Proc. of the 8^{th} International Conference on Engineering and Emerging Technologies (ICEET)
27- 28 October 2022, Kuala Lumpur, Malaysia

Automatic Tuberculosis detection in cough patterns using NLP-style cough embedding

Madhurananda Pahar
Department of Electrical Engineering, Stellenbosch University,
Stellenbosch, South Africa.
mpahar@sun.ac.za

Grant Theron
Division of Molecular Biology, Stellenbosch University,
Stellenbosch, South Africa.
gtheron@sun.ac.za

Thomas Niesler
Department of Electrical Engineering, Stellenbosch University,
Stellenbosch, South Africa.
trn@sun.ac.za

***Abstract*—Coughing is a common symptom of respiratory diseases and in the past, its audio and acoustic properties have been used to detect those diseases. In this study, we first show that cough patterns can also be successfully used to detect a respiratory disease such as tuberculosis (TB). For this purpose, we have used the vocal audio recordings of 15 TB and 33 non-TB patients, who were sick from other lung ailments. They were asked to cough, take a few deep breaths and cough again, thus producing at least two bouts of coughs. NLP-style cough embedding was invented to preserve the occurrence and sequence of every cough event with a sampling rate of 1 kHz. In total, almost 2 hours of cough embedding were used as feature vectors to train and evaluate four shallow (LR, SVM, KNN, MLP) and two deep architectures (CNN, LSTM) using a nested cross-validation scheme. Imbalance in our dataset was addressed by applying SMOTE and using AUC as the performance metric. We have also experimentally extracted MFCC, ZCR and kurtosis from the audio recording of the cough embedding and compared the performance of the classifiers while trained on these audio features. The results show that an LSTM performed the best by producing an AUC of 0.81 while using the cough embedding to discriminate TB; whereas a CNN performed the best by producing the highest AUC of 0.71 using the audio features. We show that this way of detecting TB using cough embedding due to its unique pattern preserves privacy, as they do not require the cough audio to be analysed and can be fused as an additional tool to improve the TB cough audio classification even further.**

***Index Terms*—tuberculosis, cough, NLP, machine learning, LSTM**

This research was supported by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Intramural Postdoctoral programme and the Research Capacity Development Initiative as well as the COVID-19 IMU EMC allocation from funding received from the South African National Treasury. We also acknowledge funding from the EDCTP2 programme supported by the European Union (grant SF1401, OPTIMAL DIAGNOSIS; grant RIA2020I-3305, CAGE-TB) and the National Institute of Allergy and Infection Diseases of the National Institutes of Health (U01AI152087).

We would like to thank the South African Centre for High Performance Computing (CHPC) for providing computational resources on their Lengau cluster for this research and gratefully acknowledge the support of Telkom South Africa. We also thank the Clinical Mycobacteriology & Epidemiology (CLIME) clinic team for assisting in data collection, especially Sister Jane Fortuin and Ms. Zintle Ntwana.

The content and findings reported are the sole deduction, view and responsibility of the researcher and do not reflect the official position and sentiments of the SAMRC, EDCTP2, European Union or the funders.

I. Introduction

Tuberculosis (TB) is one of the major respiratory diseases which affects our lungs and breathing systems and is responsible for 95% of deaths due to all respiratory diseases in developing countries [1]. TB is common in poorer countries and the modern diagnostic tests are costly as they rely on special equipment and laboratory procedure [2] and many suspected patients actually do not suffer from TB. Currently, programmatic screening for TB relies on self-reported symptoms, which are non-specific, resulting in vast over-referral for GeneXpert testing, and thereby unnecessary expenditure [3]. Thus, there is a need for a more specific, low-cost, point-of-care screening test which would allow more efficient application of subsequent molecular testing.

Coughing is a common symptom of respiratory disease and its audio and acoustic properties have been successfully used in the past to discriminate respiratory diseases such as TB and COVID-19 [4]–[6]. Here, in this study we show that discrimination is also possible by finding the patterns in which a patient coughs when he is asked to cough, take a few deep breath and cough again by healthcare workers at a recording site inside a healthcare clinic. We propose a new cough embedding technique, which is similar to the word embedding in a natural language processing (NLP) system. We successfully demonstrate that using cough embedding as the features to train and evaluate machine learning classifiers can be applied as a low-cost, point-of-care screening test, and deployed as a standalone application or fused to the audio-based TB discrimination [5] for improved TB classification.

