HHS Public Access

Author manuscript

Physiol Meas. Author manuscript; available in PMC 2022 November 26.

Published in final edited form as:

Physiol Meas. ; 42(10): . doi:10.1088/1361-6579/ac2fb8.

Automatic Cough Classification for Tuberculosis Screening in a Real-World Environment

Madhurananda Pahar[1], **Marisa Klopper**[2], **Byron Reeve**[2], **Rob Warren**[2], **Grant Theron**[2], **Thomas Niesler**[1]

[1]Department of Electrical and Electronic Engineering, Stellenbosch University, South Africa

[2]SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, DSI/NRF Centre of Excellence for Biomedical Tuberculosis Research, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa

Abstract

Objective: The automatic discrimination between the coughing sounds produced by patients with tuberculosis (TB) and those produced by patients with other lung ailments.

Approach: We present experiments based on a dataset of 1358 forced cough recordings obtained in a developing-world clinic from 16 patients with confirmed active pulmonary TB and 35 patients suffering from respiratory conditions suggestive of TB but confirmed to be TB negative. Using nested cross-validation, we have trained and evaluated five machine learning classifiers: logistic regression (LR), support vector machines (SVM), k-nearest neighbour (KNN), multilayer perceptrons (MLP) and convolutional neural networks (CNN).

Main Results: Although classification is possible in all cases, the best performance is achieved using LR. In combination with feature selection by sequential forward selection (SFS), our best LR system achieves an area under the ROC curve (AUC) of 0.94 using 23 features selected from a set of 78 high-resolution mel-frequency cepstral coefficients (MFCCs). This system achieves a sensitivity of 93% at a specificity of 95% and thus exceeds the 90% sensitivity at 70% specificity specification considered by the World Health Organisation (WHO) as a minimal requirement for a community-based TB triage test.

Significance: The automatic classification of cough audio sounds, when applied to symptomatic patients requiring investigation for TB, can meet the WHO triage specifications for the identification of patients who should undergo expensive molecular downstream testing. This makes it a promising and viable means of low cost, easily deployable frontline screening for TB, which can benefit especially developing countries with a heavy TB burden.

Keywords

tuberculosis; TB; machine learning; cough classification; triage test

mpahar@sun.ac.za .

1. Introduction

Tuberculosis (TB) is a bacterial infection primarily of the lungs and globally a leading cause of death (Floyd et al. 2018). Modern diagnostic tests rely on costly laboratory procedures requiring specialized equipment (Dewan et al. 2006, Bwanga et al. 2009, Konstantinos 2010, Global Health Initiative 2017). However, TB is generally most prevalent in low-income settings and is responsible for 95% of deaths due to infectious disease in developing countries (WHO 2020b). Due to a high index of suspicion in such high-incidence settings, these expensive tests are frequently conducted on patients who meet symptom criteria for TB investigation but cough due to other lung ailments. In fact, most people investigated for TB do not suffer from the disease (Chang et al. 2008).

The simplest form of TB triaging relies on self-reported symptoms. Although there is no cost involved, this has low specificity, resulting in over-testing. Furthermore, TB is also associated with stigmatisation, which may result in under-reporting of symptoms, leading to under-testing and consequent inadequate care (Nathavitharana et al. 2019). Thus, there is a need for a low-cost, point-of-care screening test, such as the automatic classification of cough sounds, which would allow a more efficient and widespread application of molecular testing. If such an objective audio-based test, which is “specimen-free”, were accurate enough, it might offer an improvement in the standard of care (Naidoo et al. 2017).

Coughing is a common symptom of respiratory disease and caused by an explosive expulsion of air from the airways (Simonsson et al. 1967). However, the effect of coughing on the respiratory system is known to vary (Higenbottam 2002). For example, lung diseases can cause the airway to be either restricted or obstructed and this can influence the cough acoustics (Chung & Pavord 2008). It has also been postulated that the glottis behaves differently under different pathological conditions and that this makes it possible to distinguish between coughs due to asthma, bronchitis and pertussis (whooping cough) (Korpáš et al. 1996). Therefore, the automatic classification of the acoustic signals associated with coughing in order to detect lung diseases like TB seems to be a reasonable avenue of investigation.

Vocal audio has been used in various disease classification studies, including the recent COVID-19 pandemic (Hassan et al. 2020, Brown et al. 2020). Aspects of speech such as phonation and vowel sounds have been used to detect Parkinson’s disease by applying machine learning (Almeida et al. 2019, Hemmerling & Sztaho 2019). Respiratory disease such as asthma bronchiale (AB) has also been successfully detected by analysing cough frequency (Marsden et al. 2016). Finally, cough sounds have been used in the diagnosis and screening of pulmonary diseases such as AB, chronic obstructive pulmonary disease (COPD) and TB (Infante et al. 2017). The voluntary coughs produced by AB and COPD patients were successfully distinguished from those produced by healthy participants using discriminant analysis with an accuracy of between 85% and 90% in (Knocikova et al. 2008). The detection of coughing associated with asthma was considered in (Al-khassaweneh & Bani Abdelrahman 2013), while pertussis was detected with good accuracy using logistic regression (LR) in (Pramono et al. 2016). The early detection of congestive heart failure (CHF) and COPD, which can increase the fatality rate in an ageing population, using a

random forest classifier was considered in (Windmon et al. 2018). Successful detection of the seal-like barking cough that can occur in children who suffer from croup or laryngotracheobronchitis has been reported in (Sharan et al. 2018). More recently, domain-specific features like mel-frequency cepstral coefficients (MFCCs) and zero crossing rate (ZCR) have shown promise when used to classify coughs (Rudraraju et al. 2020) due to pneumonia (Sotoudeh et al. 2020) and COVID-19 (Pahar, Klopper, Warren & Niesler 2021a, Pahar, Klopper, Warren & Niesler 2021b). The recently introduced generative adversarial neural network architecture has also proven to be successful in respiratory disease classification (Ramesh et al. 2020).

This work is a direct extension of our previous work in which we demonstrated that it is possible to discriminate between the coughs of TB sufferers and healthy controls using logistic regression (Botha et al. 2018). However, studies that involve only cases with a condition and healthy controls have well-known limitations (Rutjes et al. 2005). We now show that it is also possible to distinguish between the coughs of TB patients and the coughs of patients suffering from other lung ailments and for whom TB was excluded as a diagnosis. In contrast to the controlled environment in which our previous recordings were made, the audio data we use in this work were collected at a TB clinic and include substantial environmental noise, thereby directly addressing the practical scenario encountered at primary health facilities in developing countries. Patients presenting to these clinics are typically ill and it is necessary to establish the likelihood of TB disease for further referral. This referral is typically achieved by collecting an infectious and difficult to handle specimen (sputum) which is tested using an expensive (in a developing world context) test that requires laboratory expertise and specialised equipment. As in our previous work, we focus on automatic cough classification, and assume that the detection of the start and end of coughing has been reliably achieved. We acknowledge that, by sidestepping this detection step, difficult practical challenges, such as the processing of cough spasms, have been left for future work.

To perform our experiments, it was necessary to compile a new corpus of coughing sounds, gathered from patients who suffer from symptoms of TB but do not necessarily have TB. This is a priority population for the World Health Organisation (WHO) TB triage test target product profile. Our new corpus is of a similar extent to that used in our previous work, but is compiled in a more realistic environment, representative of the developing-world primary health care environment in which TB screening would likely to be performed. We therefore believe it provides a first direct affirmation that automatic sound analysis is a promising and viable approach to TB screening in high-incidence settings.

The structure of the remainder of this paper is as follows. The next section will describe our new dataset and how it was compiled. Sections 3 and 4 describes the acoustic features we consider and the machine learning techniques we evaluate. Section 5 presents our method of parameter training and hyperparameter optimisation, followed by the experimental results in Section 6. Results are discussed in Section 7 and finally Section 8 concludes the paper.

2. Data Collection

2.1. Collection Setup

Our data were collected in a busy primary health care clinic in Cape Town, South Africa, where mobile recording equipment was deployed in an outside cross-ventilated sputum collection booth, as shown in Figure 1. This setup is representative of the typical real-world clinic environment in a developing country in which low-cost, easily-deployable automatic TB screening is most urgently needed. All recordings were taken between 10 am and 4 pm by two health care workers without a technical background but who were trained to operate the recording equipment. The recording area was not fitted with additional acoustic protection and was exposed to noise from a consistently high number of patients and staff attending the clinic while the surrounding streets were busy with pedestrians, pets and vehicles. Thus, our dataset contains a considerable amount of environmental noise, and is representative of the scenario in which a TB screening test would likely be deployed. No attempts were made to de-noise the data since we were specifically interested in the performance that can be expected in a real-world scenario. Furthermore, experience in the related field of automatic speech recognition has shown that noise reduction performed before model training is often not beneficial and may reduce robustness to varying input conditions (Caballero et al. 2018). 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). Informed consent was granted for all patients in this study.

