CP acquires confidence and knowledge to be able to use the language in a natural environment

Communication support can also provide a wide variety of benefits to one's inclusion in society

Contribution

Lámh Data Acquisition

Comparison between CNN, LSTM and SVM for Lámh sign classification

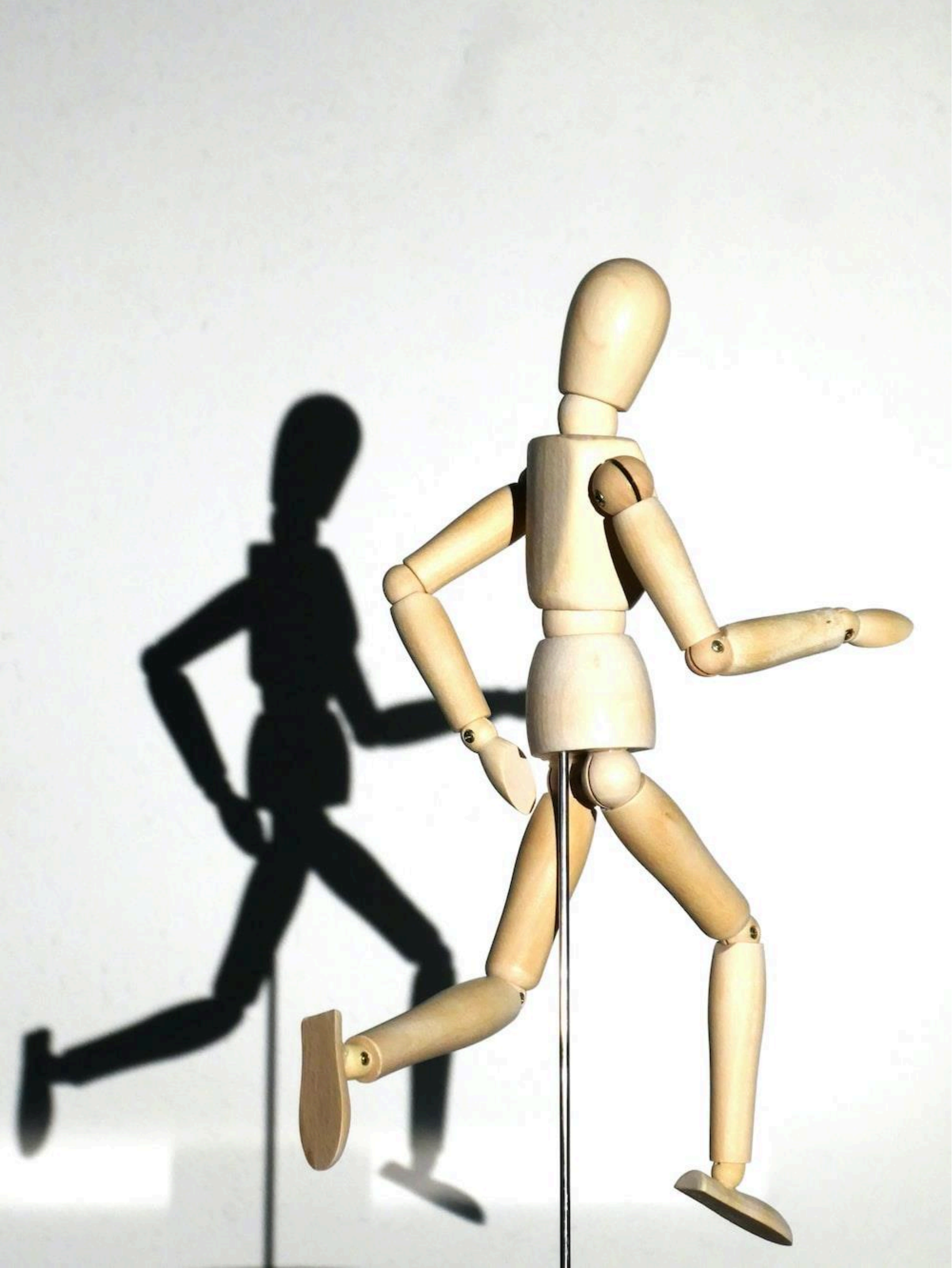
Objectives

Primary Objective

To obtain a highly accurate and generalised machine learning model that must effectively predict the movements of users through real time detection, labelling these in accordance to the respective meaning within the list of signs of the Lámh language.