II. Data

A. Recording setup

The recordings were collected between 10 am and 4 pm in an outside cross-ventilated sputum collection booth inside a busy primary healthcare clinic near Cape Town, South Africa, representing a real-world setting where an automatic TB test would likely be deployed, as shown in Fig. 1 and Fig. 2 of [5]. All participants were suffering from some sort of lung

978-1-6654-9106–8/22/$31.00

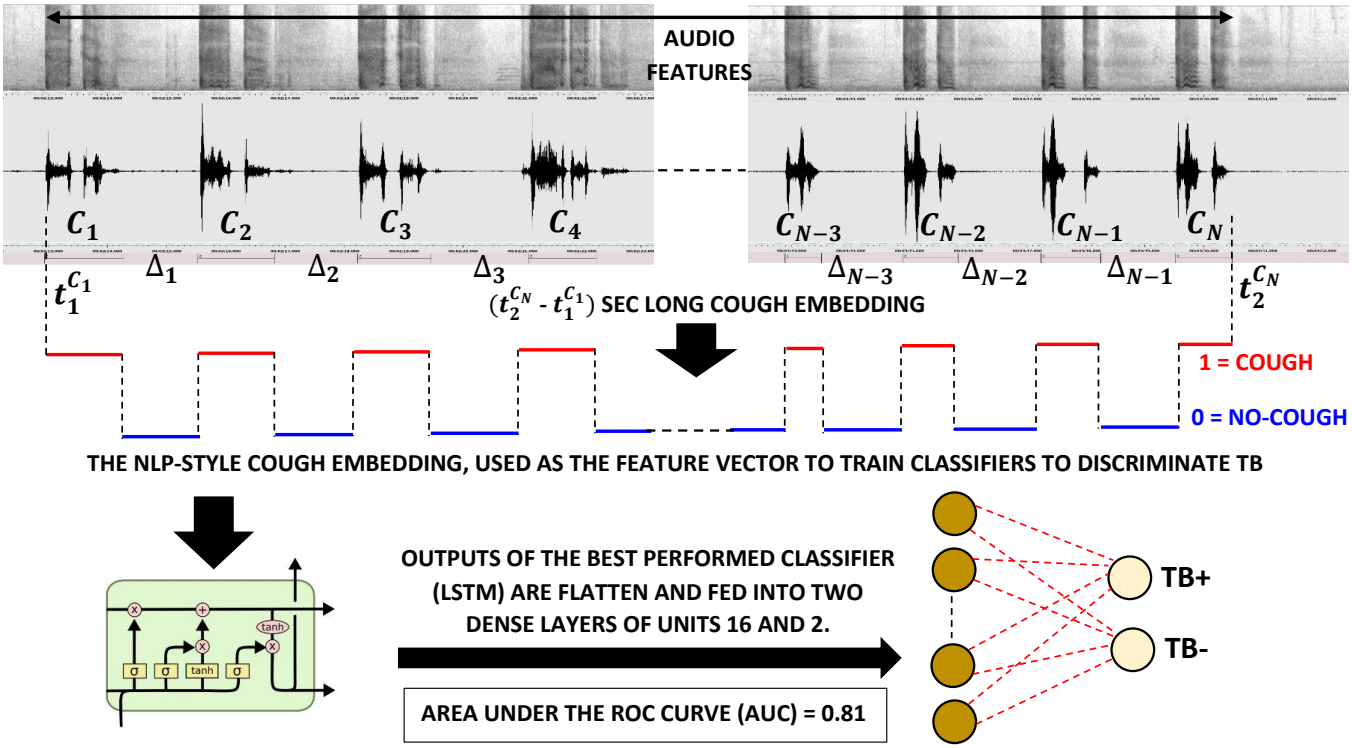

Fig. 1. **Feature extraction process.** The ELAN screenshot of an audio recording from a patient shows that the initial cough embedding before padding is $(t_2^{C_N} - t_1^{C_1})$ sec long. Each annotated cough is noted as C_i and there are total N number of coughs from the patient. Audio features such as MFCC, ZCR and kurtosis are also extracted from the audio segment of the cough embedding, where only cough audio is preserved and everything else is kept as silence (zeros). We experimentally find that an LSTM has produced the highest AUC of 0.81 while using cough embedding as features (Table IV).

TABLE I
DATASET DESCRIPTION. THIS IS THE DATASET WHICH IS USED TO CREATE COUGH EMBEDDING AND FINALLY TO TRAIN AND EVALUATE THE MACHINE LEARNING CLASSIFIERS.

