LITHUANIAN UNIVERSITY OF HEALTH SCIENCES

Haroldas Razvadauskas

EXPLORATION OF SUPERVISED MACHINE LEARNING INTEGRATION AS A PULMONARY AUSCULTATION DIAGNOSTIC TOOL UNDER DIFFERENT LEVELS OF GAUSSIAN WHITE NOISE

Doctoral Dissertation
Medical and Health Sciences,
Medicine (M 001)

Kaunas, 2025

Dissertation has been prepared at the Department of Internal Diseases of Medical Academy of Lithuanian University of Health Sciences during the period of 2020–2024 year.

Scientific Supervisor

Prof. Dr. Saulius Sadauskas (Lithuanian University of Health Sciences, Medical and Health Sciences, Medicine – M 001).

Dissertation is defended at the Medical Research Council of the Lithuanian University of Health Sciences:

Chairperson

Prof. Dr. Andrius Macas (Lithuanian University of Health Sciences, Medical and Health Sciences, Medicine – M 001).

Members:

Prof. Habil. Dr. Jolanta Vaškelytė (Lithuanian University of Health Sciences, Medical and Health Sciences, Medicine – M 001);
Assoc. Prof. Dr. Mantas Malinauskas (Lithuanian University of Health Sciences, Natural Sciences, Biology – N 010);
Prof. Dr. Vaidotas Marozas (Kaunas University of Technology, Technological Sciences, Electrical and Electronic Engineering – T 001);
Prof. Dr. Jūlija Voicehovska (Rīga Stradiņš University, Medical and Health Sciences, Medicine – M 001).

Dissertation will be defended at the open session of the Medical Research Council of the Lithuanian University of Health Sciences on 18th of June, 2025, at 11 a.m. in the Didžioji auditorium of Kaunas Hospital of Lithuanian University of Health Sciences.
Address: Josvainių 2, LT-47144 Kaunas, Lithuania.

LIETUVOS SVEIKATOS MOKSLŲ UNIVERSITETAS

Haroldas Razvadauskas

ANOTUOTO MAŠININIO MOKYMOSI PRITAIKOMUMAS PLAUČIŲ AUSKULTACINĖJE DIAGNOSTIKOJE, ESANT SKIRTINGIEMS GAUSO TRIUKŠMO LYGIAMS

Daktaro disertacija
Medicinos ir sveikatos mokslai,
medicina (M 001)

Kaunas, 2025

Disertacija rengta 2020–2024 metais Lietuvos sveikatos mokslų universiteto Medicinos akademijos Vidaus ligų klinikoje.

Mokslinis vadovas

prof. dr. Saulius Sadauskas (Lietuvos sveikatos mokslų universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Disertacija ginama Lietuvos sveikatos mokslų universiteto Medicinos mokslo krypties taryboje:

Pirmininkas

prof. dr. Andrius Macas (Lietuvos sveikatos mokslų universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Nariai:

prof. habil. dr. Jolanta Vaškelytė (Lietuvos sveikatos mokslų universitetas, medicinos ir sveikatos mokslai, medicina – M 001);
doc. dr. Mantas Malinauskas (Lietuvos sveikatos mokslų universitetas, gamtos mokslai, biologija – N 010);
prof. dr. Vaidotas Marozas (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – T 001);
prof. dr. Jūlija Voicehovska (Rygos Stradinš universitetas, Latvija, medicinos ir sveikatos mokslai, medicina – M 001).

Disertacija bus ginama viešame Medicinos mokslo krypties tarybos posėdyje 2025 m. birželio 18 dieną 11 val. Lietuvos sveikatos mokslų universiteto Kauno ligoninės Didžiojoje auditorijoje.
Disertacijos gynimo vietos adresas: Josvainių g. 2, LT-47144 Kaunas, Lietuva.

CONTENTS

ABBREVIATIONS

AdaBoost	–	Adaptive Boosting
AI	–	artificial intelligence
ANN	–	artificial neural network
AUC	–	area under the curve
CAS	–	continuous auscultated sound
CatBoost	–	Categorical data Gradient Boosting
CKD	–	chronic kidney disease
CNN	–	convolutional neural network
COPD	–	chronic obstructive pulmonary disease
CSV	–	comma-separated values
DAS	–	discontinuous auscultated sound
dB	–	decibel
el. stethoscope	–	electronic stethoscope
ER	–	emergency reception
ET	–	Extra Trees
GB	–	Gradient Boosting
GWN	–	Gaussian white noise
FN	–	false negative
FP	–	false positive
FPR	–	false positive rate
HF	–	heart failure
Histgradient	–	Histogram-based Gradient Boosting Classification Tree
Hz	–	hertz
ICF	–	informed consent form
K-NN	–	K-Nearest Neighbors
LightGBM	–	Light Gradient Boosting Machine
LR	–	Logistic Regression
MCC	–	Matthews correlation coefficient
MFS	–	Medical Faculty student
ML	–	machine learning
MLP	–	Multilayer Perceptron
ms	–	milliseconds
NAS	–	normal auscultated sounds
OI	–	organic intelligence
PIF	–	personal information form
PR-AUC	–	precision-recall area under the curve
PR curve	–	precision-recall curve
RF	–	Random Forest
ROC-AUC	–	receiver operating characteristic area under the curve
ROC curve	–	receiver operating characteristic curve
s	–	seconds
SD	–	standard deviation
SNR	–	signal-to-noise ratio
SVM	–	Support Vector Machiness
TN	–	true negative
TP	–	true positive

TPR	–	true positive rate
WHO	–	World Health Organisation
WN	–	white noise
XGBoost	–	Extreme Gradient Boosting classifier

INTRODUCTION

Lung auscultation is the most important of the four cornerstones of pulmonary system examination. The stethoscope has become ubiquitous in healthcare settings for over 200 years, yet some shortcomings of subjectivity and noise plague it [1–3].

Additionally, the levels of cardiopulmonary auscultation have decreased in recent years. Whilst third leading cause of death across the world remains pulmonary diseases [4–7].

Though science and engineering has not stood still and for decades electronic stethoscopes (el. stethoscopes) were being developed [8]. This allowed computer-aided auscultation to develop [9].

More recent rapid advancements in processing compute power on the back of Moor's Law with improved mathematical models has meant ever increasing breakthroughs and application of machine learning (ML) tools towards diagnostic field [10–13].

The synergic combination of electronic stethoscopes with artificial intelligence (AI), more specifically ML models are arisen as potential solution to improve lung auscultation diagnostic accuracy [6].

Yet there very few articles that compares human subjects' accuracy across large number of ML.

Therefore, a pivotal question when physician should seek ML assistance under varying noise conditions cannot be answered. Without Answering integration of these tools is problematic and can cause more problems than solutions it's going to resolve.

Furthermore, not all lung sounds are alike. There are two main types of auscultation sounds: normal (NAS) and pathological. Pathological auscultation sounds can be continuous (CAS) and discontinuous (DAS). The DAS' properties are heard as fine and coarse crackles, and CAS are audible as wheezes and bronchus sounds to the examiner's ear. The typical properties of CAS are typically 80 to 1600 Hz, lasting more than 250 ms, and are associated with asthma and chronic obstructive pulmonary diseases. DAS are shorter, typically less than 20 ms in duration, with a wide frequency range from 100 to 2000 Hz, and are associated with congestive heart failure (HF) and pneumonia [14].