Secondary Objectives

- Data collection;
- Data augmentation;
- Machine learning model training and comparison;
- Final artifact.



Methodology

40 Frames

45 repetitions per sign (15 on close distance to camera, 15 on medium distance and 15 on long distance)

20 Signs (Frizelle and Lyons, 2022)

Vocabulary
PLAY, TO
LOOK, TO
SIT, TO
GO, TO
YOU
GOOD
WHAT?
TIME
THANK YOU
BOOK
HELLO/HOW ARE YOU?
WANT, TO
I/ME
GIRL
BOX
SHOW, TO
TABLE
JIGSAW
GAME
PLEASE

MediaPipe

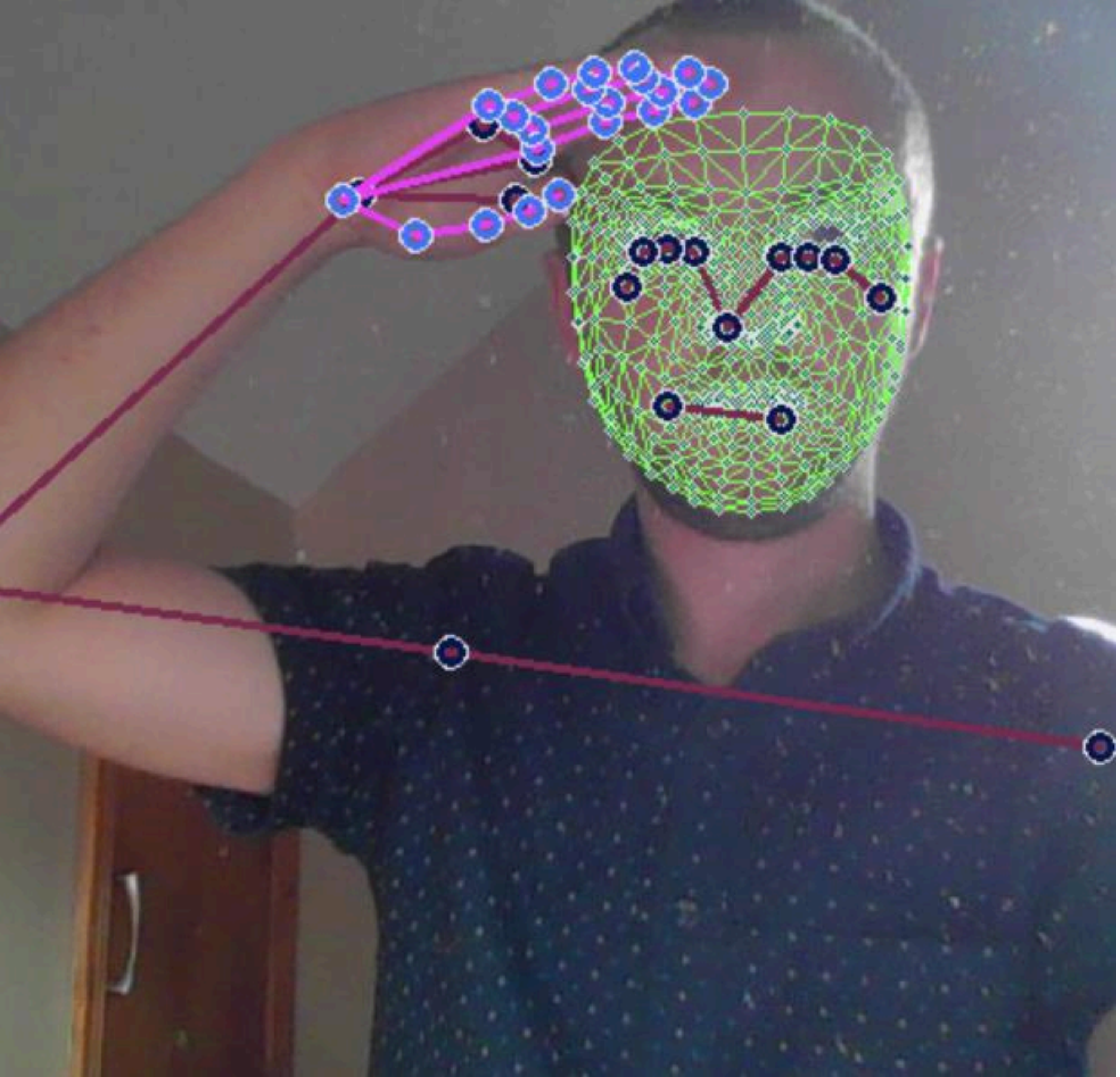
Left Hand: 21 landmarks with x, y and z coordinates