	Number of Patients	Number of Coughs	Average Number of Coughs	Total length of Coughs	Average length of Coughs	Total length of Cough Embedding	Average length of Cough Embedding
TB	15	354	23.6±10.5	4.47 min	17.86±6.81 sec	26.5 min	1.77±0.79 min
Non-TB	33	881	26.7±10.7	11.23 min	20.42±8.75 sec	1.14 hr	2.08±0.7 min
Total	48	1235	25.73±10.71	15.7 min	19.62±8.28 sec	1.59 hr	1.98±0.74 min

anomalies and all of them were TB suspects, as they had cough as their self-reported symptom. They were only diagnosed for TB by standardised methods in this study and diagnosing other diseases apart from TB were impractical to collect. Patients were formally interviewed by the healthcare workers and their inclusion and exclusion criteria are listed in Table 1 of [5]. All of them provided their informed consent and this study was approved by the Faculty of Health Sciences Research Ethics Committee of Stellenbosch University (N14/10/136) and the city of Cape Town (10483).

A mobile recording equipment consisting a RØDE M3 microphone and a ZOOM F8N field recorder was used to record each patient at a sampling rate of 44.1 kHz. A standard N95 mask, which was replaced after each patient, covered the condenser microphone and the gap between the patient and the microphone was maintained between 10 and 15 cm. Each patient was asked to count, cough, take a few deep breaths and cough again without any interruption in the audio recording. This produced at least two bursts of voluntary coughs, as all patients in our study were suffering from a respiratory disease thus producing coughs automatically due to the irritation in their respiratory system [7].

B. Annotation

The multimedia software ELAN [8] was used to manually annotate the coughs with the label ‘c’ in the audio recording, as shown in Fig. 1. The initial listening revealed that a single cough event often contains multiple cough onsets or cough episodes and the episodes which appear on a single breath out of the patients were annotated as a single cough event.

C. Data description

Table I describes the dataset used for feature extraction and classifier training. There were 354 and 881 coughs recorded

from 15 TB and 33 non-TB patients respectively i.e. 1235 coughs in total. The average number of coughs for TB and non-TB patients are 23.6 and 26.7 with a standard deviation (SD) of 10.5 and 10.7 respectively. There are 4.47 minutes of TB coughs and 11.23 minutes of non-TB coughs with an average 17.86±6.81 sec of coughs per TB patient and 20.42±8.75 sec of coughs per non-TB patient. The lengths of cough embeddings are 26.5 minutes and 1.14 hour for the TB and non-TB patients, respectively. The average length of cough embedding from TB patients is 1.77 min with a SD of 0.79 minutes and 2.08 min with SD of 0.7 minutes for non-TB patients. The shortest cough embedding has been 33.8 sec and the longest cough embedding has been 3.93 minutes, which reveals the diversity in our dataset, shown in Fig. 2.

Table I reveals an imbalance between the TB and non-TB distribution. This imbalance was addressed by using AUC as the performance metric, which has a higher degree of discriminancy than some other popular existing performance metrics such as accuracy and applying synthetic minority oversampling technique (SMOTE) [9] to create new data points inside the cross-validation training folds so that deep neural network (DNN) classifiers can perform better [6], [10].

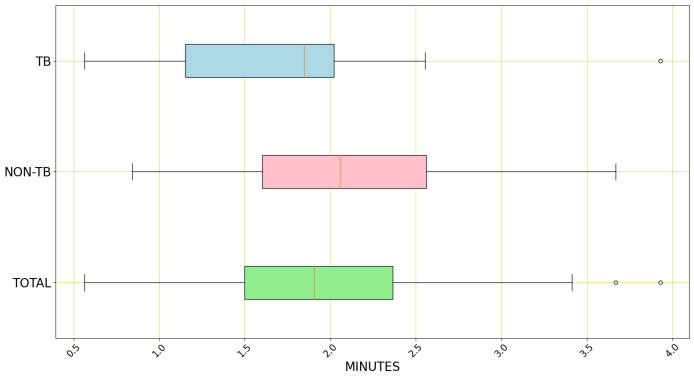

Fig. 2. **The distribution of the lengths of the TB and non-TB cough embeddings.** The medians and the quartiles of the non-TB cough embedding are longer than TB cough embedding.