2.2. Recording Setup and Annotation

A ZOOM F8N field recorder was used to record the audio captured by a RØDE M3 condenser microphone (Hsu et al. 1998, Todorović et al. 2015), covered by a standard N95 mask which was replaced after each patient. Informal listening tests indicated that the mask did not substantially affect the quality of the recorded audio signal. The health care workers ensured that the gap maintained between the patient and the microphone was 10 to 15 centimetres (Figure 2). Each patient was prompted to count from one to ten, cough, take a few deep breaths and then cough again, thus producing at least two bursts of coughs. All patients in our study were suffering from some sort of respiratory disease. So, when they were prompted to cough, they produced a bout of voluntary coughs due to the irritation in their respiratory system (Simonsson et al. 1967). In a real-world diagnostic scenario, the patient would be asked to produce a voluntary cough. Therefore, our methods are appropriate to best approximate the practical application of the TB classifier.

All audio recordings were sampled at 44.1 kHz. The portions of the resulting audio recordings that contain coughing were manually annotated using the ELAN multimedia software (Wittenburg et al. 2006) as shown in Figure 3. We note that these manually annotated stretches of voluntary coughing often contain several cough onsets. The number of these onsets for all 1358 cough events (Table 2) is 3124, indicating an average of 2.3 onsets per cough event. Among the 402 TB and 956 non-TB cough events, there are 973 and 2151 cough onsets respectively, indicating 2.42±0.83 onsets per TB cough event and 2.25±0.91 onsets per non-TB cough event. These means and standard deviations suggest that the number of cough bursts or onsets per cough event is not an influential factor in the TB

cough classification task. In this research, we make no attempt to automatically identify the boundaries of such onset subdivisions, and this remains an aspect of our ongoing work. We will refer to these stretches of audio containing coughing, such as those annotated in Figure 3, as *cough events* in the remainder of this paper.

2.3. Dataset Description

Our dataset currently contains coughs from 16 TB and 35 non-TB patients and most participants are male with an average age of 38 (Figure 4).

All participants were TB suspects that self-reported an involuntary cough suggestive of an underlying lung pathology. Clinical work was limited to bacteriological TB diagnosis for the purpose of the study. The participants were not seen by a medical doctor, but rather audio samples were collected by the health care workers. Differential diagnosis for diseases other than TB was impractical to collect since they would, in general, require several additional tests, and even then the diagnosis is often based on treatment-related symptom resolution. Therefore, patients were only tested for TB by standardised methods for the purpose of the study and no alternative diagnoses were established apart from TB. The inclusion and exclusion criteria for the participants are listed in Table 1. This information was collected during a formal interview conducted by the health care workers.

Table 2 describes the dataset and shows that there is an imbalance between the number of TB and non-TB coughs. We have used AUC as the performance measure as it has a higher degree of discriminancy than some other existing performance measures such as accuracy for imbalanced datasets (Rakotomamonjy 2004, Huang & Ling 2005, Fawcett 2006). The length of all coughs in our dataset is 1045 seconds (17.42 minutes). TB coughs are on average 0.74 seconds long with a standard deviation of 0.31, while non-TB coughs are on average 0.78 seconds long with a standard deviation of 0.39. Therefore, we note that coughs produced by TB patients are of comparable length to the coughs produced by suffers from other lung ailments. This is in contrast to our previous finding that the coughs produced by TB patients are both longer and greater in number than those produced by healthy individuals (Botha et al. 2018).

All recordings were carried out inside the same recording booth (Figures 1 and 2) and there was no link between the time or date of recording and whether the subject was TB positive or TB negative. Hence, we can assume that the SNR is independent of the TB status. This was confirmed by informal listening checks applied to the audio recordings as well as by calculating SNR estimates for the recordings of TB and non-TB coughs, as listed in Table 2. The table shows that the average SNR is 33.27 dB and 33.93 dB for TB and non-TB coughs respectively, and that the difference between these figures (0.67 dB) is much smaller than the standard deviation of the SNR estimates of both classes. We have used the Equation 1 to estimate SNR (Johnson 2006, Fgee et al. 1999).

$$\mathrm{SNR(dB)} = 10\log\frac{P_s}{P_n} \tag{1}$$

where, P_s is the signal power of the cough audio and P_n is the signal power of the background noise i.e. the entire audio recording except the coughs, breaths and spoken digits uttered by the patients.

For illustration, we show two coughs in Figure 5. The first with human chattering between its two onsets (SNR ≈ 22dB) and the second with little background noise (SNR ≈ 45dB).

3. Feature Extraction

The feature extraction process is illustrated in Figure 6. No filtering or pre-processing has been applied to the cough audio. We have considered mel-frequency cepstral coefficients (MFCCs), log-filterbank energies, zero-crossing rate (ZCR) and kurtosis as features.

3.1. MFCCs

Mel-frequency cepstral coefficients (MFCCs) have been used very successfully as features in audio analysis and especially in automatic speech recognition (Han et al. 2006, Pahar & Smith 2020). They have also been found to be useful for differentiating dry coughs from wet coughs (Chatrzarrin et al. 2011). MFCCs were computed using the following standard procedure:

- The audio signal is divided into short frames and the fast Fourier transform (FFT) is computed for each.

$$f_{mel}(f) = 2595 \times (1 + \frac{f}{700}) \tag{2}$$

- Mel-scaled filterbanks are computed (Equation 2) and the log-power spectrum is calculated.
- The discrete cosine transformation (DCT) is applied to the output of the mel-filterbanks and a certain number of the resulting coefficients are retained.
- The long-term mean of each coefficient is calculated and then subtracted.
- Inter-frame derivatives (velocity Δ) and second-order derivatives (acceleration ΔΔ) are computed for each coefficient and appended to the already computed MFCCs (Azmy 2017). Equation 3 shows the computation of the delta coefficient d_t for the frames c_{t-n} to c_{t+n} and the number of samples (N) is 2.

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2} \tag{3}$$

3.2. Log-Filterbank Energies

These features (Garreton & Yoma 2011) consist of the log energies computed after applying F linearly spaced overlapping triangular filters to the frame power spectrum $\mathfrak{S}(t)$ which is

computed using Equation 4, where $X(t)$ is the FFT of the audio frame and N is the number of samples.

$$\mathfrak{S}(t) = \frac{1}{N}\left|X(t)\right|^2 \tag{4}$$

3.3. ZCR

The zero-crossing-rate (ZCR) is the number of times the signal changes sign within a frame, as indicated in Equation 5 (Bachu et al. 2010). ZCR indicates the variability present in the signal.

$$ZCR = \frac{1}{T-1}\sum_{t=1}^{T-1}\lambda(x(t)x(t-1) < 0) \tag{5}$$

In Equation 5, $\lambda = 1$ when the sign of $x(t)$ and $x(t-1)$ differ and $\lambda = 0$ when the sign of $x(t)$ and $x(t-1)$ is the same and T is the frame length in samples.

3.4. Kurtosis

The kurtosis indicates the tailedness of a probability density (DeCarlo 1997) and specially the prevalence of higher amplitudes in an audio signal. Kurtosis has been calculated according to Equation 6, where μ is the mean and σ is the standard deviation.

$$\Lambda_x = \frac{1}{T-1}\sum_{t=1}^{T-1}\frac{x(t)-\mu}{\sigma^4} \tag{6}$$

3.5. Feature extraction hyperparameters

The feature extraction process is influenced by a number of hyperparamamters, listed in Table 3. Each cough event is first divided evenly into between 1 and 4 sections. From each section, non-overlapping consecutive frames are used to extract features which are averaged. Finally, these average feature vectors for each section are concatenated, increasing the dimensionality of the feature vector by a factor equal to the number of sections. The number of sections is a feature extraction hyperparameter, indicated in Table 3. For example, when using 13 MFCCs; $(3 \times 13 + 2) = 41$ features are extracted from each frame making the dimension of the feature vector (41×1). Therefore when four sections are used, the feature vector presented to the classifier has dimensions (164×1).

For log-filterbank energies, the number of filters in the filterbank is another hyperparameter. For MFCCs, the number of coefficients that are computed can also be varied. While 13 MFCCs are generally accepted to reflect the level of discrimination of the human auditory system, we have also considered a larger number of MFCCs in our experiments.

The frame length corresponds to the number of time-domain samples per frame. Since the audio was sampled at 44.1 kHz, by varying the frame lengths from 256 to 4096 samples, features are extracted from frame durations varying between approximately 5 and 100 msec. By varying the number of log-filterbank filters and MFCCs, the spectral resolution of the features was varied. The only form of feature normalisation applied before classifier training was cepstral mean normalisation which was performed on a per-recording basis.

4. Classifier Description

Five classifiers have been considered for discrimination between TB and non-TB coughs.

4.1. Logistic Regression (LR)

Logistic regression (LR) models have in some clinical situations been found to outperform more sophisticated classifiers (Christodoulou et al. 2019). We came to the same conclusion in our previous work into TB cough classification (Botha et al. 2018), where these models comfortably outperformed hidden-Markov model (HMM) and decision tree (DT) classifiers.