This thesis delves deep into supervise ML models' ability to identify three different classes of lung sounds under three different levels of ambient noise and compares confusion matrices precision-recall (PR), receiver operating characteristic (ROC) parameters under a scrutiny of statical validation of these models to human subjects' ability, whilst utilising same proprietary dataset.

The wealth of knowledge generated by this work aim to contribute towards advancing knowledge of cost effective, no invasive, point-of-care into the future that has potential to expand into field of ambulatory respiratory health monitoring arena [15–17].

1. AIM AND OBJECTIVES OF THE STUDY

1.1. Main aim

To evaluate and compare the diagnostic accuracy of machine learning models and medical students after training on proprietary data to identify correctly three classes of lung sounds under three different levels of Gaussian white noise (GWN).

1.2. Objectives

1. To train and evaluate machine learning models' and medical students' ability to identify three classes of lung sounds under different levels of GWN.
2. To evaluate the influence of spectrogram and scalogram on 12 different supervised ML models' ability to accurately identify different classes of lung sounds.
3. To compare the ability of machine learning models and medical students to identify three classes of lung sounds under three different levels of GWN utilising key diagnostic metric.
4. To identify the potential of machine learning model to function as diagnostic assistant under GWN conditions for three main classes of lung sounds.

2. RELEVANCE, NOVELTY AND SIGNIFICANCE OF THE RESEARCH WORK

2.1. Relevance of the research work

Stethoscopes have existed for over 205 years [18]. The use of stethoscopes has allowed auscultation of the sounds produced by the body, though exact mechanism is still poorly understood due to lack of standardisation and subjectivity of the stethoscopes use [19]. These sounds can change due to pathologies ranging from gastroenterological, cardiovascular, renal or pulmonary in nature [20–23]. Stethoscope is particularity important in cardiopulmonary screening and this research work will look specifically and lung sounds. The pulmonary sounds associated with pathologies (adventitious lung sounds) assist physicians in preliminary diagnosis and decision-making regarding further tests and treatment the patient might need [23, 24].

Regrettably, there has been a noticeable decline in the practice of auscultation in recent times, posing a potential threat to the quality of patient care [5, 25].

Additional auscultation depends on a relatively quiet room. However, increased noise levels in healthcare settings can pose another challenge for effective lung sound auscultation [26].

Yet auscultation remains a cornerstone of preliminary primary cardiopulmonary examination and is extremely widely used in clinical settings [27]. Therefore, any diagnostic improvement in stethoscope accuracy, specificity, and sensitivity to identify lung sounds can lead to more exact diagnoses and better patient treatment [28]. Whilst lung diseases remains a third leading cause world wide [7]. Hence any improvement in pulmonary auscultation can lead to great positive impact on patients' health worldwide.

Recent developments in artificial intelligence (AI), combined with electronic stethoscopes, created conditions to gather data to train ML to standardise auscultation [29]. Therefore, this seems to be an answer in the current environment.

However, the robustness of human subjects or organic intelligence (OI) compared to ML models using the same lung sound datasets under standardised ambient noise pollution conditions has not been investigated. This is especially relevant as medicine seeks to integrate ML models to assist decision-making while diagnosing lung diseases in patients in real life settings healthcare sector.

2.2. Novelty of the research work

The research project is not only unique to Baltic region, but in is one of the kind to compare human subjects and machine learning models in identifying three classes of lung sounds under three levels of GWN.

Currently, there are no studies that compared 12 ML models to human subjects using lung sounds from the same datasets.

The research that exists on human subjects' ability to identify lung sounds under different noise level conditions is seldom.

The research articles that investigated the ambient noise effect on ML models' accuracy are seldom and some are over a decade old, whilst during this time ML models have been advancing and new tools are available and not yet tested under aforementioned conditions. Therefor it is not by utilising several different models and two sound representations it is possible to add novel insights which models could be the most robust to noise impact and their applicability in decision support tools' development.

The research on human subject is also sparse, over 5 years old and performed in various environments or on paediatric patients [30, 31].

Research articles even have contradictory conclusions, such as that most examiners' ability to hear heart and lung sounds is not significantly impacted by extreme loudness found in emergency departments [31].

A review article by Wallis Rory in 2019 concluded measurement of environmental white noise levels in hospitals are inconsistent and poorly reported [32]. The above aforementioned factors make it hard to test hypotheses by replicating methodology. Therefore, application of GWN utilisation as standardised ambient noise that covers all frequencies equally to compare human subject and ML models with same dataset is another first.

Whilst research that focuses on ML under different ambient noise conditions are also very few and three articles can be uncovered in literature review [10, 30, 33].

Additional, some studies do not even have statistically significant number of data point to perform statistical analysis [10].

Therefore, the combination of use of latest ML models, GWN ambient noise at three different levels and comparison to human subjects' ability whilst utilising same dataset makes is research absolutely unique.

2.3. The application of scientific work

The study shows the statistical and clinical significance improvement of training ML models and medical students to identify three main classes of lung sound (NAS, CAS, DAS). Furthermore, the study shows an impact of GWN on the DAS class of lung sounds, indicating that noise levels could significantly affect the ability to screen pathologies associated with these lung sounds, such as HF and pneumonia. Especially in light of the fact that even though WHO recommends hospitals to be around 35 dB they are usually much lauder and values can range at day time from 37 to 88.6 dB [26].

This suggests the need to include noise in improving the auscultation accuracy of healthcare workers and ML models. This is crucial for maintaining the importance of this non-invasive, widely available, easy-to-perform, low-cost stethoscope as a pillar of objective screening in a clinical setting. Understanding and accounting for noise could significantly enhance the effectiveness of lung sound identification.

Therefore, all future healthcare workers and ML should be assessed under WN conditions to evaluate the robustness of auscultation accuracy, specificity and sensitivity in screening for adventitious lung sounds in a clinical setting. This is especially important if future medical staff use ML in AI-powered clinical decision support [34].

3. LITERATURE REVIEW

3.1. Lung auscultation

3.1.1. Definition and description of lung auscultation and general introduction

Lung auscultation is a critical clinical tool used to assess respiratory health by detecting sounds generated during breathing [35]. These sounds, traditionally classified into normal and adventitious, provide valuable insights into a variety of respiratory and systemic conditions. Adventitious sounds such as wheezes, crackles, and pleural rubs are commonly associated with specific pathologies, making their detection crucial for diagnosing diseases like chronic obstructive pulmonary disease (COPD), pneumonia, heart failure (HF), asthma, hydrothorax, and renal failure and chronic kidney disease (CKD) [36–38]. Each of these conditions presents with characteristic lung sounds that guide clinical first line decision-making [28, 39].

For instance, in COPD and asthma, wheezes – continuous, high-pitched sounds – are often heard due to airway obstruction [40]. In contrast, discontinuous sounds like crackles are typically found in patients with conditions such as pneumonia, HF, and hydrothorax, where fluid or inflammation affects lung tissue [41]. Fine crackles, commonly present in HF, are associated with alveolar fluid build-up, while coarse crackles may suggest conditions like pneumonia [41]. In renal failure, fluid overload can similarly cause fine or coarse crackles, depending on the severity of the pulmonary involvement.

Despite the clinical importance of correctly identifying adventitious lung sounds, noise pollution frequently interferes with the accuracy of lung sound detection [42, 43]. Both environmental noise, such as conversations and equipment noise, stethoscope membrane rubbing on the skin and internal bioorganic internal noise, can obscure important auscultatory findings [44–46]. This interference can make it difficult to distinguish between normal breath sounds and the pathological adventitious sounds that are key to identifying diseases like HF or pneumonia [46].