III. Feature Extraction

We have extracted both NLP-style cough embedding and audio features from the recordings. We note each annotated cough as C_i, where $i = 1, 2, \ldots, N$ and N is the total number of coughs produced by a patient in the recording. We also noted the cough C_i starts at $t_1^{C_i}$ and ends at $t_2^{C_i}$. The time gap between two consecutive coughs, C_i and C_{i+1}, is noted as Δ_j, where $j = 1, 2, \ldots, (N-1)$.

A. NLP-style cough embedding

Word embeddings in NLP are some of the most popular methods in biomedical applications [11] and they are usually the low-dimensional features used to train and evaluate machine learning classifiers [12]. The word embeddings are generated for each word in a sentence and each unique word corresponds to an unique value [13]. Being inspired by the word embedding, we have developed cough embedding, which is a technique to specifically preserve the sequence and timing information of the cough occurrences in a long audio recording. Whereas a large data corpus can contain millions of these unique words [14], our cough embedding contains only two symbolic words: 'cough' (C_i) and 'no-cough' (Δ_j), as explained in Fig. 1.

Cough embedding feature vector is created between the initial cough C_1 and the final cough C_N. For each cough C_i, '1' has been generated and '0' has been generated for each 'no-cough' Δ_j, both with the sampling frequency 1 kHz, and finally concatenated together to produce $(t_2^{C_N} - t_1^{C_1})$ sec or $(t_2^{C_N} - t_1^{C_1}) \times 1000$ sample long initial cough embedding.

In word embedding, it is also a common practice to use zero-padding due to the different lengths in the sentences [15]. We have induced the same principle and all cough embeddings have been padded by adding zeros at the end to make the final length of 3.93 minutes or 235750 samples long. Thus, the final cough embedding in samples after zero-padding is:

$$\mathcal{R}(t) = (t_2^{C_N} - t_1^{C_1}) \times 1000 + \Lambda_c \quad (1)$$

where Λ_c is the length of zeros padded at the end and $\mathcal{R}(t)$ is the feature vector fed into the classifiers.

B. Features extracted from the audio

We have also used the audio of the coughs preserving their time-domain patterns to extract features to compare the performance of cough embedding. We have used the audio for each cough C_i, sampled at 44.1 kHz and the Δ_j is replaced by silence (zeros), sampled at 8 kHz, as our initial experiments revealed that using a sampling rate much lower than the original sampling rate of 44.1 kHz does not decrease the classifier performance, rather shortens the classifier training time. Thus, the audio used for feature extraction is noted as:

$$\mathcal{A}(t) = \sum_{i=1}^{N} (t_2^{C_i} - t_1^{C_i}) \times 44100 + \sum_{j=1}^{N-1} \Delta_j \times 8000 + \Lambda_a \quad (2)$$

where, Λ_a is the zero-padding. We note that the sample length of $\mathcal{A}(t)$ is 3134170 i.e. $\acute{A} = 3134170$.

Features such as mel-frequency cepstral coefficients (MFCCs) along with their velocity and acceleration coefficients, zero-crossing rate (ZCR) and kurtosis [16] were extracted. MFCCs performed better than the linearly-spaced log filerbanks [4] in our previous TB [5] and COVID-19 [6], [10] classification tasks and have also been proved to be very useful in both detection and classification of voice audio such as speech [17] and coughs [18].

The number of lower order MFCCs ($\mathbb{M}$) and the sample length of frames used to extract features ($\mathbb{F}$) are the hyperparameters which were optimised inside a nested cross-validation scheme and are mentioned in Table II. The table shows that $\mathbb{M}$ has two values: 13 and 26. Unlike our previous experiments [6], [10], we did not use any higher MFCC dimension than 39 in this study, as the initial experiments also revealed no performance improvement for using higher

dimensional MFCCs. The frame lengths are varied between 512 and 2048 samples.

Unlike cough embedding, which is a feature vector, the extracted audio features here are a feature matrix with dimension of $(3\mathbb{M}+2, \frac{4\times\acute{A}}{\mathbb{F}})$, as the hop-length was set to $\frac{F}{4}$.

TABLE II
FEATURE EXTRACTION HYPERPARAMETERS. BY VARYING THE MFCCS BETWEEN 13 & 26 AND FRAME LENGTHS BETWEEN 512 & 2048, THE SPECTRAL RESOLUTIONS OF THE AUDIO HAVE BEEN VARIED.