The output of an LR model varies between 0 and 1, making it very useful in binary classification. It can also be considered as a single neuron neural network. The output of an LR classifier is given by:

$$P = \frac{1}{1 + e^{-(a + \mathrm{bx})}} \tag{7}$$

where, the scalar a and the vector **b** are the parameters of the model and P is the classifier probability.

We have considered the gradient descent weight regularisation strength ν_1 as well as lasso ($l1$ penalty) and ridge ($l2$ penalty) estimators to be hyperparameters which were optimised during nested k-fold cross-validation (Figure 7). We have used Equation 8 to estimate the $l1$ penalty (ν_2) and Equation 9 to estimate $l2$ penalty (ν_3) while optimising the loss function of the LR model (Yamashita & Yabe 2003, Tsuruoka et al. 2009).

$$\nu_2 = \sum_{i=1}^{n} (\mathbf{y}_i - \sum_j \mathbf{x}_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{8}$$

$$\nu_3 = \sum_{i=1}^{n} (\mathbf{y}_i - \sum_j \mathbf{x}_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|^2 \tag{9}$$

In Equations 8 and 9, **x** and **y** are the independent and dependent variables respectively.

4.2. k-Nearest Neighbour (KNN)

The k-nearest neighbour classifier bases its decision on the class labels of the k nearest neighbours in the training set. This machine learning algorithm has in the past successfully been able to both detect (Monge-Álvarez et al. 2018, Pramono et al. 2019, Vhaduri et

al. 2019) and classify (Wang et al. 2006, Pramono et al. 2016, Pahar, Klopper, Warren & Niesler 2021b) sounds such as coughs and snores. Our KNN classifier uses the Euclidean distance to calculate similarity.

4.3. Support Vector Machines (SVM)

Support vector machines (SVM) classifiers have performed well in both detecting and classifying coughing sounds (Tracey et al. 2011, Bhateja et al. 2019, Sharan et al. 2017). We have used both linear and non-linear SVM classifiers, based on the computation in Equation 10.

$$\phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - J(\mathbf{w}, b, a) \tag{10}$$

where, $\mathbf{w}$ is the weight vector, a and b are the coefficients and $J(\mathbf{w},b,a)$ is the term to minimise during hyperparameter optimisation for the parameters listed in Table 4.

4.4. Multilayer Perceptron (MLP)

The LR model described in the previous section was intended to be our baseline, and we hoped to improve classification performance using a multilayer perceptron (MLP) neural network. Unlike LR, the MLP is capable of learning non-linear relationships by using multiple layers of neurons to separate input and output. The MLP classifier is based on Equation 11, which shows the computation of a single neuron.

$$y = \phi(\sum_{i=1}^{n} w_i x_i + b) = \phi\left(\mathbf{w}^T\mathbf{x} + b\right) \tag{11}$$

Here, $\mathbf{x}$ is the input-vector, $\mathbf{w}$ is the weight-vector, b is the bias and ϕ is the nonlinear activation function. The weights and the bias are optimised during supervised training.

We have optimised the loss function by using $l2$ penalty estimator, shown in Equation 9 and stochastic gradient descent. The $l2$ penalty estimator, stochastic gradient descent learning rate and the number of hidden layers have been considered as the hyperparameters (Table 4) which were optimised using nested k-fold cross-validation (Figure 7).

4.5. Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a popular deep neural network architecture which has proved particularly effective in image classification (Krizhevsky et al. 2017, Lawrence et al. 1997). The core of a CNN can be expressed by Equation 12, where $net(t, f)$ is the output of the convolutional layer (Albawi et al. 2017).

$$\text{net}(t, f) = (\mathbf{x} * \mathbf{w})[t, f] = \sum_m \sum_n \mathbf{x}[m, n]\mathbf{w}[t - m, f - n] \tag{12}$$

In this equation, $*$ is the convolution operation, $\mathbf{w}$ is the filter or kernel matrix and $\mathbf{x}$ is the input image. The rectified linear activation function was used in the hidden layers and the softmax activation function is applied in the final layer of our CNN architecture (Qi et

al. 2017). The CNN hyperparameters that are optimised during nested cross-validation are listed in Table 4.

5. Classifier Training and Hyperparameter Optimisation

5.1. Nested cross-validation

Because our dataset is small, we consistently used nested 5-fold cross-validation in all experiments. As shown in Figure 7, an outer loop divides the dataset into training (80%) and testing (20%) partitions where it is ensured that there is no patient overlap. Within this outer loop, the training portion is again divided into two independent inner loops: one performing 4-fold and the other 2-fold cross-validation. The former is used to optimise the hyperparameter listed in Table 4, while the latter is used to determine the equal error rate which is used as part of the classifier decision. There was no patient overlap also within the inner loops and the gender balance was even.

This cross-validation strategy makes the best use of our small dataset by allowing all patients to be used for training, hyperparameter optimisation, and final testing while ensuring unbiased optimisation and a strict per-patient separation between all training, development and testing portions while all folds contain the same proportion of both classes.

5.2. Hyperparameter Optimisation

Each classifier has different hyperparameters, as shown in Table 4, which were optimised during nested k-fold cross-validation. For the LR classifier, these were ν_1, the gradient descent weight regularisation as well as ν_2 and ν_3, the lasso ($l1$) and ridge ($l2$) penalty estimators. For the KNN classifier, the number of neighbours $\varkappa_1$ and the leaf-size $\varkappa_2$ were considered as hyperparameters, while for the SVM the regularisation strength ζ_1 and the coefficient of the radial basis function kernel ζ_2 were optimised. For the MLP, the number of hidden neurons ξ_1, the $l2$ penalty ξ_2 and the stochastic gradient descent learning rate ξ_3 were considered. Finally, for the CNN, the number of convolutional layers α_1, the dropout rate α_2 and the batch size α_3 were optimised.

5.3. Classifier Evaluation

The area under the receiver operating characteristic (ROC) curve (AUC) has been used as the primary evaluation metric of the classifier's performance due to its higher degree of discriminancy for an imbalanced dataset such as ours and because it is widely used in medical diagnosis since 1970s (Rakotomamonjy 2004, Huang & Ling 2005, Fawcett 2006). It also indicates how well the classifier has performed over a range of decision thresholds.

Receiver operating characteristic (ROC) curves were calculated within the inner loop of nested cross-validation. From these ROC curves, the decision threshold that achieves an equal-error-rate (γ_{EE}) was computed. If the mean per-frame TB probability for a cough is $\hat{P}$ and the number of frames in a cough is K, then the cough is labelled as a TB cough, when $\hat{P} > \gamma_{EE}$, where:

$$\hat{P} = \frac{\sum_{i=1}^{K} P(Y = 1 \mid X, \theta)}{K} \tag{13}$$

Defining the indicator variable C as:

$$C = \begin{cases} 1 & \text{if } \hat{P} \geq \gamma_{EE} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

We define two TB index scores: TBI_1 and TBI_2 by following Equation 15 and 16.

$$\text{TBI}_1 = \frac{\sum_{i=1}^{N_1} C}{N_1} \tag{15}$$

$$\text{TBI}_2 = \frac{\sum_{i=1}^{N_2} P(Y = 1 \mid X)}{N_2} \tag{16}$$

In Equation 15, N_1 is the number of coughs obtained from the patient, while in Equation 16, N_2 indicates the total number of frames of cough audio gathered from the patient. Hence, Equation 15 computes a per-cough average probability while Equation 16 computes a per-frame average probability. Finally, a patient is classified as having TB when either the per-cough average probability is greater than 0.5, i.e. more than half of all coughs were classified as TB, or the per-frame average probability over all coughs is greater than γ.

$$\text{TB} = \begin{cases} 1 & \text{if } \text{TBI}_1 > 0.5 \\ 1 & \text{if } \text{TBI}_2 > \gamma \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Several variations of Equation 17 were considered, for example using only TBI_1 or only TBI_2 or including a threshold also for TBI_1. However, the presented formulation was found to be the most effective.

Accuracies, positive predictive values (PPV) and negative predictive values (NPV) have also been calculated at the outer loop of the cross-validation scheme using Equation 18, 19 and 20 respectively.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP} + \text{TN} + \text{FN}} \tag{18}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP+FP}} \tag{19}$$

$$\text{NPV} = \frac{\text{TN}}{\text{FN+TN}} \tag{20}$$

Here, TP = true positives; TN = true negatives; FP = false positives and FN = false negatives.

6. Results

6.1. Classifier Performance

The performance of classifiers trained and evaluated using the nested cross-validation procedure, described in Section 5.1, is shown in Table 5. The mean and standard deviation of the AUC was calculated over the outer cross-validation folds. The feature-extraction hyperparameters (Table 3) as well as the associated optimal classifier hyperparameters determined within the inner loop of nested cross-validation (Table 4) are also given for each classifier architecture. Classifier hyperparameters producing the highest AUC in the outer folds have been noted as the 'optimal classifier hyperparameters' in Table 5. Even for our small dataset, this procedure was computationally intensive.