Right Hand: 21 landmarks with x, y and z coordinates

Pose: 33 landmarks with x, y, z coordinates and visibility

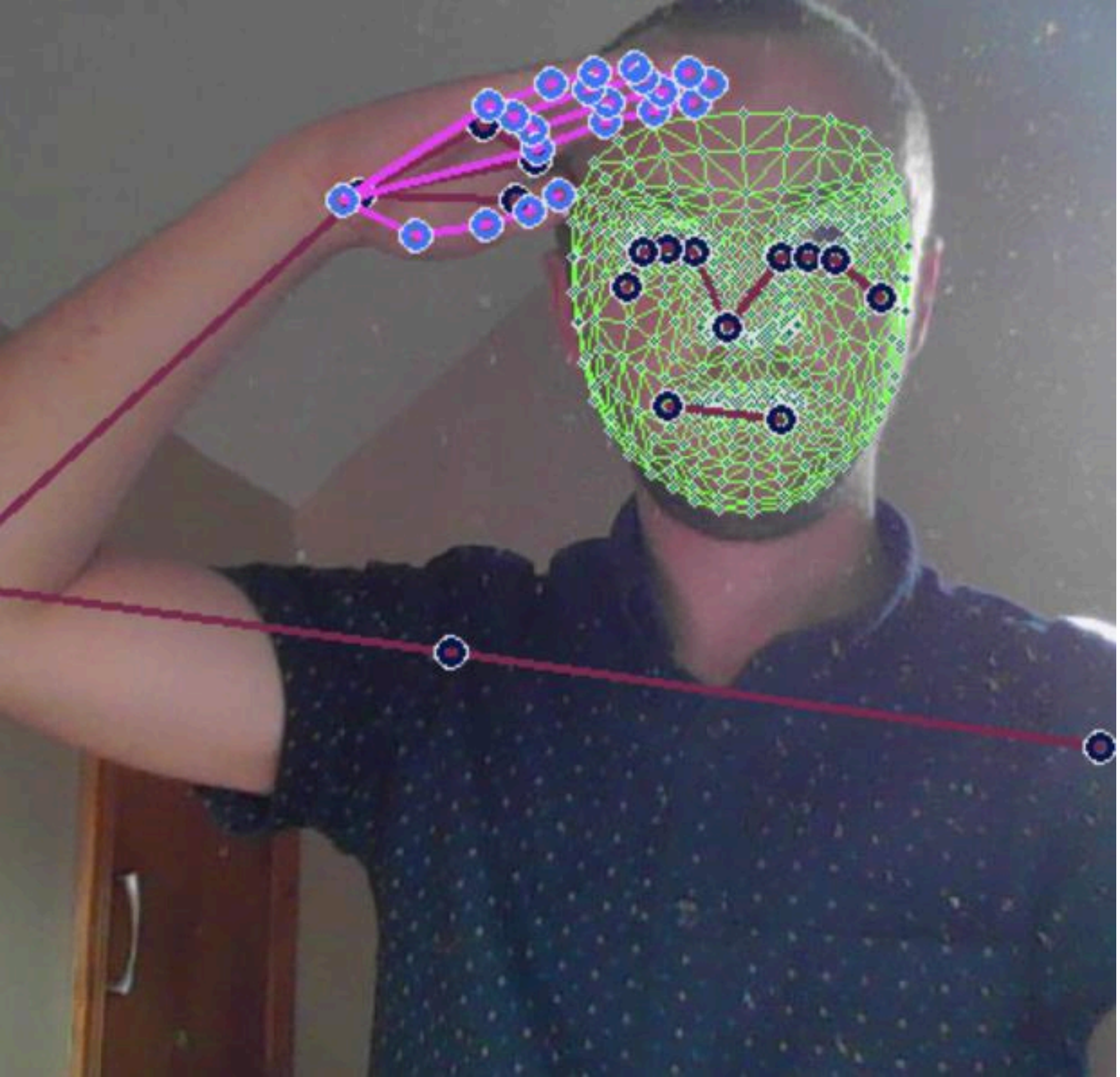
Face: 468 landmarks with x, y and z coordinates