Hyperparameters	Description	Range
MFCCs ($\mathbb{M}$)	lower order MFCCs to keep	$13 \times k$, where $k = 1, 2$
Frame length ($\mathbb{F}$)	Length of frames (in samples) used to extract features	2^k, where $k = 9, 10, 11$

IV. CLASSIFICATION PROCESS

A. Classifier Architectures

We have used logistic regression (LR), k-nearest neighbors (KNN), support vector machine (SVM) and multilayer perceptron (MLP) as shallow classifiers due to their effectiveness in detecting and classifying cough events [19]–[24]. We have experimentally found that shallow classifiers outperform the deep architectures when the data are small [5]. LR is a simple classifier but has been proved to be more effective than more complex classifiers such as SVM, random forests and classification trees in clinical prediction tasks [4], [25].

We have also used convolutional neural network (CNN) [26] and long short-term memory (LSTM) [27] as two deep architectures, as they were successfully applied both in classifying and detecting TB and COVID-19 coughs from both the healthy coughs, sick coughs and from each other [6], [28], [29]. As cough embedding is a feature vector, we applied 1-dimensional CNN (CNN-1D) and as audio features are feature matrix, we applied the traditional 2-dimensional CNN (CNN-2D).

B. Hyperparameter optimisation

As our dataset is small, a stratified k-fold (nested) cross-validation scheme [30], [31] was implemented to make the best use of our dataset along with GridSearchCV to optimise the hyperparameters, mentioned in Table III. The DNN classifiers were created using the TensorFlow pipelines [32] and optimised using Keras GridSearchCV [33]. We have used a five-fold cross-validation, due to its effectiveness in medical applications [34].

C. Classifier evaluation

The area under the receiver operating characteristic (ROC) curve (AUC) score has been the optimisation criteria among the cross-validation folds and the performance-indicator of the classifiers [35]. The average AUC score along with its standard deviation (σ_{AUC}) and the F1-scores [36] over the outer folds during cross-validation are shown in Tables IV. Hyperparameters producing the highest AUC over the inner folds of the cross-validation scheme have been noted as the 'best classifier hyperparameters' in Table IV.

TABLE III
CLASSIFIER HYPERPARAMETERS, OPTIMISED USING A NESTED CROSS-VALIDATION SCHEME.

Hyperparameters	Classifier	Range
Shallow classifiers		
Regularisation (α_1)	LR, SVM	10^{-3} to 10^{3}
$l1$ penalty (α_2)	LR	0 to 1 in steps of 0.05
$l2$ penalty (α_3)	LR	0 to 1 in steps of 0.05
No. of neighbours (α_4)	KNN	10 to 100 in steps of 10
Leaf size (α_5)	KNN	5 to 20 in steps of 5
Kernel Coefficient (α_6)	SVM	10^{-3} to 10^{3}
No. of neurons (α_7)	MLP	10 to 100 in steps of 10
$l2$ penalty (α_8)	MLP	10^{-4} to 10^{4}
DNN classifiers		
Conv-1D filters (β_1)	CNN	3×2^k where $k = 5, 6, 7$
Conv-2D filters (β_2)	CNN	3×2^k where $k = 6, 7, 8$
Kernel size (β_3)	CNN	2 and 3
Dropout rate (β_4)	CNN, LSTM	0.1 to 0.5 in steps of 0.2
Dense layer units (β_5)	CNN, LSTM	2^k where $k = 4, 5$
LSTM units (β_6)	LSTM	2^k where $k = 5, 6, 7$
Learning rate (β_7)	LSTM	10^{-k} where, $k = 2, 3, 4$
Batch Size (β_8)	CNN, LSTM	2^k where $k = 5, 6, 7$

V. RESULTS

All our classifiers are trained and evaluated using both cough embedding and audio features. The results are shown in Table IV and the ROC curves of the best performances are shown in Fig. 3.

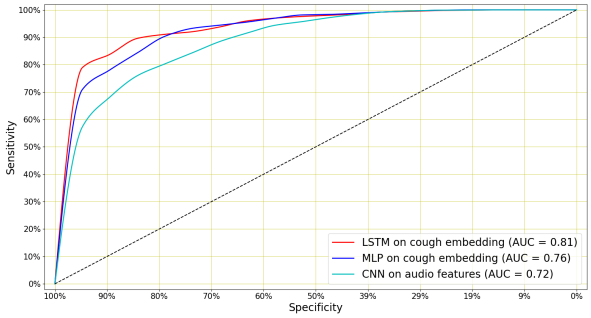

Fig. 3. **The ROC curves for diagnosing TB in cough patterns:** The AUC of 0.81 and 0.76 are achieved from the LSTM and MLP while using cough embedding. The highest AUC of 0.72 is achieved from the CNN while using the audio features such as MFCC, ZCR and kurtosis.