6.2. Feature Selection

As an additional experiment, sequential forward selection (SFS) (Devijver & Kittler 1982) was applied to discover the best performing individual features responsible for distinguishing between TB and non-TB coughs. SFS is a greedy selection procedure that, starting from a single feature, sequentially finds the additional feature that contributes the most to classification performance. SFS was applied within the inner cross-validation loop to the best performing system in Table 5, which uses 26 MFCCs with appended velocity (Δ) and acceleration (ΔΔ) coefficients, and therefore 78 features in total. The results of this selection are shown in Figure 8.

We see in Figure 8 that optimal performance is achieved using 23 of the total 78 features and near-optimal performance is achieved using as few as four features. These best-four features are the 3^{rd}, the 11^{th}, the velocity (Δ) of the 14^{th}, and the 12^{th} MFCCs. It is interesting to note that the first acceleration (ΔΔ) feature to be chosen appears only in the 9^{th} position, after the best-four and the 5^{th}, 17^{th}, 7^{th}, and 18^{th} MFCCs. Figure 8 also shows that there are several features that lead to deteriorated performance and should therefore be omitted from the classifier.

Finally, Figure 9 shows the ROC curves for the classifier subjected to SFS, both when using all 78 and when using the best 23 features. We see that SFS affords better classification performance across a wide range of operating conditions.

7. Discussion

The results in Table 5 show that LR outperforms the other four classifiers, achieving an AUC of 0.86 and an accuracy of 84.54%. Figure 8 shows that, by applying SFS to this LR classifier and retaining the top 23 features, the AUC is further improved to 0.94. Among these 23 selected features, velocity and acceleration coefficients appear only once in the top nine features. Hence near-optimal performance can be achieved by using MFCCs

without added velocity or acceleration. Figure 9 shows that this system is able to achieve a sensitivity of 93% at a specificity of 95% which exceeds the minimal requirement for a community-based TB triage test of 90% sensitivity at 70% specificity set by the WHO (WHO 2014).

The LR model was intended to be our baseline, and we had hoped to improve classification performance using more complex models, such as the MLP and CNN which both contain multiple neurons and can model non-linear relationships. However, it seems that our dataset is too small and perhaps too noisy for the greater flexibility of neural networks to be of benefit. We note that other researchers have also reported the superiority of LR in some clinical prediction tasks (Christodoulou et al. 2019). We also note that all our classifiers have been evaluated within the same nested k-fold cross-validation scheme. Hence, even though the more complex neural network architectures such as the CNN might in future benefit from more training data, our comparison between the classifiers is justifiable and the achieved performance is a reflection of what is currently possible.

Table 5 also shows that, for all classifiers, splitting the cough audio signal into multiple sections does not lead to improved performance, which suggests that the acoustic information in all phases of a cough is equally important for the purposes of TB classification. However, increasing the number of MFCCs does provide consistent improvements, as does the use of longer frames. MFCCs are shown to outperform linearly spaced log-filterbank energies as features, which is in contrast to the indications in our previous work (Botha et al. 2018). However, in our previous work, only a classical MFCC configuration (13 coefficients, with appended velocity and acceleration) was considered. Here, we find that better performance can be achieved using a larger number of MFCCs than is necessary to model the discriminatory characteristics of human hearing. This again leads us to conclude that the classifier is to some extent basing its decision on acoustic information not perceivable by a human listener.

When compared with the simpler log-filterbank energies, MFCCs have the additional advantage of providing a simple and effective way to compensate for convolutional channel variability by means of mean normalisation. The dataset we use in this work is more noisy and less controlled than the one we used in our previous work. For example, the microphone position is generally a little different between recordings since it does not remain in the collection booth overnight. Hence the advantages of channel normalisation may weigh more strongly for the dataset we consider here.

Finally, we have seen that the number of features can be reduced in a greedy fashion to optimise performance. The highest AUC is achieved when using 23 of the possible 78 features, and near-optimal performance can be achieved using as few as four features. This is of particular importance with a view to implement audio TB screening on mobile computing devices, such as smartphones, since computational effort is saved by reducing the dimensionality of the feature vector. Implementation on a consumer mobile device would make the algorithm portable, inexpensive and easy to apply, which makes it attractive in under-resourced environments.

8. Conclusion and Future Work

We have shown for the first time that it is possible to automatically distinguish between the forced coughing sounds of tuberculosis (TB) patients and the coughing sounds of patients with other lung ailments. This strengthens our previous work which indicated that such discrimination is possible between the coughs of TB patients and healthy controls. In contrast to diseases such as croup coughs, which can be classified with higher accuracy (Sharan et al. 2017, Sharan et al. 2018), the sounds of coughs by TB sufferers do not appear to possess obviously identifiable characteristics. This view is based on informal listening tests of our data, as well as the personal opinions of several medical practitioners who we consulted during the course of this research. The identification of precisely which aspects of the cough signal are important for TB classification are a subject of our ongoing work.

Our experiments are based on a newly-compiled dataset recorded in a noisy primary healthcare clinic. Hence, we also show that TB cough classification is possible in the type of real-world environment that may be expected at a screening facility in a developing country. Using nested cross-validation, five machine learning classifiers were evaluated. By applying logistic regression (LR) and performing sequential forward selection (SFS) to select the top 23 of 78 high-resolution MFCC features, an area under the ROC curve (AUC) of 0.94 was achieved which shows that MFCCs without velocity or acceleration can produce near-optimal performance. This classifier achieves 93% sensitivity at 95% specificity, which exceeds the 90% sensitivity at 70% specificity specification considered by the World Health Organisation (WHO) as a minimal requirement for community-based triage testing (WHO 2014).

The proposed screening by automatic analysis of coughing sounds is non-intrusive, can be applied without specialist medical expertise or laboratory facilities, produces a result quickly and can be implemented on readily-available and inexpensive consumer hardware, such as a smartphone. It therefore may represent a useful tool in the fight against TB especially in developing countries where the TB burden is high, such as our own setting in Cape Town, South Africa (Blaser et al. 2016, Mulongeni et al. 2019). Recent studies have shown that, in South Africa, there are currently on average between 600 and 700 TB cases per 100,000 people (WHO 2020a, NICD n.d., Kanabus n.d.).

Our study has several limitations and we aim to improve those in our future work. Firstly, although our dataset is unique, it is also rather small compared to some other datasets used for cough detection (Pahar, Miranda, Diacon & Niesler 2021) and classification (Sharma et al. 2020). We believe this is why more advanced classifiers, such as a convolutional neural network (CNN), did not offer any performance advantage in our experiments (Pahar, Klopper, Warren & Niesler 2021b). We are extending the dataset, hoping that this will allow such more advanced classifiers to perform better than the LR baseline. Secondly, we are currently using only the recordings of the cough sounds as a basis of classification. The speech audio which was also recorded as part of our data collection might allow classifier accuracy to be improved and this investigation is currently ongoing. Thirdly, the manually annotated cough events sometimes contain multiple bursts of cough onsets and methods that identify such bursts within a cough event automatically are also a subject of

our ongoing investigation. Fourthly, we have evaluated our classifier using only a single dataset. To better establish the ability of the classifier to correctly process truly unseen data, an additional validation-only dataset is required. We are in the process of planning such a data collection effort, where recordings will be made in different but also noisy primary healthcare environments. Fifthly, participants were recruited into this study based on a self-reported cough. This approach may miss patients who have a cough but do not report it. To address this, further studies without cough as an eligibility criterion are required. Finally, the proposed system is not yet ready for practical implementation. The automatic detection of the cough within the recorded audio must be considered, as well as the practical integration of our classifier on a mobile device, as well as the consideration of additional audio captured by a stethoscope (Pasterkamp et al. 1997), also form part of our ongoing work.

Acknowledgements

This project was partially supported by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Intramural Postdoctoral programme 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. The authors have confirmed that any identifiable participants in this study have given their consent for publication.

10. References

Al-khassaweneh M & Bani Abdelrahman R. (2013). A signal processing approach for the diagnosis of asthma from cough sounds, Journal of Medical Engineering & Technology 37(3): 165–171. [PubMed: 23631519]

Albawi S, Mohammed TA & Al-Zawi S (2017). Understanding of a convolutional neural network, 2017 International Conference on Engineering and Technology (ICET), IEEE, pp. 1–6.

Almeida JS, Rebouças Filho P. P., Carneiro T, Wei W, Damaševičius R, Maskeliūnas R & de Albuquerque VHC (2019). Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques, Pattern Recognition Letters 125: 55–62.

Azmy MM (2017). Feature extraction of heart sounds using velocity and acceleration of MFCCs based on support vector machines, 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), IEEE, pp. 1–4.

Bachu R, Kopparthi S, Adapa B & Barkana BD (2010). Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy, Advanced Techniques in Computing Sciences and Software Engineering, Springer, pp. 279–282.

