Covid-19 Fake News Detection using RNNs

Andrei Meșină, Raluca Ioniță, Cristina Vultur and Florin Brad*
University of Bucharest, Romania
*Bitdefender, Romania

andreimesina98@gmail.com, raluca.ionita04@gmail.com, c.vultur@yahoo.com, fbrad@bitdefender.com

UNIVERSITY OF BUCHAREST
VIRTUTE ET SAPIENTIA

1. Introduction

Motivation: In the era of fast-travelling information, news about the ongoing pandemic are spreading even faster than the virus itself. Social Media is playing a huge role in informing us nowadays, which is generally a very helpful thing. Having news, funny memes, cat pictures and your friend's posts in the same place sound wonderful, doesn't it? The bad part, though, is that when it comes to sensitive subjects, nowadays Covid-19 and its effects, the more shocking the more likely we are to believe it. But how can we filter the news that we are being fed to constantly?
Proposal: We have made it our target to help people filter the information that they can trust. We have analyzed tweets about Covid-19 and the pandemic using neural networks to separate them into real ones, that we can trust, and fake ones, which should be reported. This little project ca be very useful in times like those that we live, where news are everywhere, but answers nowhere.

2. Dataset and Preprocessing

Dataset Description: We use the training and validation split from the The Contraint@AAAI 2021 Covid-19 Fake news detection dataset, which has media articles and posts acquired from multiple platforms. The possible labels are: "real" -tweets which are from verified sources and give useful information on COVID-19, or "fake" -tweets, posts, articles which make claims and speculations about COVID-19 which are verified to be not true.

Dataset stats	
AVERAGE WORDS PER REAL TWEET	31.74
AVERAGE WORDS PER FAKE TWEET	21.87
AVERAGE CHARS PER REAL TWEET	215.37
AVERAGE CHARS PER FAKE TWEET	143.51
AVERAGE URLS PER REAL TWEET	0.94
AVERAGE URLS PER FAKE TWEET	0.40
AVERAGE HASHTAGS PER REAL TWEET	1.06
AVERAGE HASHTAGS PER FAKE TWEET	0.49