A. TB Classification using cough embedding

LR has achieved the AUC of 0.69 with the σ_{AUC} of 0.04, whereas SVM and KNN have performed better by producing an AUC of 0.73 and 0.74 respectively and an F1-score of 0.72. An MLP has performed the best amongst all the shallow classifiers by producing the highest AUC of 0.76 (Fig. 3).

Both CNN-1D and LSTM classifiers have performed better than the shallow classifiers. CNN has produced an AUC of 0.78 and this is significantly outperformed by the LSTM as it produced the highest AUC of 0.81 with a σ_{AUC} of 0.08 while detecting TB in cough patterns using cough embedding.

TABLE IV

DETECTING TB IN COUGH PATTERNS: THE LSTM HAS PERFORMED THE BEST BY PRODUCING THE MEAN AUC OF 0.81 WITH σ_{AUC} OF 0.08 USING COUGH EMBEDDING. THE CNN HAS OUTPERFORMED THE OTHER CLASSIFIERS BY ACHIEVING THE HIGHEST MEAN AUC OF 0.72 ALONG WITH THE σ_{AUC} OF 0.05 USING THE AUDIO CLASSIFIERS.

Classifier	Best Feature Hyperparameters	Best Classifier Hyperparameters (Optimised inside nested cross-validation)	Performance		
			AUC	σ_{AUC}	F1-score
Cough Embedding					
LR	—	$\alpha_1 = 0.01$, $\alpha_2 = 0.6$, $\alpha_3 = 0.4$	0.69	0.04	0.67
SVM	—	$\alpha_1 = 0.1$, $\alpha_6 = 10$	0.73	0.06	0.72
KNN	—	$\alpha_4 = 20$, $\alpha_5 = 5$	0.74	0.06	0.72
MLP	—	$\alpha_7 = 60$, $\alpha_8 = 0.001$	0.76	0.05	0.75
CNN-1D	—	$\beta_1 = 64$, $\beta_3 = 2$, $\beta_4 = 0.3$, $\beta_5 = 16$, $\beta_8 = 64$	0.78	0.07	0.78
LSTM	—	$\beta_4 = 0.3$, $\beta_5 = 16$, $\beta_6 = 64$, $\beta_7 = 0.01$, $\beta_8 = 64$	***0.81***	***0.08***	***0.79***
Audio Features					
LR	$\mathbb{M} = 13, \mathbb{F} = 1024$	$\alpha_1 = 0.001$, $\alpha_2 = 0.45$, $\alpha_3 = 0.55$	0.61	0.04	0.60
SVM	$\mathbb{M} = 13, \mathbb{F} = 1024$	$\alpha_1 = 10$, $\alpha_6 = 0.01$	0.61	0.05	0.61
KNN	$\mathbb{M} = 13, \mathbb{F} = 1024$	$\alpha_4 = 40$, $\alpha_5 = 15$	0.64	0.05	0.62
MLP	$\mathbb{M} = 13, \mathbb{F} = 1024$	$\alpha_7 = 50$, $\alpha_8 = 0.01$	0.66	0.05	0.63
LSTM	$\mathbb{M} = 13, \mathbb{F} = 512$	$\beta_4 = 0.3$, $\beta_5 = 16$, $\beta_6 = 128$, $\beta_7 = 0.001$, $\beta_8 = 128$	0.71	0.06	0.71
CNN-2D	$\mathbb{M} = 26, \mathbb{F} = 512$	$\beta_2 = 128$, $\beta_3 = 2$, $\beta_4 = 0.3$, $\beta_5 = 32$, $\beta_8 = 64$	*0.72*	*0.05*	*0.71*

B. TB Classification using audio features

Here, the LR and SVM have produced the AUC of 0.61, whereas the KNN and MLP have produced the AUC of 0.64 and 0.66 respectively with the σ_{AUC} of 0.05. Again, DNN classifiers performed better than the shallow classifiers. This time the CNN-2D has produced the highest AUC of 0.72 along with σ_{AUC} of 0.05 and this performance is marginally better than the LSTM, as it produced the AUC of 0.71.

