I suggest limit the Literature Review to other research on the classification of disease on the basis of cough audio, and possibly, for TB, also other machine learning approaches that have been tried, eg. x-rays.

Chapter 3

(So some of the material currently here can be moved to chapter 4 "Background")

Literature Review

This chapter aims to provide an overview of the existing literature concerning disease classification based on cough audio in general and TB cough classification in particular. In addition, attention will be given to ~~and~~ contrastive learning. A brief review is provided on tuberculosis, followed by a summary of the existing research in the field of cough classification. A summary of the mainstream contrastive learning architectures and loss functions in the literature is provided. Lastly, a few other resources are discussed that are related to the field of cough classification using contrastive learning.

this should go in background ...

3.1. Tuberculosis

Tuberculosis (TB) is an infectious disease that is caused by a bacteria known as Mycobacterium tuberculosis. ~~There are~~ different types of TB ~~based on the~~ affect different parts of the body ~~part that is affected~~, with pulmonary (lung) TB ~~being~~ the most common ~~type of TB~~. Pulmonary TB is spread when an infectious person coughs, sneezes or spits [1]. According to the 2023 Global Tuberculosis Report [17], 7.5 million people were newly diagnosed with TB in 2022, which is the highest since monitoring by the World Health Organisation ~~has~~ started ~~to monitor TB~~. Globally, an estimated 10.6 million people contracted TB in 2022, of which An estimated 1.3 million ~~deaths were recorded~~ died? ~~in 2022~~, ~~with~~ making Tuberculosis ~~being~~ the 10th leading cause of death world wide [18].

number

reduce repetition of "TB" (your text was correct, just lots of "TB" going on)

? These two sentences seem to contradict each other?

To address ~~the~~ global TB ~~health issue~~ infectious, the World Health Organisation ~~proposed~~ has stated the need to develop a low-cost and accessible TB screening solution, known as ~~the~~ a triage test. The triage test has to be simple and low-cost, require minimal samples and maintenance. The aim is to reduce the use of expensive tests and make tests more accessible to patients who previously could not access TB tests. The triage test has to who states that such a have an overall minimum sensitivity of 90% (with an optimal sensitivity of ~~more than~~ 95%) and overall minimum specificity of 70% (with an optimal specificity of ~~more than~~ 80%) [19].

or more

The aim is to reduce the number of people referred to the expensive tests (unnecessarily)

The purpose of this project is to contribute to the development of such a triage test based on cough audio.

note a triage test is a general medical term

trī΄age noun
1 Sorting out
2 In war, etc, the selection for treatment of those casualties most likely to survive
3 Allocation of resources to where they will have the most effect, rather than to where the need is most urgent or severe
4 Broken coffee beans

!!

Also accessible in the sense that test can be performed at locations where expensive tests not available

I would suggest review other cough classification work first (including eg pneumonia and COVID) and then describe what has been done for TB by our group and others,

for TB detection for a number of years

3.2. Cough audio classification

The Digital Signal Processing Group (DSP) Group at Stellenbosch University has been engaged in the development of cough sound classifiers [4], [6], [5], [20], [21], [22]. The aim of this ongoing research project is to build a mobile application can can allow first-contact personnel at primary health-care clinics to administer a cost free/low cost test to patients as a TB screening process. The test uses a cough audio sample as input to a machine learning model that then classifies the pateint as TB positive or TB negative. This project ~~is a part of this group and~~ contributes to this effort by aims ing to improve the performance of the existing ~~model~~ system to ensure that the model's specificity and sensitivity are within the guidelines as set out by ~~triage test~~ the WHO. In the following sections, the chronological results of the work done by DSP group are reviewed, as well as other related TB classification papers.

distinguish between the coughs of patients with

The first paper aimed to identify whether one could use a machine learning model to ~~accurately predict whether a patient has~~ TB ~~or not~~ [4]. ~~A logistic regression model was used to classify whether patients have TB or not~~. It was found that classification of short-term spectral features is possible using logistic regression with an accuracy of 78%. The work also found that the classifier is using some spectral distinction beyond the limit of human auditory perception. This shows that this field of research has promise.