Given the reliance on the stethoscope for primary screening of lung pathologies and its ubiquitous use equipment, whilst at the same time extremely poorly utilisation can lead to misdiagnosis, especially in a noisy healthcare settings [47, 48]. This literature review aims to examine the classification of lung sounds in relation to specific pathologies and explore the effects of noise pollution on auscultation ability of human subjects and introduce exploration of various ML models as potential solutions.

3.1.2. Lung sound classification

The lung can be classified into normal and abnormal (adventitious lung sounds).

Normal lung sounds can be further divided by their sound qualities and locations: tracheal, vesicular sounds, bronchial, bronchovesicular [49].

Tracheal breath sounds are characterised by their high pitch and loudness, with a hollow quality. They are most audible in the neck area. Vesicular sounds are primarily heard over the peripheral lung fields, especially over the lung bases posteriorly and laterally on both sides of the chest. The sound quality is soft, low-pitched, and rustling, with the inspiration phase being longer than the expiration.

Bronchial sounds are primarily heard over the trachea and near the manubrium of the sternum. They are characterised by loud, high-pitched, and tubular, with expiration often louder and longer than inspiration.

Bronchovesicular sounds are heard in the first and second intercostal spaces anteriorly and between the scapulae posteriorly, where the bronchi are close to the chest wall. The bronchovesicular sounds are a mixture of bronchial and vesicular, with inspiration and expiration almost equal in length and intensity [50].

These lung sounds are normal if they are heard in their normal location. However, if, for example, the bronchial sound is heard where only vesicular sound should be audible, this can be a sign of pathology.

The adventitious lung sounds can be classified into two broad groups: continuous and discontinuous [51]. Continuous auscultated sounds, such as wheezes, rhonchi, and stridor, are characterised by a musical quality and are typically associated with airway obstruction. These sounds can vary in pitch and duration, often lasting over 250 ms.

Discontinuous auscultated sounds, such as crackles or rales, are non-musical and characterised by brief, intermittent sounds typically lasting less than 20 ms [52].

Sadly, inconsistency persists in terminology, and even in English literature, “crackles” and “rales” are used interchangeably by pulmonary physicians.

The discontinuous lung sounds have a very broad range of frequencies, and in combination with their short duration, poorer identification of these sounds by healthcare workers as compared to continuous lung sounds.

Another example of inconsistency has been historical concerning continuous lung sounds with the terms like “wheeze” and “rhonchus” sometimes used interchangeably [53]. Therefore, even though the research will use the current definitions of normal, continuous and discontinuous lung sounds, it is important to understand that these terms are quite broad.

3.1.3. Adventitious lung sounds and associated pathologies epidemiology

Continuous lung sounds are associated with pathologies such as asthma and chronic obstructive pulmonary disease (COPD) [54].

Discontinuous lung sounds are often associated with conditions like pneumonia and HF, crackles (a type of discontinuous lung sound) can be found in early onset of COPD [54, 55].

3.1.4. Current diagnostic accuracy of physicians and AI models the stethoscope according to lung sound classes

The sensitivity and specificity for three classes of lung sounds vary quite significantly between AI models and physicians.

The sensitivity is poor, and specificity is suitable for normal (NAS) lung sound detection by physicians and AI models. CAS sounds like wheezing are much easier to detect for physicians and AI models with good sensitivity and specificity.

Finally, DAS lung sound detections by both AI and physicians have shown good sensitivities but poor specificity [56].

3.2. Machine learning models

3.2.1. Machine learning model definition and description

Machine learning (ML) models are algorithms that can trained on data to make correction decisions [57, 58]. ML models involve three main steps; training, validation and testing. The classical ML models can be categorised into two broad groups supervised learning and unsupervised learning [56]. In the latter group, an ML model tries to find structure in the data by itself via clustering or dimensionality reduction. Supervised models work on usually human-labelled data that allows the ML to know the target at the start of training; this type of method is very taxing on human resources and can have mislabel data; however, with well-prepared data, which is of paramount importance, it can lead to highly accurate models [59, 60].

3.2.2. Supervised machine learning model types

The main focus the thesis is to evaluate the performance of 12 supervised ML models compared to human subjects in recognising three classes of lung sounds under three levels of GWN.

Though there are a significant number of ML models that can be used to for lung sound analysis we will discuss the most commonly used.

Logistic Regression (LR) is one of the simplest models, Logistic Regression can still be very effective for binary classification problems [61]. Though for situations that needs model to classify outcomes into three or more categories a multinomial Logistic Regression can be applied.

Support Vector Machiness (SVM) is a classifier that finds an optimal hyperplane that separates different classes. Features such as spectral content or temporal characteristics of lung sounds are used for effective classification. This method has been has been widely use and achieved reasonable diagnostic accuracy of in a number of studies [62, 63].

K-Nearest Neighbors (KNN) is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The model was previously utilised in pulmonary sound research [64].

Random Forest (RF) is ensemble learning method uses multiple decision trees to classify data. For lung sound recognition, Random Forest can provide robust predictions by averaging multiple tree outcomes, reducing overfitting and handling noise in data effectively [65].

Extra Trees or Extremely Randomised Trees (ET) is an ensemble ML model that combines extensive number of different trees in similar fashion to RF but with additional randomisation. The ET has an advantage of working well with high dimensionality which can come in forefront whilst working with lung sounds. Though major drawback that this model has as we step away from simpler model such as RF it gets harder to understand why it works well or does not, so transparency is reduced and becomes less interpretable. This model has been used in categorising non speech sounds signals into seven different classes – breath, cough, cry laugh, sneeze, and yawn [66].

Extreme Gradient Boosting classifier (XGBoost) is a more sophisticated ensemble learning model than Random Forest. It belongs to family of gradient boosting algorithms [67]. The model has been successfully utilised in previous studies [68, 69].

Gradient Boosting (GB) or Gradient Boosting Machine is another ensemble learning model. GB produced a prediction model in the form of an ensemble of weak prediction models, typically decision trees. The main parameters that require fine tuning in this model are number of trees, maximum interaction between the independent variables and learning rate (shrinkage) [70]. The main advantage's is high accuracy and suitability with complex pattern recognition. The disadvantage is it is prone to overfitting especially if model is note fine-tuned.

Light Gradient Boosting Machine (LightGBM) in contrast to horizontal growth in XGBoost it carries out vertical growth that can reduce loss reduction and can lead to higher diagnostic accuracy. It is also fast and efficient, though this is more important in larger datasets. It works well with categorical data which can be advantageous with pulmonary long sounds. Though it suffers from overfitting. Research shows successful application of LightGBM in lung sound recognition using ICBHI-2017 database [71].

Categorical data Gradient Boosting (CatBoost) is an algorithm for gradient boosting on decision trees. The main advantages is its robustness to overfitting and applicability to work with categorical features. The previous research has shown to be more accurate than LightGBM and XGBoost [67].

Adaptive Boosting (AdaBoost) is also an ensemble learning method, it works by combining many weak classifiers into a single strong one. It shown to be less prone to overfitting which can be of benefit with lung sounds identification [72]. Though AdaBoost is sensitive to noisy data, outliers, data imbalance can cause a drop in performance [73]. It has been successfully utilised in lung sound classification [68].

Histogram-based Gradient Boosting Classification Tree (Histgradient) is another type of ensemble learning algorithm that builds tree models sequentially, with each new tree focusing on correcting the prediction errors of the previous trees in the ensemble. Histgradient has shown to have great robustness when it comes to concern of missing data in datasets, though it can be fidgety and requires optimisation through fine tuning and takes time to acquire most out of the model in terms of diagnostic accuracy. Histgradient has been utilised in emotion recognition from speech pattern changes by Nasifa T. Ira [74]. Therefore, this model holds a great potential in our research.