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

Blaser N, Zahnd C, Hermans S, Salazar-Vizcaya L, Estill J, Morrow C, Egger M, Keiser O & Wood R (2016). Tuberculosis in Cape Town: an age-structured transmission model, Epidemics 14: 54–61. [PubMed: 26972514]

Botha G, Theron G, Warren R, Klopper M, Dheda K, Van Helden P & Niesler T (2018). Detection of tuberculosis by automatic cough sound analysis, Physiological Measurement 39(4): 045005. [PubMed: 29543189]

Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, Xia T, Cicuta P & Mascolo C (2020). Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3474–3484.

Bwanga F, Hoffner S, Haile M & Joloba ML (2009). Direct susceptibility testing for multi drug resistant tuberculosis: a meta-analysis, BMC Infectious Diseases 9(1): 67. [PubMed: 19457256]

Caballero FV, Ives D, Laperle C, Charlton D, Zhuge Q, O'Sullivan M & Savory SJ (2018). Machine learning based linear and nonlinear noise estimation, Journal of Optical Communications and Networking 10(10): D42–D51.

Chang A, Redding G & Everard M (2008). Chronic wet cough: protracted bronchitis, chronic suppurative lung disease and bronchiectasis, Pediatric Pulmonology 43(6): 519–531. [PubMed: 18435475]

Chatrzarrin H, Arcelus A, Goubran R & Knoefel F (2011). Feature extraction for the differentiation of dry and wet cough sounds, IEEE International Symposium on Medical Measurements and Applications, IEEE, pp. 162–166.

Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY & Van Calster B (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models, Journal of Clinical Epidemiology 110: 12–22. [PubMed: 30763612]

Chung KF & Pavord ID (2008). Prevalence, pathogenesis, and causes of chronic cough, The Lancet 371(9621): 1364–1374.

DeCarlo LT (1997). On the meaning and use of kurtosis., Psychological Methods 2(3): 292.

Devijver PA & Kittler J (1982). Pattern Recognition: A statistical approach, Prentice Hall.

Dewan PK, Grinsdale J, Liska S, Wong E, Fallstad R & Kawamura LM (2006). Feasibility, acceptability, and cost of tuberculosis testing by whole-blood interferon-gamma assay, BMC Infectious Diseases 6(1): 47. [PubMed: 16539718]

Fawcett T (2006). An introduction to ROC analysis, Pattern Recognition Letters 27(8): 861–874.

Fgee E-B, Phillips W & Robertson W (1999). Comparing audio compression using wavelets with other audio compression schemes, Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No. 99TH8411), Vol. 2, IEEE, pp. 698–701.

Floyd K, Glaziou P, Zumla A & Raviglione M (2018). 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 6(4): 299–314. [PubMed: 29595511]

Garreton C & Yoma NB (2011). Telephone channel compensation in speaker verification using a polynomial approximation in the log-filter-bank energy domain, IEEE Transactions on Audio, Speech, and Language Processing 20(1): 336–341.

Global Health Initiative (2017). GLI practical guide to TB laboratory strengthening. Last accessed: 15th July, 2021. URL: http://stoptb.org/wg/gli/assets/documents/GLI_practical_guide.pdf

Han W, Chan C-F, Choy C-S & Pun K-P (2006). An efficient MFCC extraction method in speech recognition, IEEE International Symposium on Circuits and Systems, IEEE, pp. 4–pp.

Hassan A, Shahin I & Alsabek MB (2020). COVID-19 Detection System using Recurrent Neural Networks, 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), IEEE, pp. 1–5.

Hemmerling D & Sztaho D (2019). Parkinson's disease classification based on vowel sound, Models and Analysis of Vocal Emissions for Biomedical Applications, p. 29.

Higenbottam T (2002). Chronic cough and the cough reflex in common lung diseases, Pulmonary Pharmacology & Therapeutics 15(3): 241–247. [PubMed: 12099771]

Hsu P-C, Mastrangelo C & Wise K (1998). A high sensitivity polysilicon diaphragm condenser microphone, Proceedings MEMS 98. IEEE. Eleventh Annual International Workshop on Micro Electro Mechanical Systems. An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems (Cat. No. 98CH36176, IEEE, pp. 580–585.

Huang J & Ling CX (2005). Using AUC and accuracy in evaluating learning algorithms, IEEE Transactions on knowledge and Data Engineering 17(3): 299–310.

Infante C, Chamberlain D, Fletcher R, Thorat Y & Kodgule R (2017). Use of cough sounds for diagnosis and screening of pulmonary disease, 2017 IEEE Global Humanitarian Technology Conference (GHTC), IEEE, pp. 1–10.

Johnson DH (2006). Signal-to-noise ratio, Scholarpedia 1(12): 2088.

Kanabus A (n.d.). TB Statistics South Africa - National, Incidence, Provincial. Last accessed: 15th July, 2021. URL: https://tbfacts.org/tb-statistics-south-africa/

Knocikova J, Korpas J, Vrabec M & Javorka M (2008). Wavelet analysis of voluntary cough sound in patients with respiratory diseases, Journal of Physiology and Pharmacology 59(Suppl 6): 331–40. [PubMed: 19218657]

Konstantinos A (2010). Diagnostic tests: Testing for tuberculosis, Australian Prescriber 33(1): 12–18.

Korpáš J, Sadloňová J & Vrabec M (1996). Analysis of the cough sound: an overview, Pulmonary Pharmacology 9(5–6): 261–268. [PubMed: 9232662]

Krizhevsky A, Sutskever I & Hinton GE (2017). Imagenet classification with deep convolutional neural networks, Communications of the ACM 60(6): 84–90.

Lawrence S, Giles CL, Tsoi AC & Back AD (1997). Face recognition: A convolutional neural-network approach, IEEE Transactions on Neural Networks 8(1): 98–113. [PubMed: 18255614]

Marsden PA, Satia I, Ibrahim B, Woodcock A, Yates L, Donnelly I, Jolly L, Thomson NC, Fowler SJ & Smith JA (2016). Objective Cough Frequency, Airway Inflammation, and Disease Control in Asthma, Chest 149(6): 1460–1466. [PubMed: 26973014]

Monge-Álvarez J, Hoyos-Barceló C, Lesso P & Casaseca-de-la Higuera P (2018). Robust Detection of Audio-Cough Events Using Local Hu Moments, IEEE Journal of Biomedical and Health Informatics 23(1): 184–196. [PubMed: 29994432]

Mulongeni P, Hermans S, Caldwell J, Bekker L-G, Wood R & Kaplan R (2019). HIV prevalence and determinants of loss-to-follow-up in adolescents and young adults with tuberculosis in Cape Town, PLoS One 14(2): e0210937. [PubMed: 30721239]

Naidoo P, Theron G, Rangaka MX, Chihota VN, Vaughan L, Brey ZO & Pillay Y (2017). The South African tuberculosis care cascade: estimated losses and methodological challenges, The Journal of Infectious Diseases 216(supplement 7): S702–S713. [PubMed: 29117342]

Nathavitharana RR, Yoon C, Macpherson P, Dowdy DW, Cattamanchi A, Somoskovi A, Broger T, Ottenhoff THM, Arinaminpathy N, Lonnroth K, Reither K, Cobelens F, Gilpin C, Denkinger CM & Schumacher SG (2019). Guidance for Studies Evaluating the Accuracy of Tuberculosis Triage Tests, The Journal of Infectious Diseases 220(Supplement 3): S116–S125. [PubMed: 31593600]

NICD (n.d.). Microbiologically confirmed tuberculosis 2004 −15. Last accessed: 15th July, 2021. URL: https://www.nicd.ac.za/wp-content/uploads/2019/11/National-TB-Surveillace-Report_2004_2015_NICD.pdf

Pahar M, Klopper M, Warren R & Niesler T (2021a). COVID-19 cough classification using machine learning and global smartphone recordings, Computers in Biology and Medicine 135: 104572. [PubMed: 34182331]

Pahar M, Klopper M, Warren R & Niesler T (2021b). COVID-19 Detection in Cough, Breath and Speech using Deep Transfer Learning and Bottleneck Features, arXiv preprint arXiv:2104.02477.

Pahar M, Miranda I, Diacon A & Niesler T (2021). Deep Neural Network based Cough Detection using Bed-mounted Accelerometer Measurements, ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8002–8006.

Pahar M & Smith LS (2020). Coding and Decoding Speech using a Biologically Inspired Coding System, 2020 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, pp. 3025–3032.

Pasterkamp H, Kraman SS & Wodicka GR (1997). Respiratory sounds: advances beyond the stethoscope, American Journal of Respiratory and Critical Care Medicine 156(3): 974–987. [PubMed: 9310022]

Pramono RXA, Imtiaz SA & Rodriguez-Villegas E (2016). A cough-based algorithm for automatic diagnosis of pertussis, PloS one 11(9): e0162128. [PubMed: 27583523]

Pramono RXA, Imtiaz SA & Rodriguez-Villegas E (2019). Automatic cough detection in acoustic signal using spectral features, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 7153–7156.