and healthy coughs

repeated in next section

that consisted of coughs recorded from patients who were all sick

The next study expanded on the work done by including new data ~~to investigate whether the machine learning model can detect TB among sick patients~~ (but not necessarily with TB) [6]. Logistic regression (LR), k-nearest neighbors (kNN), support vector machines (SVM), multi-layer perceptions (MLP), and a convolutional neural network (CNN) were considered as possible classifiers. The study found that logistic regression again performed the best among the five classifiers considered. ~~Nested cross-validation was used since the data set is still small~~. This paper shows that it is possible to distinguish TB ~~from~~ and those with other pulmonary illnesses.

no need for all the detail

between the coughs by patients with

Since the COVID-19 pandemic had begun, members of the group subsequently considered whether

~~The study described~~ in [5] ~~investigated whether~~ voice recordings on a smartphone can be used to diagnose COVID-19 patients remotely. Seven machine learning algorithms were implemented namely logistic regression (LR), k-nearest neighbor (KNN), support vector machine (SVM), multi-layer perceptron (MLP), convolutional neural network (CNN), long short-term memory (LSTM) and a residual-based neural network architecture (ResNet). ~~The study~~ In this case it was found that the ResNet had the best performance, followed by the CNN and LSTM classifiers. Similar to [6], this paper found that the classifiers use some information imperceptible to ~~the~~ human ~~ear~~ hearing to classify the input data. This paper shows that the Resnet50 classifier can accurately predict whether patients have COVID-19 using only a smartphone recording. of cough?

?

classifies

Next, The research team looked at distinguishing between TB and COVID-19 using deep ~~learning techniques~~ [20]. The existing dataset was extended to include coughs from 47 TB patients, 229 COVID-19 patients, and 1498 healthy patients. The deep learning models evaluated included a CNN, LSTM, and ResNet50. SMOTE was used to generate data points to balance the dataset. The initial results indicated that the CNN delivered the best results. Based off previous research where transfer learning has been shown to improve performance, the architectures were then pre-trained using sneeze, speech, and noise data. After the addition of transfer learning, the LSTM model delivered the best results, exceeding the specificity and sensitivity required by the WHO. This research shows that transfer learning can be used to improve the performance of existing models that are used to perform TB and COVID-19 classification. for TB detection

was further

The use of recurrent neural network~~s~~ architectures ~~were~~ explored following the results from the research discussed above [21]. Three classifiers were compared ~~to perform TB classification~~ namely logistic regression (used a the baseline classifier), a basic BiLSTM ad a BiLSTM with attention. Furthermore, Three types of acoustic features were considered namely mel-spectrograms, linear filter-bank energies and MFCCs. The results indicate that a BiLSTM model delivers the best performance results ~~compared to the previous results~~. It was also shown that feature selection improves the generalisation across the different cross-validation folds.

concerns

An audio-based monitoring system as used above can lead to ~~some~~ privacy ~~issues~~. As a result the team investigated using an accelerometer to monitor coughing [22]. The accelerometer is on-board an inexpensive smartphone with an external microphone that is attached to a patient's bed. The aim is identify cough from other activites such as sneezing, throat clearing and movement in the bed. Six different classifiers were consider. Three baseline classifiers namely logistic regression (LR), support vector machine (SVM) and a multi-layer perceptron (MLP). Three deep learning architecures were also considered, namely a concolutional neural network (CNN), long short-term memory (LSTM) and a ResNet50. All the models were trained and evaluated using leave-one-patient-out cross-validation. For both the accelerometer-based cough detector and the audio-based cough classifier the multi-layer perceptron model performed the best of the baseline classifiers with an accelerometer-based accuracy of 85.67% and an audio-based accuracy of 89.78%. The ResNet50 also performed the best for both cases of the deep classifiers with an accelerometer-based accuracy of 96.71% and an audio-based accuracy of 98.13%. The results indicate that audio-based cough detection is more accurate than the accelerometer based cough detection, but that the performance is very similar when using deeper architectures, such as a ResNet50.

The work in this paragraph does cough identification, but does not classify disease. I think it can be omitted.

Ask Grant for some background info - he was the SA PI.