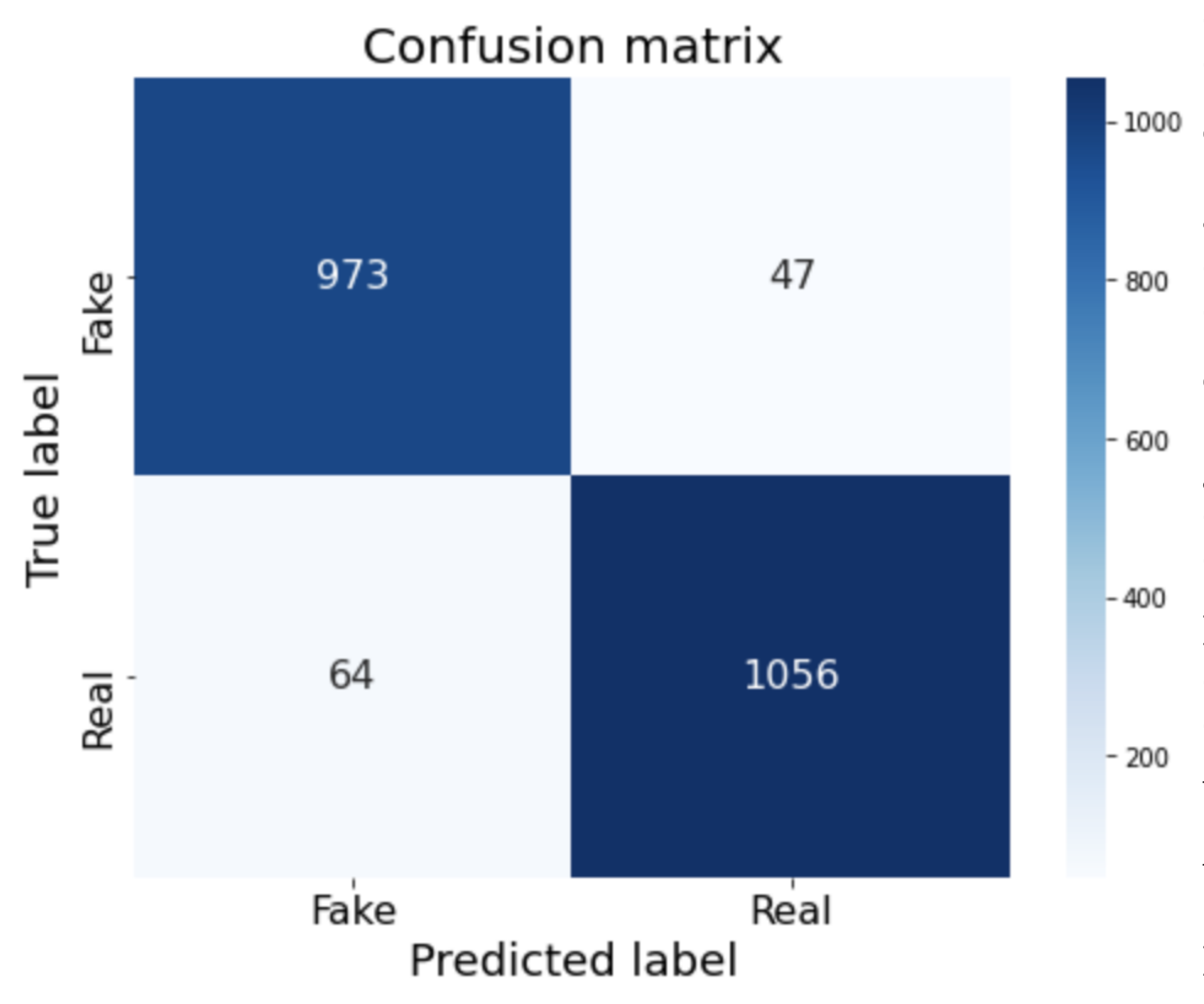
Data preprocessing: The first steps we made to prepare our dataset was to split the tweets into tokens, and to parse the data by expanding contractions and removing special characters, being useful for all the architectures that we tested. For our RNN model, which performed the best, we padded each sample with a <pad> token until the desired length was achieved, and replaced rare words with an <unk> token. Inputs that were too long were cropped by dropping words based on the medium lenght so that we can maintain a balance.