VI. DISCUSSION

Table IV demonstrates that it is possible to discriminate TB in cough patterns using both cough embedding and audio features, although the performance is significantly better while the prior is used. In both cases, DNN classifiers outperformed the shallow classifiers and a high σ_{AUC} can be noticed in all cases. It is also noticeable that the DNN classifiers have performed better using the cough embedding, but it came with the cost of higher σ_{AUC}. This indicates that although the DNN classifiers are capable of producing higher AUC, they are also prone to overfitting.

While using the audio features, both DNN and shallow classifiers produced similar σ_{AUC} and the DNN classifiers have performed better than the shallow classifiers. Although the performance is worse than cough embedding, using audio features produce a more stable performance across the data indicating better generalisation over the folds and the classifiers are more robust. Recurrent neural networks such as an LSTM performs better on feature vectors such as cough embedding, whereas CNN performs better on feature matrix such as audio features, consisting MFCCs, ZCR and kurtosis.

VII. CONCLUSION AND FUTURE WORK

Here in this study, we have used the cough patterns in the audio recordings taken at a outside recording booth to discriminate TB from other lung ailments. Our dataset contained 15 TB and 33 non-TB patients, who were suffering from other respiratory disease. They asked to cough, breathe and then cough again, thus producing at least two bouts of coughs while standing in front of a standing microphone and field recorder representing a real-world environment where a TB test would likely be deployed. NLP-style cough embedding was created in such a way that the time-domain information for each cough occurrence and sequence are recorded. This cough embedding is used as the features to train and evaluate both shallow and deep architectures using a nested cross-validation scheme. The results show that although TB discrimination was possible in all cases, LSTM performs the best by producing the highest AUC of 0.81. This performance is also compared with the features such as MFCC, ZCR and kurtosis extracted from the audio recordings of the cough embedding. This time a CNN produced the highest AUC of 0.72. Although the performance was worse, classifiers demonstrated better robustness and generalisation across the cross-validation folds.

As for the future work, we are constantly increasing the number of patients and their recordings so that the DNN classifiers and more advanced architectures such as a transformer network can be trained for better performance. We are also in the process of deploying the TensorFlow-based models on an Android and iOS platform.

REFERENCES

[1] K. Floyd, P. Glaziou, A. Zumla, and M. Raviglione, "The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era," *The Lancet Respiratory Medicine*, vol. 6, no. 4, pp. 299–314, 2018. [Online]. Available: https://doi.org/10.1016/S2213-2600(18)30057-2

[2] A. Konstantinos, "Testing for tuberculosis," 2010. [Online]. Available: https://doi.org/10.18773/austprescr.2010.005

[3] P. K. Dewan, J. Grinsdale, S. Liska, E. Wong, R. Fallstad, and L. M. Kawamura, "Feasibility, acceptability, and cost of tuberculosis testing by whole-blood interferon-gamma assay," *BMC Infectious Diseases*, vol. 6, no. 1, p. 47, 2006. [Online]. Available: https://doi.org/10.1186/1471-2334-6-47

[4] G. Botha, G. Theron, R. Warren, M. Klopper, K. Dheda, P. Van Helden, and T. Niesler, "Detection of Tuberculosis by Automatic Cough Sound

Analysis," *Physiological Measurement*, vol. 39, no. 4, p. 045005, 2018. [Online]. Available: https://doi.org/10.1088/1361-6579/aab6d0

[5] M. Pahar, M. Klopper, B. Reeve, R. Warren, G. Theron, and T. Niesler, "Automatic cough classification for tuberculosis screening in a real-world environment," *Physiological Measurement*, vol. 42, no. 10, p. 105014, oct 2021. [Online]. Available: https://doi.org/10.1088/1361-6579/ac2fb8

[6] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 cough classification using machine learning and global smartphone recordings," *Computers in Biology and Medicine*, vol. 135, p. 104572, 2021. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2021.104572

[7] B. Simonsson, F. Jacobs, J. Nadel *et al.*, "Role of autonomic nervous system and the cough reflex in the increased responsiveness of airways in patients with obstructive airway disease," *The Journal of clinical investigation*, vol. 46, no. 11, pp. 1812–1818, 1967.

[8] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: a professional framework for multimodality research," in *5th International Conference on Language Resources and Evaluation (LREC 2006)*, 2006.