But first read info of Paul Ellis's Slurpsie

The CODA TB DREAM Challenge set out to develop a tool to diagnose TB that is low-cost and non-invasive [23]. The data for this challenge consists of TB cough recordings that were collected at clinics using the Hyfe application. The data was collected in seven countries namely India, the Philippines, South Africa, Uganda, Vietnam, Tanzania, and Madagascar. A total of 9772 coughs were recorded, of which 2930 coughs were from TB-positive patients, and the remaining 6842 coughs were from TB-negative patients with persistent coughs. The dataset also included other meta-data that could be useful in developing a model including age, sex, reported duration of cough, hemoptysis, heart rate, fever, weight loss, etc. The meta-data was appended to the flattened log-mel spectrogram and MFCCs for training and testing purposes. The results showed that the models performed the best when trained per participant as opposed to training per cough. The paper shows that the data collected using the Hyfe application can be used to train and test models.

In 2022, researchers in India published an article that considers the development of an AI platform for pulmonary tuberculosis (PTB) screening using cough samples [24]. The data for this research, which was collected at Andhra Medical College in India, consists of 278 positive and 289 negative PTB patients. The algorithm used is based on a multi-modal convolution al neural network and a tabular model based on the collected features. The model combines the logic of the CNN and the tabular model - consisting of primary and secondary features - to perform classification. The model achieved 86.82% accuracy, 90.36% sensitivity, and 84.67% specificity, meeting the minimum requirements for the triage diagnosis as set out by the World Health Organisation. A

Another research group from India at the University of Mumbai ~~in India~~ considered tuberculosis detection using X-rays [25]. The group used a convolution al neural network (CNN) to perform image classification. The model consists of four phases namely pre-processing, image segmentation, feature extraction, and classification. The dataset consists of X-ray images collected ~~from~~ at a medical clinic in Shenzhen, China, where 80 X-rays are from patients without TB and 58 X-rays are from patients who showed indications of TB. The CNN model consists of 3 hidden layers, 9 convolutional layers, 10 ReLU layers, 4 drop-out layers, and 3 max pooling layers. The group used an Adam optimiser to achieve an accuracy of 82.09%, a precision of 95%, and a recall of 85%. A

A major problem that is often encountered when developing classification models, especially in the biomedical domain, is the lack of quality data samples. As a result, Generative Adversarial Networks (GAN) are proposed to generate images to train deep learning models [26]. The work suggests using GAN to generate novel synthetic chest X-rays of TB patients. An image translation architecture is used to convert TB-negative chest X-rays into TB-positive X-rays, and visa versa. The generated images are then added to the original dataset. This process helps to mitigate small and unbalanced datasets. The model uses a deep learning architecture consisting of 5 convolutional layers, with a global average pooling layer, a fully connected layer, and a softmax layer with two outputs. Adversarial loss and PatchNCE loss is combined to train the generative model. The results show that a data generation technique improves the effectiveness and robustness of the model.

3.3. Contrastive learning

Suggest put this in "Background" chapter.

Contrastive Learning has recently seen an increase in interest due to its self-supervised learning approach. As a result, many different architectures and approaches have been proposed to implement contrastive learning. A framework and review was ~~created~~ by Le-Khac, et al. [27] to bring together the different approaches.

Contrastive learning is usually (?) implemented using a Siamese network architecture in conjunction with ~~and~~ a contrastive loss function. A Siamese network takes two inputs and compares the similarity between the two inputs using a distance function. A contrastive loss function is used to minimise the distance between similar inputs and maximise the distance between dissimilar inputs.

The different contrastive loss functions [27], [12], [16], [28] that can be used to maximise agreement are discussed in Section 2.4. The mainstream contrastive learning architectures [8], [9], [29], [30], [10], [11], [12], [13], [14], [15] that use Siamese networks to perform contrastive learning are discussed in Section 3.3.1.

One key aspect of contrastive learning is generating a sample to contrast the true sample to. To generate this sample several different augmentation processes can be considered. A transformation refers to the augmented sample that is generated using these processes. Different image and audio augmentation techniques have been proposed in the literature [8], [11], [31], [9], [15], [32], [28].