3. Models

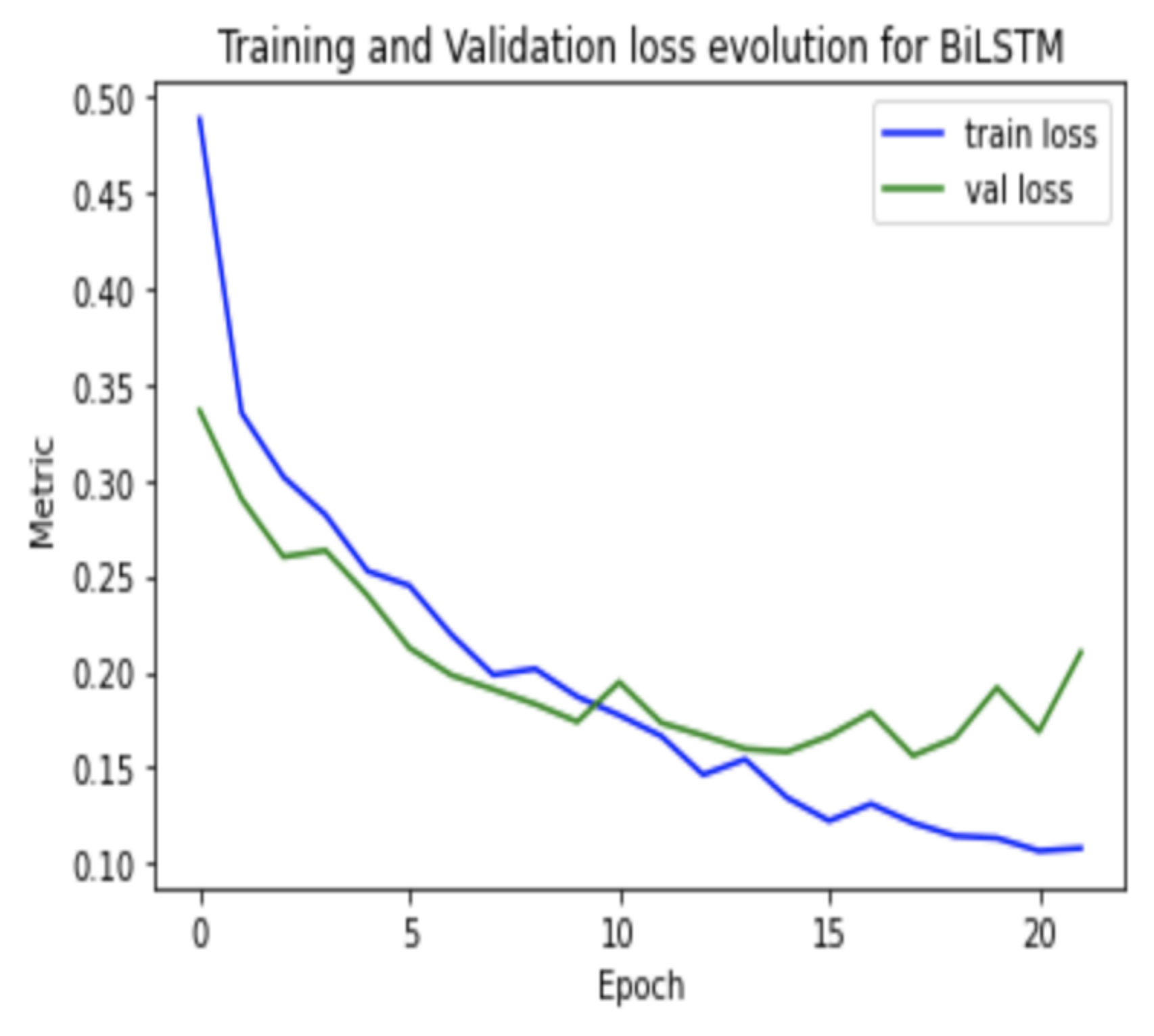
Baseline: We trained a feedforward neural network with one hidden layer where the input is the average of the GloVe word embeddings associated to words in the input sequence. If a word was not represented in GloVe, we have simply eliminated it.
RNN Approaches: We have trained an LSTM model which is using the last hidden layer. The parameters that we chose were Adam optimizer, with a learning rate of 0.001. LSTM has hidden size 64, Tanh activation and trains on 20 epochs.
LSTM with GloVe: We have first given as input to the LSTM both embeddings from GloVe and our own trained ones.
LSTM without GloVe: We have given as input to the LSTM model only embeddings that the network has learned.
BiLSTM with Max-Pooling Over All Hidden States: Left-to-right LSTM misses context on the right side, so we also trained a bidirectional LSTM. We apply max-pooling over all the hidden states and concatenate the resulting vectors from both directions.