Multilayer Perceptron (MLP) is a neural network that is based on universal approximators. The network is created of components called nodes (neurons) otherwise known as perceptrons. The MLP classifier effectively models complex non-linear relationships, making it well-suited for capturing the intricacies of lung sounds. Though it requires a larger amount of data as before mentioned models, yet it has a lot of potential and will be a good reference point. Additionally, it has been used in previous research. MLP have demonstrated high accuracy in lung sound recognition tasks. For instance, MLP ML model achieved an accuracy of 99.22% on a publicly available respiratory sounds dataset, outperforming other machine learning classifiers [75].

3.2.3. Ambient noise and ML models

Several research studies have been performed on applying ML models in identifying lung sounds with different pathologies. However, they have been mostly done without adding white noise (WN). Few studies investigated various types of WN impact on lung sound recognition by human subjects or ML models utilising ambient noise ranging from ambulance car, babble noise, Gaussian white noise (GWN), background talking, crying, electronic, interference and artefacts produced by intentional, unintentional stethoscope displacements. Only three articles analyse specifically the impacts of WN on ML models' ability to classify lung sounds correctly in two research articles, the first by Gwo-Ching Chang and Yi-Ping and the second paper by Cheng Gwo-Ching Chang and Yung-Fa Lai [10, 33].

The third research article by Dimitra Emmanouilidou and SVM classifier concluded that the model was overwhelmed by background noises containing a weaker interference component and transient bursts of audio energy led to added confusion to the classification [76]. Results were summarised in Table 3.2.3.1 below.

Table 3.2.3.1. *Summary of studies investigating machine learning lung sounds under ambient noise conditions*

Study Title	Noise Types Examined	Methodology	Main Findings	Reference
Performance evaluation and enhancement of lung sound recognition system in two real noisy environments	The ambulance car noise and babble	Autoregressive and mel-frequency cepstral coefficients for feature representation; dynamic time warping for classification	The results indicate that additive noise produces a mismatch between training and recognition environments and deteriorates the classification performance with a decrease in the SNR levels	Chang G., Lai Y., 2010 [33]
Investigation of noise effect on lung sound recognition	Gaussian and two real noises (babble and car noises)	Autoregressive coefficients, mel-frequency cepstral coefficients, bispectrum diagonal slices; dynamic time warping for classification	Results showed that the bispectrum diagonal slices was more immune to noise interference in lung sound recognition, but the Mel-frequency cepstral coefficients was more vulnerable to noise disturbance	Chagn G., Cheng Y., 2008 [77]
Characterization of noise contaminations in lung sound recordings	Ambient noise, background talking, crying, electronic interference and artifacts produced by intentional or unintentional stethoscope displacements	A spectro-temporal signal representation followed by a standard SVM classifier	The results show that noise contamination in recordings have distinct features and can be discriminated	Emmanouilidou. D., Elhilali M., 2013 [78]

3.2.4. Lung sound representations

The lung sounds can be directly processed by machine learning models, though previous research shows that best results are obtained by extracting features from the representations and using this datasets to train the models [79]. The two main representations that can be used for visualisation and extraction of a biological audio signal are spectrograms and scalograms. ML models mainly use spectrograms [80]. However prior research indicates that scalogram's using the Wavelet function compared to Fourier transformation in spectrograms can localise representation of time and frequency in a better fashion, hence, leading to better ML model accuracy [81].

Though, it is worth mentioning that machine learning models most commonly utilise spectrograms. Scalograms, therefore, might require more fine tuning or even bigger datasets to produce similar or superior diagnostic results, whilst using same data.

An illustration bellow shows representations of three main classes of lung sounds that our research project will focus on (Fig. 3.2.4.1).

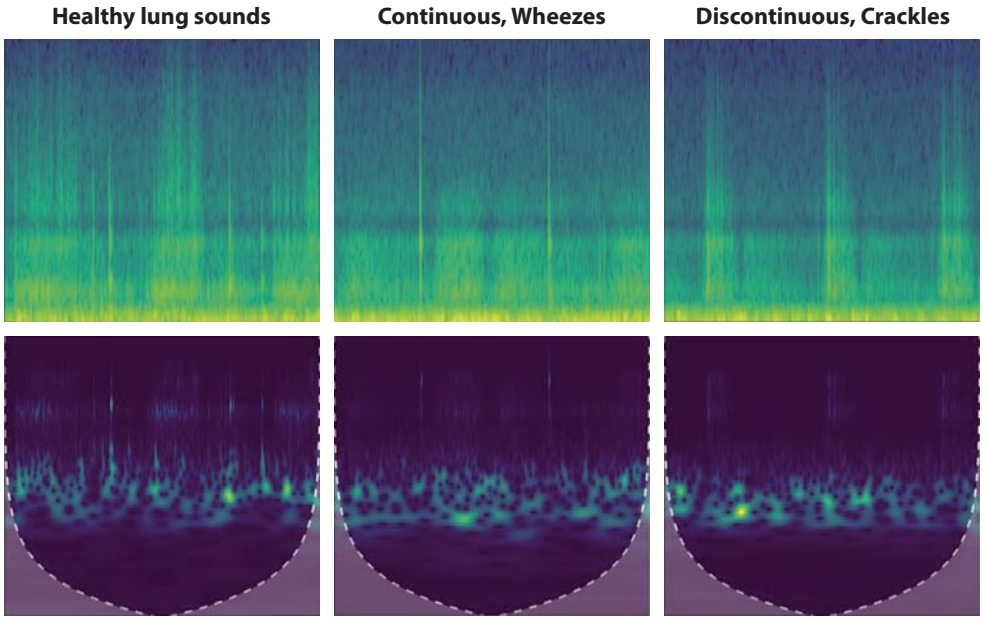


Fig. 3.2.4.1. *Illustration two types of representations of lung sounds (scalogram and spectrogram) for normal (healthy), and pathological (continuous and discontinuous lungs sounds)*

3.3. Ambient noise

3.3.1. Background noise definition, description and impact on auscultation accuracy

Ambient noise, in our case, is any sound other than lung sounds. They can be caused by background talking, stethoscope movement [82]. This type of noise reduces physicians' or machine learning models' ability to discriminate between different classes and types of lung sounds. Therefore, reduces the overall accuracy of the stethoscope as a diagnostic tool [83].

3.3.2. Background noise types and levels in healthcare settings

Background noise can have extracorporeal and intracorporeal origins and can be organic or inorganic in nature. The internal organic background noise is biological sounds in source, such as, heart sounds, active peristalsis, patient starting to talk or cough during auscultation process.

The external factors can be organic and inorganic in nature. The organic external nature sounds can be healthcare staff, children crying, other patients having a conversation in the corridor. An external inorganic sounds can be trollies, ambulance car or medical equipment sounds. Additionally, noise can be produced by chest piece of the stetscopes (diaphragm) rubbing against dry skin or hairs of the body [84].

The prior research has used several type of noise, babble, emergency room noise, ambulance vehicles, fake crackles produce by membrane, Military equipment helicopters and Gaussian white noise (GWN) [77, 85–87].
From the above-mentioned, ambient sounds used in research, GWN stands out as it is not only the only synthetic noise but also an ideal candidate for the sonic pollution impact of lung sound identification, as it pollutes each frequency evenly. Hence, even though it is synthetic, it can provide a great baseline to evaluate human and machine learning models' accuracy before moving to more specific sounds that could be applicable at particular settings, such as ambulance sirens, conversation or helicopter blades spinning.