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. [Online]. Available: https://doi.org/10.1613/jair.953

[10] M. Pahar, I. Miranda, A. Diacon, and T. Niesler, "Deep Neural Network based Cough Detection using Bed-mounted Accelerometer Measurements," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8002–8006. [Online]. Available: https://doi.org/10.1109/ICASSP39728.2021.9414744

[11] Y. Wang, S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen, P. Kingsbury, and H. Liu, "A comparison of word embeddings for the biomedical natural language processing," *Journal of Biomedical Informatics*, vol. 87, pp. 12–20, 2018. [Online]. Available: https://doi.org/10.1016/j.jbi.2018.09.008

[12] Y. Li and T. Yang, "Word Embedding for Understanding Natural Language: A Survey," in *Guide to Big Data Applications*. Springer, 2018, pp. 83–104. [Online]. Available: https://doi.org/10.1007/978-3-319-53817-4_4

[13] S. Lai, K. Liu, S. He, and J. Zhao, "How to Generate a Good Word Embedding," *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 5–14, 2016. [Online]. Available: https://doi.org/10.1109/MIS.2016.45

[14] S. Ghannay, B. Favre, Y. Esteve, and N. Camelin, "Word Embedding Evaluation and Combination," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 300–305.

[15] A. J. Masino, D. Forsyth, and A. G. Fiks, "Detecting Adverse Drug Reactions on Twitter with Convolutional Neural Networks and Word Embedding Features," *Journal of Healthcare Informatics Research*, vol. 2, no. 1, pp. 25–43, 2018. [Online]. Available: https://doi.org/10.1007/s41666-018-0018-9

[16] R. Bachu, S. Kopparthi, B. Adapa, and B. D. Barkana, "Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy," in *Advanced Techniques in Computing Sciences and Software Engineering*. Springer, 2010, pp. 279–282. [Online]. Available: https://doi.org/10.1007/978-90-481-3660-5_47

[17] M. Pahar and L. S. Smith, "Coding and Decoding Speech using a Biologically Inspired Coding System," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 3025–3032. [Online]. Available: https://doi.org/10.1109/SSCI47803.2020.9308328

[18] I. D. Miranda, A. H. Diacon, and T. R. Niesler, "A comparative study of features for acoustic cough detection using deep architectures," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 2601–2605. [Online]. Available: https://doi.org/10.1109/EMBC.2019.8856412

[19] V. Bhateja, A. Taquee, and D. K. Sharma, "Pre-processing and classification of cough sounds in noisy environment using SVM," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*. IEEE, 2019, pp. 822–826.

[20] B. H. Tracey, G. Comina, S. Larson, M. Bravard, J. W. López, and R. H. Gilman, "Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 6017–6020. [Online]. Available: https://doi.org/10.1109/IEMBS.2011.6091487

[21] R. V. Sharan, U. R. Abeyratne, V. R. Swarnkar, and P. Porter, "Cough sound analysis for diagnosing croup in pediatric patients using biologically inspired features," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 4578–4581.

[22] L. Sarangi, M. N. Mohanty, and S. Pattanayak, "Design of MLP based model for analysis of patient suffering from influenza," *Procedia Computer Science*, vol. 92, pp. 396–403, 2016.

[23] J. Amoh and K. Odame, "Deepcough: A deep convolutional neural network in a wearable cough detection system," in *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2015, pp. 1–4.

[24] J.-M. Liu, M. You, Z. Wang, G.-Z. Li, X. Xu, and Z. Qiu, "Cough detection using deep neural networks," in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2014, pp. 560–563.

[25] E. Christodoulou, J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel, and B. Van Calster, "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models," *Journal of Clinical Epidemiology*, vol. 110, pp. 12–22, 2019.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. [Online]. Available: https://doi.org/10.1145/3065386

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[28] M. Pahar, M. Klopper, B. Reeve, R. Warren, G. Theron, A. Diacon, and T. Niesler, "Automatic Tuberculosis and COVID-19 cough classification using deep learning," *arXiv preprint arXiv:2205.05480*, 2022.

[29] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features," *Computers in Biology and Medicine*, vol. 141, p. 105153, 2022. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2021.105153

[30] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[32] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[33] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[34] A. Rácz, D. Bajusz, and K. Héberger, "Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification," *Molecules*, vol. 26, no. 4, p. 1111, 2021. [Online]. Available: https://doi.org/10.3390/molecules26041111

[35] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[36] D. Fourure, M. U. Javaid, N. Posocco, and S. Tihon, "Anomaly Detection: How to Artificially Increase Your F1-Score with a Biased Evaluation Protocol," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2021, pp. 3–18. [Online]. Available: https://doi.org/10.1007/978-3-030-86514-6_1