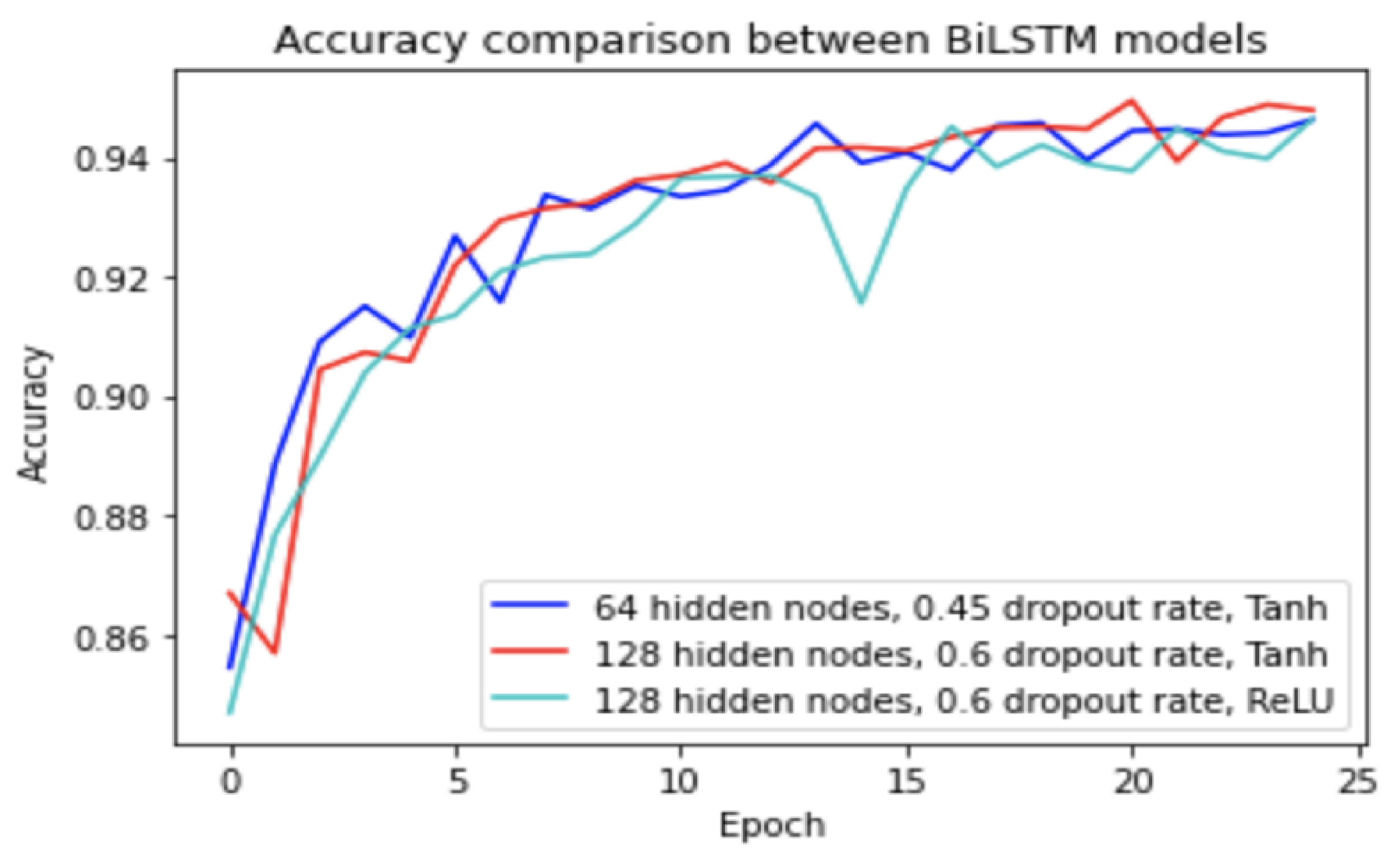
4. Results

Architecture	Accuracy	Precision	F1	Recall
FEED-FORWARD	0.82	0.79	0.83	0.88
LSTM WITH GLOVE	0.93	0.92	0.93	0.93
LSTM WITHOUT GLOVE	0.91	0.91	0.91	0.92
BILSTM MAX-POOLING OVER ALL HIDDEN STATES:	**0.95**	**0.94**	**0.95**	**0.95**



Conf. matrix for BiLSTM all hidden maxpool.
Predictions examples:
Real tweet predicted as real
coronavirus uk records 20 deaths of people with covid 19 and 3991 new daily cases
Real tweet predicted as fake
coronavirus sir keir starmer says second lockdown would be government failure not an act of god
Fake tweet predicted as fake
novel coronavirus affects only elderly not young people
Fake tweet predicted as real
the number of deaths from respiratory illnesses in ireland this year is comparable to previous years and covid 19 is not causing any more deaths than would usually be happening
[h] 0.5





5. Conclusion

We have explored several neural network models that use preprocessed data along with words embeddings. Using a combination of Glove vectors and our own words representations in the embeddings layer offered the best results.
The highest F1 score (0.95) on the real-fake classes has been achieved with a BiLSTM model, as opposed to a F1 score of 0.83 achieved with the baseline model. Making the LSTM model bidirectional and adding MaxPool over all the hidden states brought important improvements (about 6% for loss and 3% for F1 and accuracy). Selecting only the final states which correspond to the endings of sentences did not increase the results of our network.
The confusion matrix shows there are more Real predictions.
Future work involves applying other classes of neural networks (convolutional neural networks, temporal convolutional networks) that might better exploit local clues about the validity of the message.