Therefore, GWN assist in achieving standardisation of noise pollution across all frequencies evenly in our methodology.

3.4. Human subjects selection for the study

The auscultation is still usually performed inside hospitals by physicians, nurses, resident doctors, medical nurse and medical students. The main two reasons why acoustic stethoscopes are still preferred to electronic stethoscopes (el. stethoscopes) are due to costs, and synthetic sounds transmitted by variety of el. stethoscopes, whilst not showing clear advantages in diagnosing pathologies [88]. Physicians could be potentially used in the study to identify lung sounds, though, there are several factors that do not make them ideal candidate. First of all, physicians are trained already to identify lung sounds, they have different work hours, are different ages and with age there is an increased risk of hearing impairment that would need to be accounted in the study [89]. The study needs motivated human subjects, that have no prior auscultation skills, but are willing to learn the lung sounds, and in large enough numbers. Hence, medical students are ideal subjects for this type of study.

Previous study shows, though students diagnostic accuracy is lower than physicians, it follows similar pattern where both physicians and medical students have lower ability to identify crackles (DAS class of lung sounds) compared to wheezes (CAS class of lung sounds) therefore, results can provide as with inference in understanding how physicians and nurses could be affected too [90].

3.5. Studies examining impact of ambient noise on human subjects ability to accurately auscultate

There are only a few studies examining impact of ambient noise on human subjects ability to accurately auscultate. The first study by Peitao Ye in 2022, assesses 56 participants' ability to correctly identify a discontinuous class of lung sound, whilst auscultating in the presence of fake crackles [86]. The fake crackles are generated when the stethoscope membrane glides over the skin. The article concludes that these crackles can lead to misdiagnosis.

A review paper by Jun J. Seah in 2023, primarily focuses on the advancements of stethoscopes in auscultation. While it acknowledges the impact of extreme noise in disaster zones, chaotic situations, and helicopters it falls short of providing detailed insights into the effects of different classes of respiratory lung sounds, leaving a gap in our understanding of white noise's influence on auscultation, especially in medium levels of ambient noise that are experience inside hospitals [91].

Further literature investigation on topic of ambient noise effects on human subjects' ability to auscultate revealed only two more older articles focusing on specific ambient noise. The first by Steven Gaydos in 2011 looked at military helicopter's spinning blades and concluded that extreme noise produced in-flight makes pulmonary auscultation a futile endeavour [92]. A more applicable research to civilian settings was performed by Jörg D. Leuppi in 2005 research on 137 patients (though only male), showed low value of the stethoscope in a noisy emergency reception (ER) as a diagnostic tool, but at the same time surrounding ambient noise did not impact final diagnosis. Nonetheless, the article concluded that normal lung auscultation results are a valuable predictor for not having a lung or heart disease. In contrast, wheezing was a predictor of having a lung disease [93].

These studies are not satisfactory enough to understand at what levels the ambient sound will start influencing the auscultator's ability to classify lung sounds correctly. Therefore, it emphasises the importance of standardised conditions, with set noise self-pollution at exact signal-to-noise ratios (SNR). The literature review was summarised in Table 3.5.1.

Table 3.5.1. *Summary of studies investigating lung sounds recognition by health workers under white noise conditions*

Study Title	Noise types examined	Methodology	Main Findings	Reference
Clinical auscultation in noisy environments	Various sounds but mainly military helicopter's spinning blades	Case report: a patient with blast injury developed hemodynamic instability of unclear aetiology during transport in the combat aviation environment	Auscultation of breath sounds while in-flight would be futile due to the high ambient noise of the helicopter	Gaydos S., 2011 [92]
Regularity and mechanism of fake crackle noise in an electronic stethoscope	Fake crackles	Prospective double-blind study	Stethoscope membrane generated fake crackles gliding over the skin lead to misdiagnosis of discontinuous lung sounds	Ye P., et al. 2022 [86]
Diagnostic value of lung auscultation in an emergency room setting	Various natural emergency reception sounds	Prospective study	Abnormal lung auscultation does not appear to contribute considerably to the final diagnosis in these patients. However, normal lung auscultation is a valuable predictor for not having lung or heart disease, whereas wheezing is a predictor for having a lung disease and not having a heart disease	Leuppi J.D., et al., 2005 [93]

4. RESEARCH METHODOLOGY

4.1. Study design, study location, inclusion and exclusion criteria

Prospective study carried out in Lithuania in 2020–2024.

Study subjects: Patients hospitalized with adventitious lung sounds and diagnoses confirmed for pneumonia, HF, COPD, asthma, kidney failure or CKD, hydrothorax. Patients were diagnosed according to international protocols [95–104].

Location of the study for lung sound collection and medical student enrolment: the study was conducted in the Cardiology and Internal Medicine Diagnostic Departments of Lithuanian University of Health Sciences Kaunas Hospital (Josvainių 2 and Hipodromo 13 Kaunas).

The total inpatient bed fund in 2020 was 1,620 beds. Forty-two thousand sixty-four patients were treated. During the pre-pandemic period (2019), the hospital provided about 60,000 inpatient healthcare services, and this number is projected to return post-pandemic [105].

Partial research was performed in collaboration with Prof. Evaldas Vaičiukynas and his colleagues from Kaunas Technology University (KTU), with sponsorship from the education and research funds of Kaunas University of Technology (Grant No. PP2023/39/4) and Lithuanian University of Health Sciences.

Inclusion criteria for lung sounds recording:

1. patient diagnosed with pneumonia;
2. patient diagnosed with asthma;
3. patient diagnosed with heart failure;
4. patient diagnosed with kidney failure;
5. patient diagnosed with COPD exacerbation;
6. patient has adventitious lung sounds;
7. patient 18 years or older;
8. patient with no mental disorder;
9. the patient was conscious and able to answer questions correctly;
10. the patient has signed the personal information form (PIF) and the informed consent form (ICF).

Inclusion criteria for medical students:

1. LSMU medical students in their second or third year;
2. participants 18 years or older;
3. medical students that have no prior experience with auscultation and agree to participate by signing ICF.

Exclusion criteria for lung sounds recording:

1. patients that refused to participate in the study;
2. patients who could not speak Lithuanian and provide consent;
3. patient that could not stand, sit still for auscultation to be performed.

Exclusion criteria for medical students:

1. students with hearing impairment or loss;
2. students over 40-year-old;
3. students that did not sign the consent forms.

4.2. Study sample size calculation

The sample size for the lung sounds recordings and the medical students was calculated using G*Power software (ver. 3.1.9.4; Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany) [106, 107].

Due to a lack of studies, the sample size calculations for the medical students were based on a proprietary pilot study. The software utilised the following settings to calculate the means: Wilcoxon signed-rank test (matched pairs) function. The following assumptions were applied: power ($1 - \beta$ error probability) at 0.95 and an α error probability of 0.05. The effect size (Cohen dz) from the pilot study was 0.61, based on pre-and post-training means and standard deviations (SD) of 4.80 ± 0.49 and 5.07 ± 0.36, respectively. These values were inputted into the function, resulting in a sample size of 33 subjects. The pilot study had an attrition rate of 30%. Therefore, accounting for attrition, the total number of subjects required was 48.