Final array: 900 x 40 x 1,662



Methodology - Models

- 1-D CNN layer (32 nodes), Dense layer (32 nodes);
- Two 1-D CNN layers (32 nodes and 64 nodes), Dense layer (32 nodes);
- Two 1-D CNN layers (32 nodes and 64 nodes), 0.2 Dropout, Dense layer (32 nodes);
- Three 1-D CNN layers (32 nodes, 64 and 128 nodes), Dense layer (32 nodes);
- Single LSTM layer (32 nodes), Dense layer (32 nodes);
- Two LSTM layers (32 nodes and 64 nodes), Dense layer (32 nodes);
- Two LSTM layers (32 nodes and 64 nodes), Two Dense layers (32 nodes and 64 nodes);
- Three LSTM layers (32 nodes, 64 and 128 nodes), Dense layer (32 nodes);

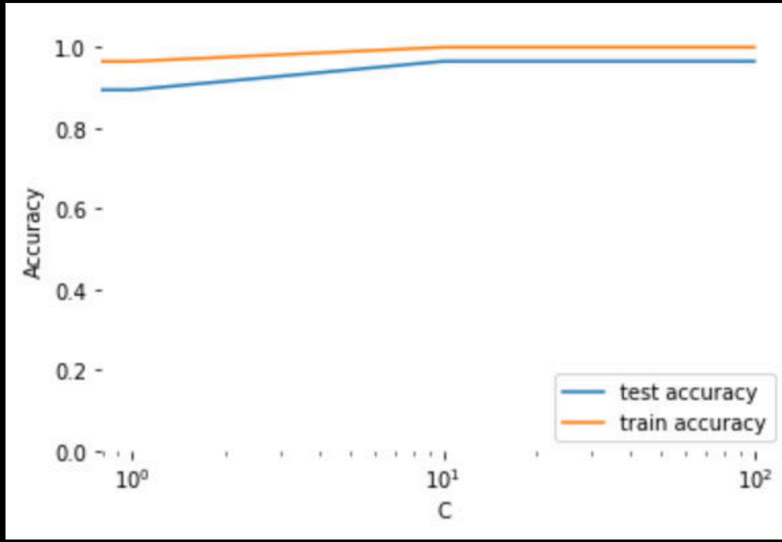


Methodology - Models

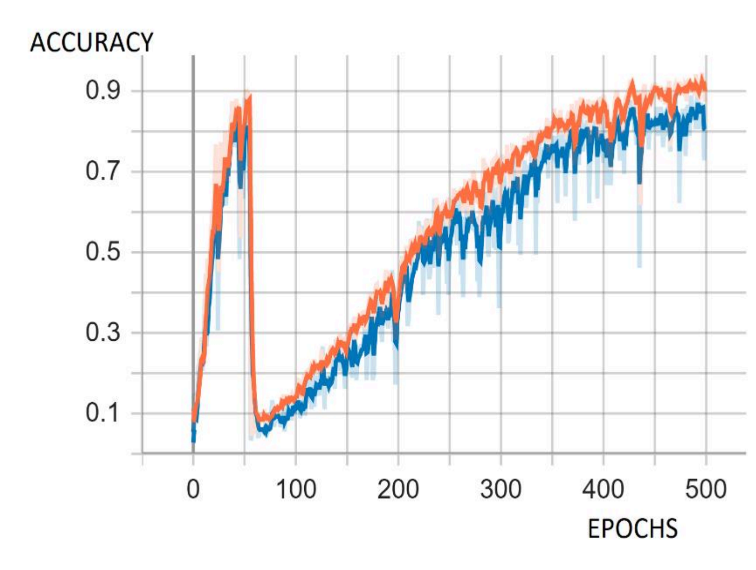
- 1-D CNN layer (32 nodes), LSTM layer (32 nodes), Dense layer (32 nodes);
- 1-D CNN layer (32 nodes), 0.2 Dropout, LSTM layer (32 nodes), 0.2 Dropout, Dense layer (32 nodes), 0.2 Dropout;
- 1-D CNN layer (32 nodes), 0.2 Dropout, LSTM layer (32 nodes), Dense layer (32 nodes);
- Two 1-D CNN layers (32 nodes and 64 nodes), Two LSTM layers (32 nodes and 64 nodes), Dense layer (32 nodes);
- Two 1-D CNN layers (32 nodes and 64 nodes), 0.2 Dropout, Two LSTM layers (32 nodes and 64 nodes), Dense layer (32 nodes);
- SVM – Cross Validation with 10 Folds
  - Gamma: [10, 1, 0.1, 0.01, 0.001];
  - Kernel: [poly, rbf, sigmoid, linear]
  - C: [0.1, 1, 10, 100]