Qi X, Wang T & Liu J (2017). Comparison of support vector machine and softmax classifiers in computer vision, 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE), IEEE, pp. 151–155.

Rakotomamonjy A (2004). Optimizing area under ROC curve with SVMs., 1st International Workshop on ROC Analysis in Artificial Intelligence (ROCAI), pp. 71–80.

Ramesh V, Vatanparvar K, Nemati E, Nathan V, Rahman MM & Kuang J (2020). CoughGAN: Generating Synthetic Coughs that Improve Respiratory Disease Classification, 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, pp. 5682–5688.

Rudraraju G, Palreddy S, Mamidgi B, Sripada NR, Sai YP, Vodnala NK & Haranath SP (2020). Cough sound analysis and objective correlation with spirometry and clinical diagnosis, Informatics in Medicine Unlocked 19.

Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS & Bossuyt PM (2005). Case–control and two-gate designs in diagnostic accuracy studies, Clinical Chemistry 51(8): 1335–1341. [PubMed: 15961549]

Sharan RV, Abeyratne UR, Swarnkar VR & Porter P (2017). Cough sound analysis for diagnosing croup in pediatric patients using biologically inspired features, 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 4578–4581.

Sharan RV, Abeyratne UR, Swarnkar VR & Porter P (2018). Automatic croup diagnosis using cough sound recognition, IEEE Transactions on Biomedical Engineering 66(2): 485–495. [PubMed: 29993458]

Sharma N, Krishnan P, Kumar R, Ramoji S, Chetupalli SR, Nirmala R, Ghosh PK & Ganapathy S (2020). Coswara-A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis, Proceedings of Interspeech, pp. 4811–4815.

Simonsson B, Jacobs F & Nadel J (1967). 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 46(11): 1812–1818. [PubMed: 6070326]

Sotoudeh H, Tabatabaei M, Tasorian B, Tavakol K, Sotoudeh E & Moini AL (2020). Artificial intelligence empowers radiologists to differentiate pneumonia induced by COVID-19 versus influenza viruses, Acta Informatica Medica 28(3): 190. [PubMed: 33417642]

Todorović D, Matković A, Milićević M, Jovanović D, Gajić R, Salom I & Spasenović M (2015). Multilayer graphene condenser microphone, 2D Materials 2(4): 045013.

Tracey BH, Comina G, Larson S, Bravard M, López JW & Gilman RH (2011). Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 6017–6020.

Tsuruoka Y, Tsujii J & Ananiadou S (2009). Stochastic gradient descent training for l1-regularized log-linear models with cumulative penalty, Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pp. 477–485.

Vhaduri S, Van Kessel T, Ko B, Wood D, Wang S & Brunschwiler T (2019). Nocturnal cough and snore detection in noisy environments using smartphone-microphones, 2019 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, pp. 1–7.

Wang J-C, Wang J-F, He KW & Hsu C-S (2006). Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor, The 2006 IEEE International Joint Conference on Neural Network Proceedings, IEEE, pp. 1731–1735.

WHO (2014). High priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting, 28–29 april 2014, Geneva, Switzerland.

WHO (2020a). Global Tuberculosis Report 2020. Last accessed: 15th July, 2021. URL: https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf

WHO (2020b). Tuberculosis; who is most at risk? Last accessed: 15th July, 2021. URL: https://www.who.int/news-room/fact-sheets/detail/tuberculosis

Windmon A, Minakshi M, Bharti P, Chellappan S, Johansson M, Jenkins BA & Athilingam PR (2018). Tussiswatch: A smart-phone system to identify cough episodes as early symptoms of chronic obstructive pulmonary disease and congestive heart failure, IEEE Journal of Biomedical and Health Informatics 23(4): 1566–1573. [PubMed: 30273159]

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

Yamashita H & Yabe H (2003). An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization, SIAM Journal on Optimization 14(2): 479–499.

Figure 1: Recording booth:

Cough recordings were made in a standard cross-ventilated sputum collection booth situated outside on the premises of a primary health care clinic in a high-density urban neighbourhood of Cape Town, South Africa. The first panel shows the inside of the booth while the second shows a member of the research staff explaining the recording process to a study participant. The recording environment can be considered challenging due to a constant and considerable level of environmental noise. Co-authors Dr. Byron Reeve (left panel) and Dr. Marisa Klopper (right panel) gave their consent to have their visible faces appear in this figure.

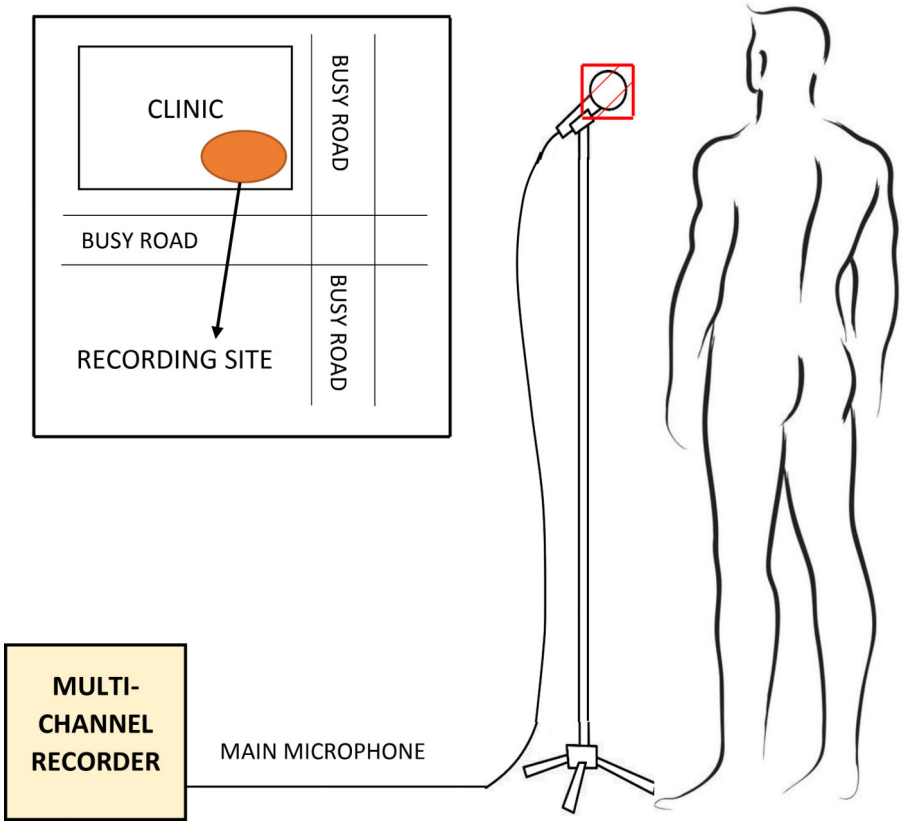


Figure 2: Recording setup:
During recording, the patient stands in front of a microphone covered by a standard N95 mask at a distance of approximately 10 to 15 centimetres. For each patient, the recording session lasted approximately 5 minutes on average. Data collection took place next to a busy road, as shown in the inset.

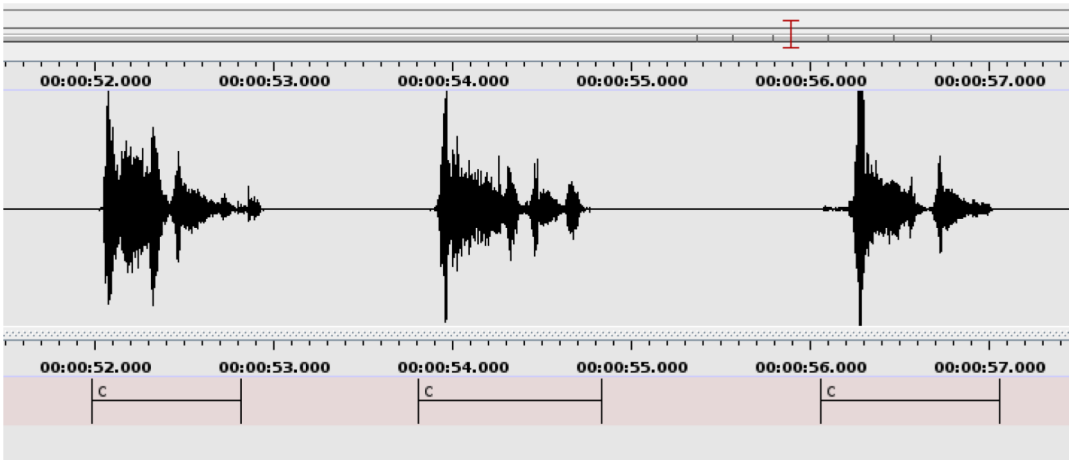


Figure 3: Cough annotation process:

The start and end times of all cough events were manually annotated by the label ‘c’ in the audio recording using the ELAN software. This figure shows the manual annotation of three consecutive cough events. The three labelled cough events include two, three and two individual cough onsets respectively. The manual annotation process indicated the start and the end times of audio containing coughing, but did not label such onsets. However, it revealed that the average number of such onsets is almost equal across the TB and non-TB classes. Each manually annotated cough event was subsequently automatically divided into sections, as described in Section 3.5.