The sample size for lung sounds recordings was calculated based on the assumption of the effect size to be 0.50, power ($1 - \beta$ error probability) at 0.95 and an α error probability of 0.05, with the number of groups set at 3. The G*Power software (ver. 3.1.9.4; Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany) function was set at ANOVA: fixed effect. The input resulted in the value of 85 subjects (including the control). The recording of lung sounds had to undergo a double-blind review, assuming that the screening group was out of around 30%, which means that around 122 subjects needed to be enrolled in the study.

4.3. Study enrolment methodology

Individuals who agreed to participate in the study were briefed and signed the PIF and ICF. During the study, data were collected according to the patient survey questionnaire (see Annex 1). All subject data was gathered by doctoral student. All subject medical data was coded and accessible only to the principal investigator (doctoral student).

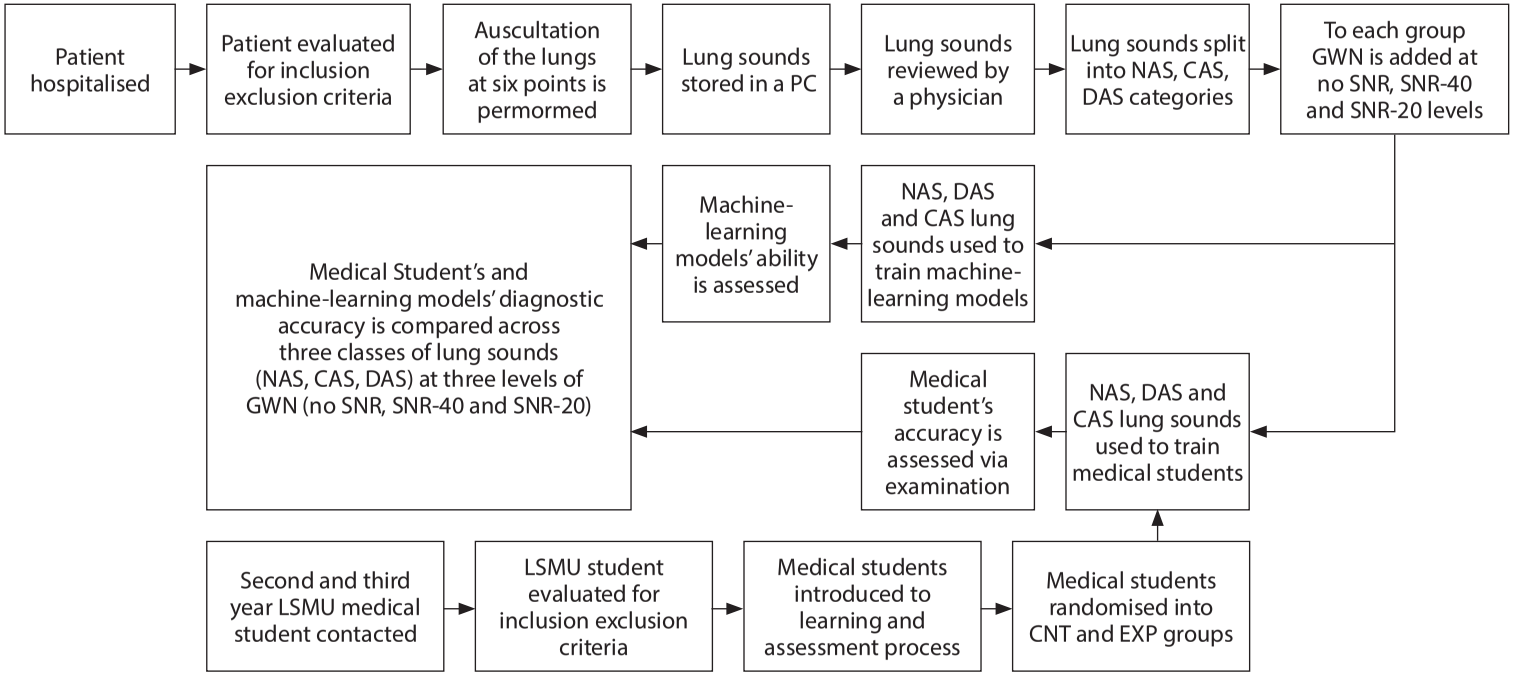


Fig. 4.3.1. *Flowchart for inclusion of lung sounds and medical students into the study, which compares machine learning models' and medical students' diagnostic accuracy across three classes of lung sounds under three different levels of Gaussian white noise*

CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, NAS – normal auscultated sound, SNR – signal-to-noise ratio, GWN – Gaussian white noise.

4.4. Pulmonary sound recording

A 3M™ Littmann® CORE digital stethoscope (3M Company, St Paul, Minnesota, United States), Microsoft®, Windows® 10 Operating System software (Microsoft Corporation, Redmond, Washington, United States) based HP ProBook 450 G4 (HP Inc., Palo Alto, California, United States) Intel® Core™ i5 i5-7200U (Intel Corporation, Santa Clara, California, United States) laptop was used to store audio files via 3M™ Littmann® StethAssist-1.3.230 (3M Company, St. Paul, Minnesota, United States) software.

The auscultation sound recordings were performed over approximately three months. The electronic stethoscope settings were as follows: mode was set to the diaphragm, and sound amplification was set to level 3 (the implication is up to level 9). The investigator performed the recordings in the wards, usually containing 2 to 4 patients. All patients were in stable condition and treated in the department for their underlying disorders. Patients with pancreatitis or severe hypertension were primarily the sources for normal lung sound recordings. Patients with pathological lung sounds were diagnosed and being treated for: Pneumonia, COPD, Asthma, HF, Hydrothorax, CKD. When the noise levels rose to hinder auscultation due to reasons such as a trolley passing, the nurse entered the room, the lung sound was rerecorded. Audio recordings were 15 s long each and stored in a waveform audio file format (WAV). Six recordings for each patient were performed from the back of the chest (Fig. 4.4.1).

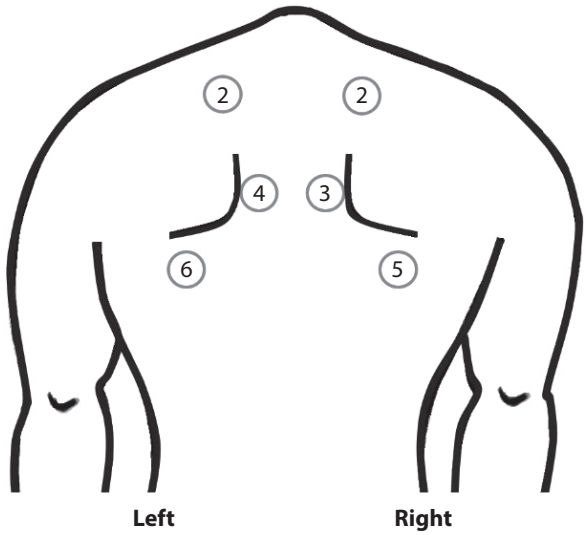


Fig. 4.4.1. *Illustration with six sites on the back of the chest where the 15 s lung sounds were recorded from*

4.5. Lung sound processing

In a double-masked review, a team of family and internal medicine physicians assessed the quality of the sounds and whether they were normal or pathological in nature. The quality of the sound recording was rated: “audible” or “inaudible” for training and assessing ML models. The internal and family doctors independently had to agree on the sound being “audible” and their type and class so that they could be included in the respiratory sounds database. Medical specialists reviewed lung sounds using the same Sennheiser HD 560S headphones (Sennheiser Electronic GmbH & Co. KG, Wedermark, Germany).