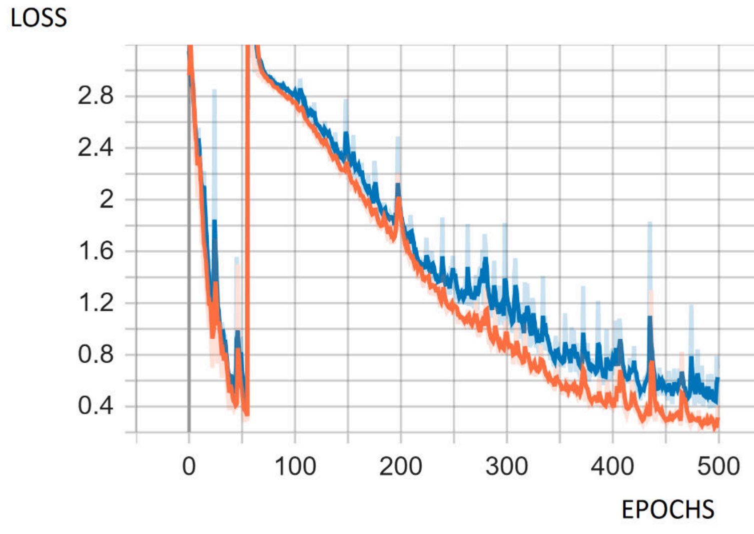
Results

SVM RBF
Gamma = 0.001

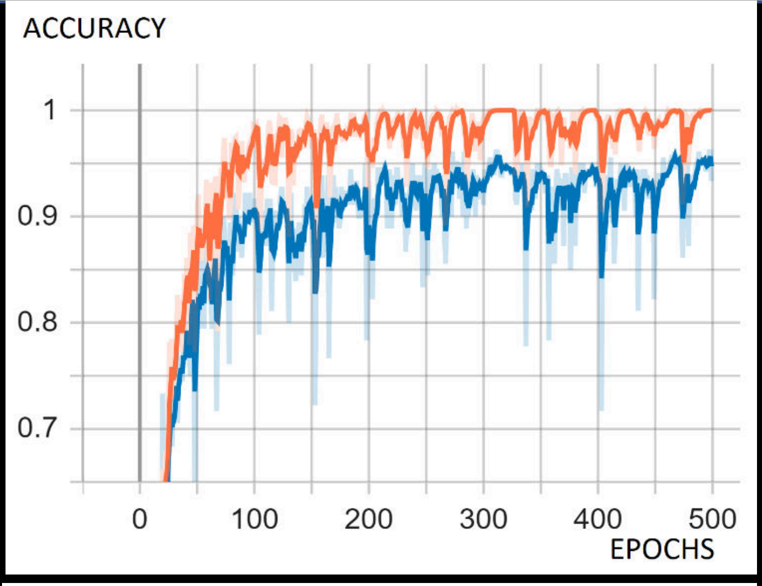


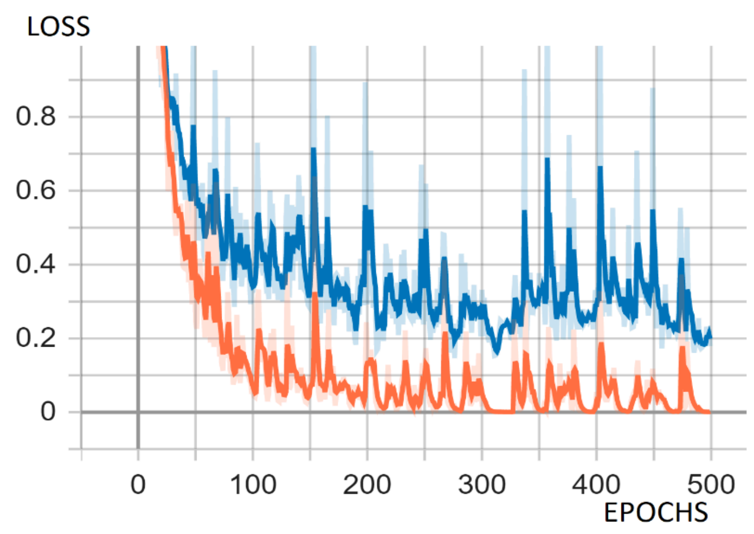
Double LSTM
(32 and 64 nodes)