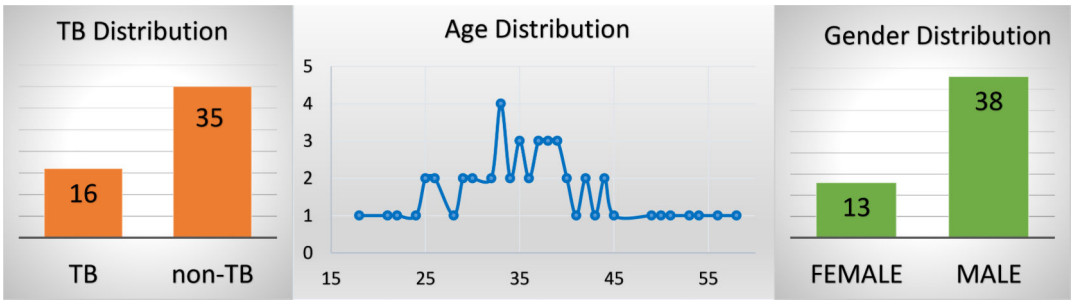
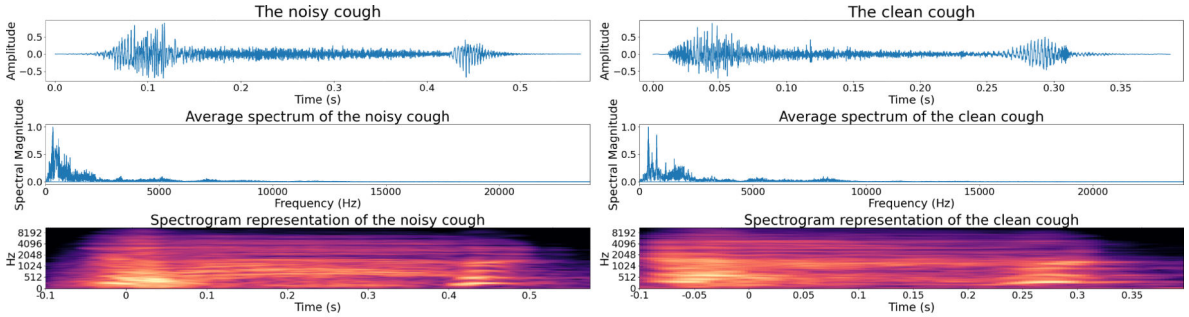


Figure 4:

Demographic distribution of the patients show that most of the participants are middle-aged and male patients dominant over the female patients.



(a) A noisy cough of SNR ≈ 22 dB. (b) A clean cough of SNR ≈ 45 dB.

Figure 5: Examples of a noisy and a clean cough:
The noisy cough contains human chattering between its two phases and the clean cough contains minimal background noise.

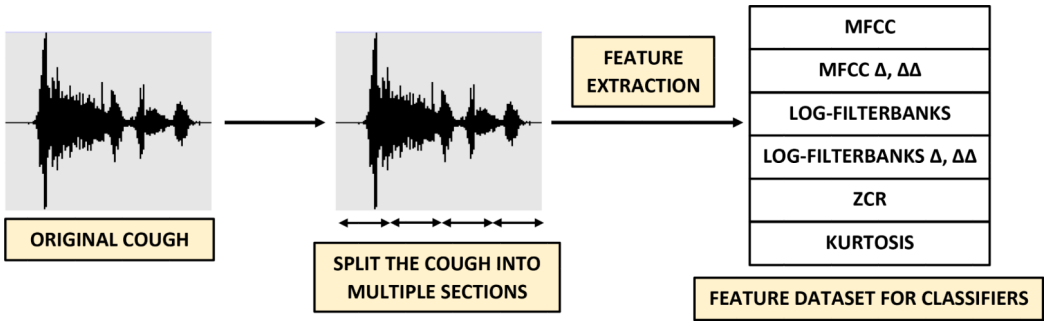


Figure 6: Feature extraction:

Each raw cough recording, as shown in Figure 3, is automatically split into individual sections after which features including MFCCs (including velocity Δ and acceleration ΔΔ), linearly spaced log-filterbank energies (including velocity Δ and acceleration ΔΔ), zero crossing rate and kurtosis are extracted. For example, when using 13 MFCCs, (13 × 3 + 2) = 41 features including Δ, ΔΔ, zero crossing rate and kurtosis are extracted for each section. The number of sections, MFCCs and linearly spaced filters are used as feature extraction hyperparameters listed in Table 3.

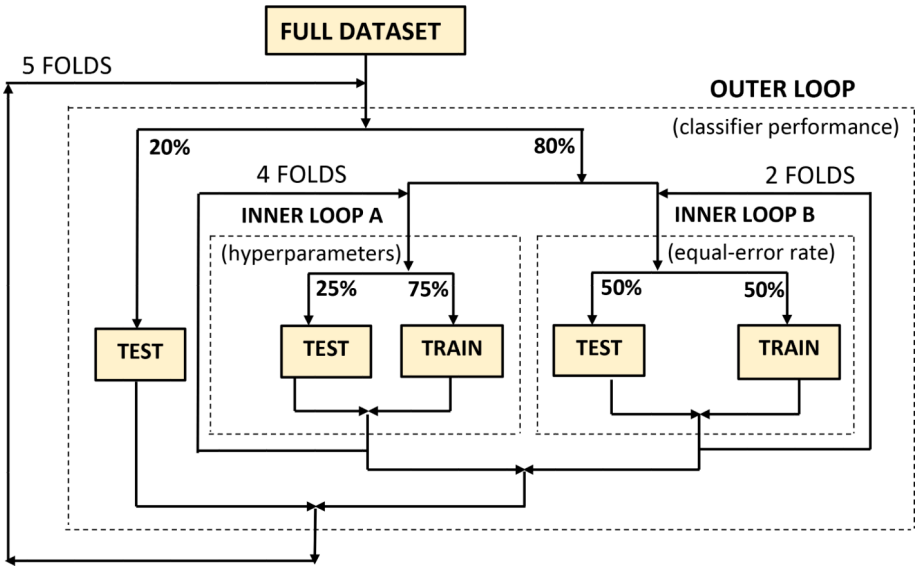


Figure 7:

Nested k-fold cross-validation was used for hyperparameter optimisation as well as the training and evaluation of all classifiers.

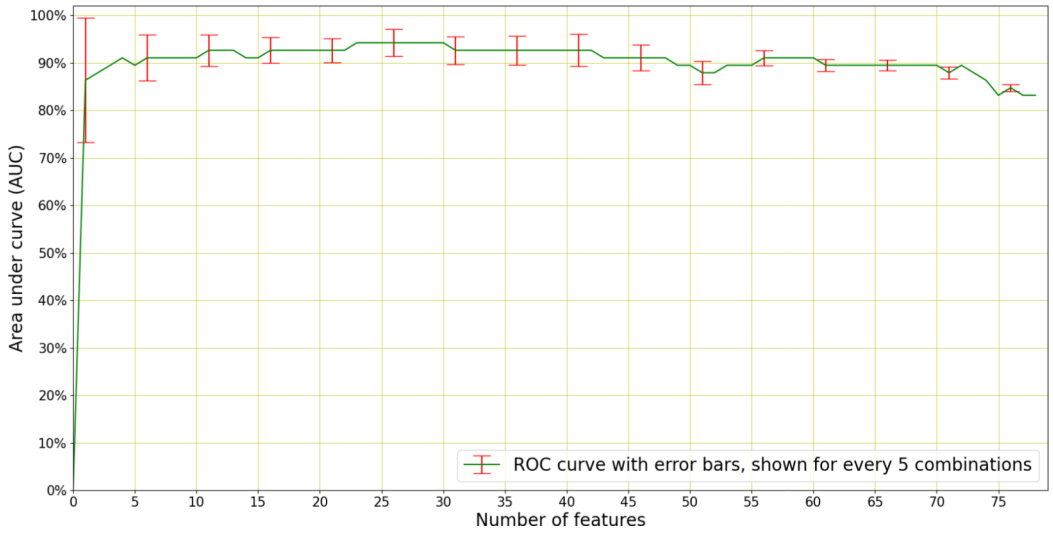


Figure 8: Sequential Forward Selection

(SFS) applied to the best performing system (LR) in Table 5 which uses a total 78 features (26 MFCCs with appended velocity and acceleration). A maximum AUC of 0.94 is achieved using the best 23 features to discriminate TB patients from non-TB patients. The error bars (standard deviation) of the AUC is indicated every 5 features.