Of the 124 subjects and 744 recordings, only 250 recordings were suitable for ML models: 130 recordings for NAS, 70 for CAS and 50 for DAS. The sound descriptions and WAV files were securely stored in an encrypted Microsoft® Excel® (Microsoft Corporation, Redmond, Washington, United States) software database and audio folder, respectively. This database only contained essential patient information: age, gender, clinical diagnosis, audio file name and lung sound description. The data was held on the Internal Medicine Clinic’s password-locked laptop, ensuring its safety and confidentiality.

To evaluate medical students and ML models robustness to different levels of signal-to-noise ratio (SNR), Gaussian white noise was added to each recording according to Samit Ari methodology [108]. The assessment of the classification performance can be based on class indices, such as sensitivity, specificity and precision, which describes the classification results achieved on each modelled class. However, in several situations, it is useful to represent the global classification performance with a single number. Therefore, several measures have been introduced in literature to deal with this diagnostic assessment problem. These metrics have been proposed to generally face binary classification tasks and can behave differently depending on the classification scenario. In this study, different global measures of classification performances are compared by means of results achieved on an extended set of real multivariate datasets. The systematic comparison is carried out through multivariate analysis. Further investigations are then derived on specific indices to understand how the presence of unbalanced classes and the number of modelled classes can influence their behaviour.

Finally, this work introduces a set of benchmark values based on different random classification scenarios. These benchmark thresholds can serve as the initial criterion to accept or reject a classification model on the basis of its performance according to Samit Ari methodology [108].

The following three levels of GWN levels were used: no GWN, SNR-40 (medium GWN level at 5 dB) and SNR-20 (high GWN level at 25 dB). The SNRs were as follows: no GWN had a signal of approximately 45 dB and noise at 0 dB; SNR-40 had a signal of approximately 45 dB and noise at 5 dB; and SNR-20 had a signal of approximately 45 dB and noise at 25 dB. The GWN was added across the frequency spectrum from 31.25 to 1968.75 Hz. Audacity® (Muse Group, Limassol, Cyprus) was used to visualise waveforms and spectrograms (Fig. 4.5.1).

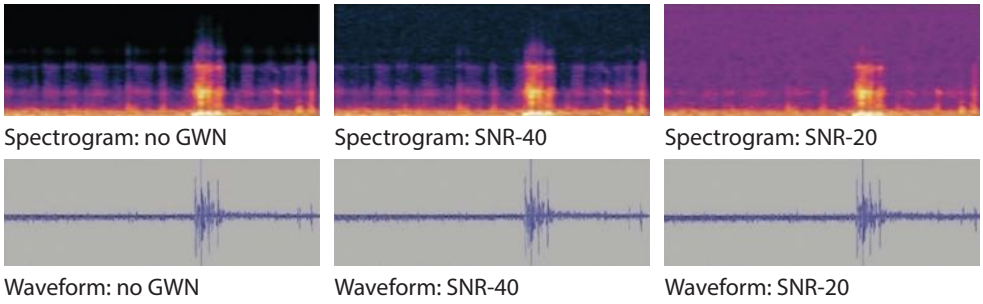


Fig. 4.5.1. *Flowchart visualisation of GWN levels added to lung sounds. Spectrogram (top row) and waveform (bottom row) analysis from one 15 s recording. Brighter backgrounds in the spectrogram indicate increasing GWN intensity from lowest (no GWN), medium SNR-40, to highest SNR-20 (left to right column)*

SNR – signal-to-noise ratio, GWN – Gaussian white noise.

4.6. Lung sound preparation for training and assessing human subjects

A website was created with training and examination sections for the MFS subjects and this platform had been successfully utilised in previous pilot study [109]. The training section of the website featured a pictogram of a chest with six clickable points, allowing students to listen to lung sounds, effectively creating web-based virtual simulated patients (Fig. 4.6.1). The information presented to the students was anonymised; only the patient's age and gender were included, along with details regarding the lung sound. The training section contained 101 lung sound recordings, of which 54% were DAS and CAS. The examination section was randomised and included 54 sound recordings, comprising equal proportions of NAS, CAS and DAS classes of lung sounds. Prior to the pilot study, the website was tested with 15 students to assess its functionality during a dry run and to collect data for sample size calculation. A pulmonologist reviewed the website. Enrolment involved 52 second- and third-year medical students who met specific

enrolment criteria and provided informed consent. After 4 days of training subjects were assessed for ability to correctly identify NAS, CAS, DAS via 3 exams, each having different levels of GWN (no GWN, GWN at SNR-40, GWN at SNR-20).

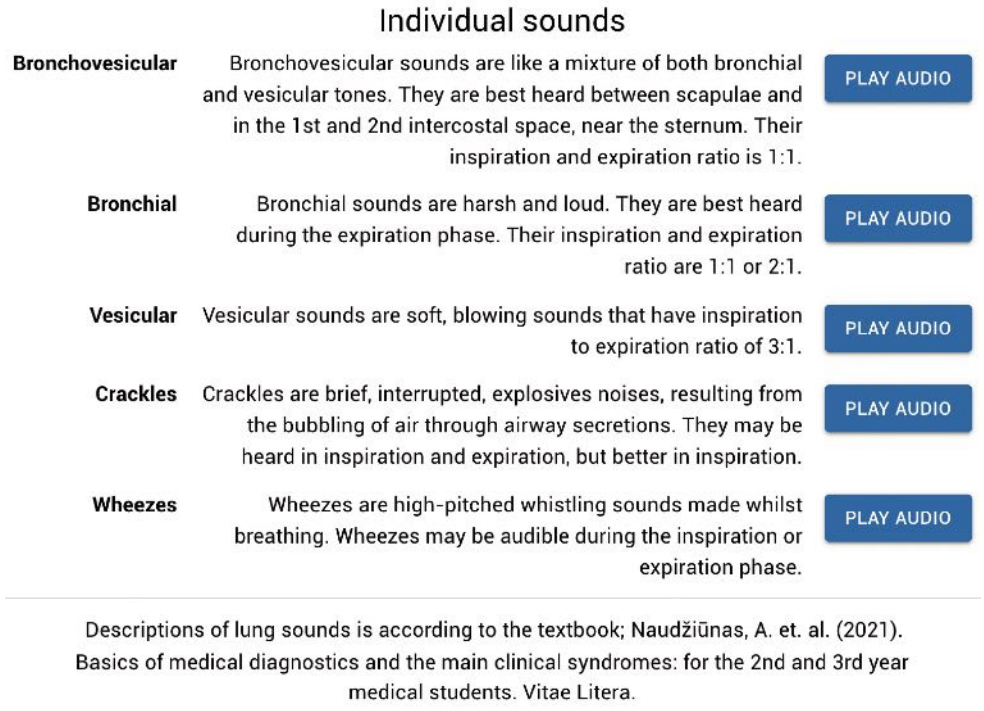


Fig. 4.6.1. *General website layout*

4.7. Lung sound preparation for training and assessing machine learning models

The sound files were converted to spectrograms and scalograms. Features were extracted at 3-second intervals from spectrograms and scalograms (the approximate average expiration/inspiration time ratios being 1.0 and 3.4, respectively) [110]. The following features were extracted from each of 250 recordings: average value of scalogram coefficients (mean), variability of scalogram coefficients (standard deviation), tailedness of the distribution of coefficients (kurtosis), asymmetry of the distribution of coefficients (skewness), central value of scalogram coefficients (median), most frequently occurring value in the coefficients (mode), smallest coefficient value (minimum) and most considerable coefficient value (maximum). The 450 extrac-

ted features each 15 s recording. Then, the each of 250 legers were categorised with a class label columns (NAS, CAS or DAS) and saved to a comma-separated values (CSV) file ready to be fed into ML models.