Double CNN (32 and 64 nodes) + 0.2 Dropout + Double LSTM (32 and 64 nodes)





train
validation

Results

Training and Testing Accuracy and Loss – Best Models

Model	Training Accuracy	Testing Accuracy	Training Loss	Testing Loss
Double CNN + Dropout + Double LSTM	99.86%	93.33%	0.0035	0.1791
Double LSTM	87.36%	80.00%	0.4069	0.6955
SVM RBF Gamma = 0.001 C = 10	100%	96.5%	-	-

Classifications in Real-time Detection

Model	Detected	Detected with reservations	Not Detected
Double CNN + Dropout + Double LSTM	8	6	6
Double LSTM	3	10	7
SVM RBF Gamma = 0.001 C = 10	8	3	9

Results

Classifications in Real-time Detection

Double CNN + 0.2 Dropout + Double LSTM

Sign	Detection Status	Comments
To play	Detected	
To look	Detected with reservations	Detected easily when the angle of the webcam is greater than 45 degrees, if the angle is 45 degrees or less, the model picks the word 'hello'
To sit	Detected	
To go	Detected	
You	Not detected	The model tends to pick either 'please' or 'to want', which are not similar models
Good	Not detected	The model tends to pick either 'please' or 'to want', which are not similar models
What	Not detected	The model chooses other two-hand signs, 'to play' or 'to show'
Time	Detected with reservations	Detected in a specific position. Until this position is found, other two-hand signs are detected, like 'to play', 'to show' or 'box'
Thank you	Detected	
Book	Detected	
Hello	Detected	
To want	Not detected	The model's probabilities are distributed between 'to sit', which is a very similar sign, and 'to play' another two-hand sign
Please	Detected	

Sign	Detection Status	Comments
Me	Detected with reservations	Since the model tends to pick 'me' when there is no movement of the hands at chest level, this detection may be compromised
Girl	Not detected	The model tends to pick the sign 'hello'
Box	Detected with reservations	Detected when the laptop angle is greater than 45 degrees
To show	Detected	
Table	Detected with reservations	Detected when the laptop angle is equal or less than 45 degrees
Jigsaw	Not Detected	Other two-hand signs chosen like 'to show' or 'to play'
Game	Detected with reservations	Detected when the angle of the laptop is equal or less than 45 degrees

Future Work

- Additional data collection is required to differentiate some similar signs;
- More comprehensive data is required adding more hips and legs landmarks;
- Data augmentation involving small coordinate offsets;
- A solution must be provided to stabilize the real-time classifications (e.g.: assess the impact of a “no action” classification).

Achievements

- 8 signs easily detected + 6 detected with reservations;
- Creation of a data collection and machine learning processing methodology that is able to read the body movements related to Lámh language;
- Real-time detection assessment to differentiate the best models;
- Exchange of knowledge with professionals of the Lámh language, education, speech therapy and social care sectors.

References

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV*. Beijing: O'Reilly.

Byrne, Á., Pyne, J. and Sheehan, V. (2019). Use of key word signing for children and adults with intellectual disability in an Irish context. *Tizard Learning Disability Review*, 24(3), pp.113–120.

Dolly, A. and Noble, E. (2018). ‘Lámh Signs Combined’ – Investigating a Whole School Approach to Augmentative and Alternative Communication (AAC) Intervention Through Research in Practice. *REACH: Journal of Inclusive Education in Ireland*, 31(1), pp. 53–68.

Frizelle, P. and Lyons, C. (2022). The development of a core key word signing vocabulary (Lámh) to facilitate communication with children with down syndrome in the first year of mainstream primary school in Ireland. *Augmentative and Alternative Communication*, pp.1-14.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.L., Yong, M.G., Lee, J. and Chang, W.T. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.

Sigafoos, J., O’Reilly, M., Lancioni, G. and Sutherland, D. (2014). Augmentative and Alternative Communication for Individuals with Autism Spectrum Disorder and Intellectual Disability. *Current Developmental Disorders Reports*, 1(2), pp.51–57.

Tan, X., Trembath, D., Bloomberg, K., Iacono, T. and Caithness, T. (2014). Acquisition and generalization of key word signing by three children with autism. *Developmental Neurorehabilitation*, 17(2), pp.125-136.