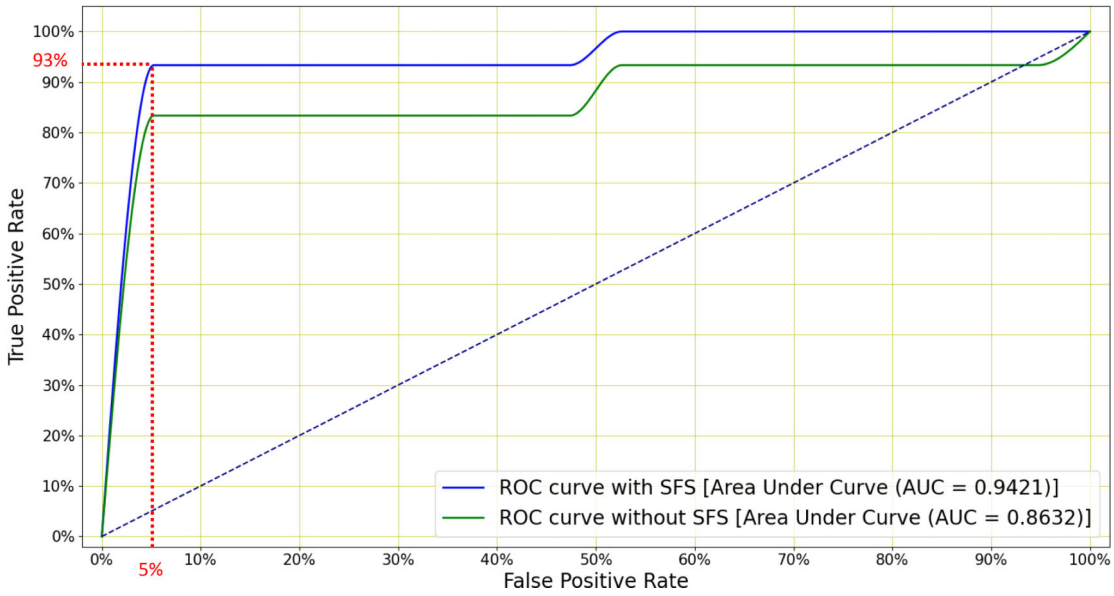


Figure 9:

Mean ROC curves for the LR classifier when distinguishing between TB and non-TB patients with and without SFS (Figure 8). The former uses all 78 features while the latter retains only the 23 best features. A sensitivity of 93% is achieved at 95% specificity and this exceeds the minimum WHO specification for community-based TB triage testing.

Table 1:

Inclusion and exclusion criteria for participants who are included in our dataset summarised in Table 2.

Inclusion	Exclusion
• Age > 18 years, AND	• No consent
• if HIV negative:	• On TB treatment during the 60 days prior to enrolment
– Cough for > 2 weeks, AND	
– Additional symptoms (any of night sweats, fever, weight loss, coughing blood)	• Unable to provide sputum specimens for testing to confirm TB
	• Unable to provide cough audio
• if HIV positive:	
– Cough for any duration, AND	
* Additional symptoms (any of night sweats, fever, weight loss, coughing blood), OR	
* In regular contact with known TB case	

Table 2:

Dataset description: Composition of the dataset used for experimentation. All recorded patients were ill, either with tuberculosis (TB) or with a different lung ailment.

	No. of patients	No. of coughs	Avg coughs per patient	Avg length of coughs	Avg SNR of coughs	Total length of coughs
TB	16	402	25.1	0.74 sec	33.27±15.11 dB	299 sec
Non-TB	35	956	27.3	0.78 sec	33.93±19.24 dB	746 sec
Total	51	1358	26.6	0.77 sec	33.72 dB	1045 sec

Table 3:

Feature extraction hyperparameters for which the results are shown in Table 5.

Hyperparameter	Description	Range
Frame length ($\mathscr{F} =$)	Length of the frames (in samples) from which features were extracted	2^k where k = 8, 9, …, 12
No. of sections ($\mathscr{S} =$)	Number of sections into which frames were grouped	1, 2, 3, 4
No. of linearly spaced filters ($\mathscr{B} =$)	Number of filters used to extract log-filerbank energies	40 to 200 in steps of 20
No. of MFCCs ($\mathscr{M} =$)	Number of lower order MFCCs coefficients retained	13, 26, 39

Table 4:

Classifier hyperparameters which are optimised using nested k-fold cross-validation.

Hyperparameter	Classifier	Range
Regularisation Strength (ν_1)	LR	10^{-7} to 10^{7}
*l*1 penalty (ν_2)	LR	0 to 1 in steps of 0.05
*l*2 penalty (ν_3)	LR	0 to 1 in steps of 0.05
No. of Neighbours (χ_1)	KNN	10 to 100 in steps of 10
Leaf size (χ_2)	KNN	5 to 30 in steps of 5
Regularisation Strength (ζ_1)	SVM	10^{-7} to 10^{7}
RBF kernel coefficient (ζ_2)	SVM	10^{-7} to 10^{7}
No. of hidden neurons (ξ_1)	MLP	10 to 100 in steps of 10
*l*2 penalty (ξ_2)	MLP	10^{-7} to 10^{5}
Stochastic gradient descent learning rate (ξ_3)	MLP	0 to 1 in steps of 0.05
No. of Convolutional Layers (a_1)	CNN	2^k where $k = 4, 5, 6$
Dropout Rate (a_2)	CNN	0.1 to 0.5 in steps 0.2
Batch Size (a_3)	CNN	2^k where $k = 6, 7, 8$

Table 5:

Classifier performance in discriminating TB patients from non-TB patients: The performance achieved by each considered classifier architecture in terms of the area under the ROC curve (AUC). Optimal classifier hyperparameters, determined during cross-validation, are also shown. Mean and standard deviations of AUC are calculated over the five outer folds of the nested k-fold cross-validation, shown in Figure 7.

Classifier	Feature Hyperparameters	Mean AUC	Mean SD	Mean Accuracy	Mean PPV	Mean NPV	Optimal Classifier Hyperparameters
LR	$\mathscr{M} = 13; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.8000	0.0519	78.82%	75.41%	83.29%	ν_1=100, ν_2=0.15, ν_3=0.85
LR	$\mathscr{M} = 13; \mathscr{F} = 2^{10}; \mathscr{S} = 1$	0.7842	0.0645	75.93%	72.59%	80.43%	ν_1=0.001, ν_2=0.25, ν_3=0.75
LR	$\mathscr{M} = 13; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7477	0.0402	72.67%	70.36%	75.56%	ν_1=10, ν_2=0.15, ν_3=0.85
LR	$\mathscr{M} = 26; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	*0.8632*	*0.0601*	*84.54%*	*80.56%*	*89.71%*	*ν_1 =0.01, ν_2 =0.3, ν_3 =0.7*
LR	$\mathscr{M} = 26; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7845	0.0491	76.22%	72.9%	80.67%	ν_1=0.00001, ν_2=0.45, ν_3=0.55
LR	$\mathscr{M} = 39; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7458	0.0531	72.11%	70.14%	74.51%	ν_1=0.0001, ν_2=0.7, ν_3=0.3
LR	$\mathscr{M} = 39; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7402	0.0455	73.02%	70.57%	76.14%	ν_1=0.0001, ν_2=0.4, ν_3=0.6
LR	$\mathscr{B} = 60; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7526	0.0507	72.89%	70.26%	76.31%	ν_1=0.1, ν_2=0.45, ν_3=0.55
KNN	$\mathscr{M} = 26; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7701	0.0505	75.09%	71.89%	79.38%	χ_1=80 χ_2=20
KNN	$\mathscr{M} = 26; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7394	0.0385	70.76%	68.95%	72.96%	χ_1=60 χ_2=15
SVM	$\mathscr{M} = 26; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7435	0.0543	71.91%	69.69%	74.7%	ζ_1=0.01 ζ_2=100
SVM	$\mathscr{M} = 39; \mathscr{F} = 2^{10}; \mathscr{S} = 1$	0.7291	0.0495	70.05%	67.95%	72.71%	ζ_1=0.001 ζ_2=0.0001
MLP	$\mathscr{M} = 13; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7389	0.0457	71.16%	68.91%	74.02%	ξ_1=80, ξ_2=0.0001, ξ_3=0.65
MLP	$\mathscr{M} = 26; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.8000	0.0391	77.87%	75.13%	81.28%	ξ_1=50, ξ_2=0.001, ξ_3=0.55
MLP	$\mathscr{M} = 39; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7742	0.0409	76.47%	73.63%	80.09%	ξ_1=30, ξ_2=0.01, ξ_3=0.35
CNN	$\mathscr{M} = 26; \mathscr{F} = 2^{11}; \mathscr{S} = 1$	0.7109	0.0409	68.89%	67.7%	70.25%	a_1=32 a_2=0.3, a_3=128
CNN	$\mathscr{M} = 39; \mathscr{F} = 2^{10}$ $\mathscr{S} = 1$	0.7001	0.0301	68.71%	67.52%	70.07%	a_1=64 a_2=0.1, a_3=128