4.8. Machine learning models

A total of 12 supervised learning ML models were used: AdaBoost, CatBoost, ET, GB, Histgradient, K-NN, LightGBM, LR, MLP, RF, SVM, XGBoost. The models were chosen due to their potential and being previously used in researches related to lung sounds or other auditory bio signals recognition along with their ability to be applied to smaller datasets.

The models were trained using methodology that utilises extracted features from scalograms and spectrograms [111].

The models selection ranged from simplest such as K-NN to more sophisticated models such as XGBoost classifier. The K-NN is instance-based learning model that uses a nonparametric classification algorithm and is relatively efficient with small datasets. Another model that held a great potential was SVM because in past research it done well with small datasets that have more features than cases, as in our project that has 250 lung sounds with 450 features. The SVM model also shows robustness. LR is a classical statistical method that uses a linear model to predict binary classes. RF uses multiple decision trees during training. It also shows excellent robustness. More sophisticated models such as XGBoost and Histgradient implements gradient boosting parameters to improve performance over models such as RF regarding model accuracy. This model uses non-linear relations in modelling and has the potential to identify more subtle differences in lung sounds this could assist the model in making the correct predictions.

4.9. Hardware utilised for training and assessing machine learning models

A custom-built PC running Windows® 10 operating system (Microsoft Corporation, Redmond, WA, USA) was used, equipped with an Intel® Core™ i7-12700K processor, 64 GB of RAM, and an NVIDIA GeForce RTX 3060 graphics card with 12 GB of VRAM (NVIDIA Corporation, Santa Clara, CA, USA).

4.10. Model training and testing

GWN was added using Anaconda® (Austin, TX, USA) with Jupyter Notebook 6.4.7 utilising Python packages for machine learning training and assessment. Audio features were extracted using the Python library on to a comma-separated values (CSV). Lung sounds were labelled in double-blind setting. The datasets were split into 80/20 ratio for training and testing [112].

The split data contained even proportional number of NAS, CAS, DAS lung sounds at three different levels of GWN (no GWN, SNR-40, SNR-20). During cross-validation, the training data were partitioned into nine folds to ensure a similar distribution of the target classes in each fold and to improve the ML models. Due to class imbalance, stratification of the dataset was critical to ensure that each block of data contained representatives from each category [113]. The performance metric for each fold was collected and averaged at the end to provide the best evaluation of the model's performance.

In total, 30 iterations (runs) were performed for each model, including handling class imbalance, performing cross-validation, training the models, and calculating the performance metrics [114]. Once the best model was selected out of the 24 potential variations (12 ML models based on spectrograms and 12 based on scalograms), the model was fine-tuned again and 45 runs were performed, for average MMC calculations.

4.11. Performance metrics

The following performance markers were used to assess the validity of the models. Particular importance on ROC-AUC, PR-AUC, MCC, F1-score, TP, TN, as the study examined the best ML models' diagnostic validity.

True positive (TP): is a measure of classification of adventitious sound identified in patients with true adventitious lung sounds. False positive (FP) measures the classification of normal lung sounds as adventitious lung sounds. True negative (TN) measures the classification of normal sounds as normal, and false negative (FN) states adventitious sounds as normal.

Sensitivity (Sens), also known as recall, is a measure of the ability of the model to correctly detect positive cases of pathological lung sounds.

$$\text{Sens} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad [115]$$

Specificity (Spec), is the measure of the model's ability to identify a negative test result given that lung sounds correctly are normal.

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \qquad [116]$$

False positive rate (FPR) is the opposite of specificity:

$$\text{FPR} = 1 - \text{Spec} \qquad [117]$$

ROC-AUC: measures the overall performance of a model to classify lung sounds as normal or pathological. This is a robust measure, relying on all possible classification thresholds. The ideal classifier would produce a point in the top-left corner of the ROC space, indicating maximum sensitivity and specificity (minimal false positive rate). In contrast, a random classifier would generate points along the diagonal of the ROC space, extending from the bottom-left to the top-right corner [118]. The AUC can be approximated by summing up the areas of the trapezoids formed between points on the receiver operating characteristic (ROC) curve:

$$\text{AUC} \approx \sum_{i=1}^{n-1} \frac{(\text{TPR}_i + \text{TPR}_i + 1)}{2} \times (\text{FPR}_{i+1} - \text{FPR}_i) \qquad [119, 120]$$

Accuracy (Acc), is a measure of the overall correctness of the model by calculating the proportion of correct predictions (both true positives and true negatives) out of all predictions. This is a good measure if classes are balanced. Though if dataset suffers from imbalance MCC is preferred [118].

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad [121]$$

Positive predictive value (PPV) also called precision, is a measure of the correct positive predictions:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad [122]$$

F1-score – measures the harmonic mean of precision (PPV) and recall (sensitivity). This measurement is widely use in machine learning and is especially useful when trying to understand models diagnostic accuracy trained and assessed on imbalanced datasets.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad [123]$$

Cohen's Kappa is a statistical measure designed to evaluate the level of agreement between two evaluators or classifiers, adjusting for the possibility of agreement due to random chance. It is commonly employed in classification tasks to gauge model performance, particularly in cases involving imbalanced datasets.

$$\kappa = \frac{1 - Pe}{Po - Pe}$$

where: κ – Cohen's Kappa;
Po – is the proportion of instances where the evaluators agree;
Pe – the proportion of agreement expected by chance [124].

Matthews correlation coefficient (MCC) was mainly used in binary classifications, but it has been adapted in multivariant studies too. It is another excellent measurement for use when datasets are imbalanced. It incorporates TN, TP, FP, FN in calculations and gives values ranging from –1 to 1. The interpretation of MCC values is provided bellow, as adapted from Natarjan Meghanathan article [125].

Table 4.11.1. *MCC Value interpretation*

MCC value	Interpretation
0.80 to 1.0	Very strong positive (model almost always has correct prediction)
0.60 to 0.79	Strong positive
0.40 to 0.59	Moderate positive
0.20 to 0.39	Weak positive
0.00 to 0.19	No better than random guessing
–0.19 to –0.01	Very weak negative
–0.39 to –0.20	Poor classification
–0.59 to –0.40	Moderate negative
–0.79 to –0.60	Strong negative
–1.00 to –0.80	Very strong negative (model predicts completely opposite direction from expected results)

MCC – Matthews correlation coefficient.

4.12. Statistical analysis

The data for ML models and LSMU medical students were analysed using a Microsoft® Excel® (Microsoft Corporation, Redmond, Washington, United States) spreadsheet and the JASP (ver. 0.18.3; Jeffreys' Amazing Statistics Programme, The Jamovi project, Sydney, Australia) statistical package [126]. Additionally, IBM® SPSS® ver. 29 (IBM Inc., Armonk, New York, United States) was also utilised to complement analysis via JASP. A *P*-value of 0.05 or below was considered statistically significant. The results were presented in tables and summarised in a box-and-whisker plot.

During data cleaning, seven subjects were excluded from further statistical analysis for not completing all three assessments. Therefore, statistical analysis was performed on 45 out of 52 subjects.

The results did not adhere to a normal distribution; therefore, nonparametric tests were used for further analysis of median values. The Wilcoxon rank-sum test assessed the effect of training on students' ability to discern lung sounds accurately, whilst Friedman's test was used to analyse the impact of the three GWN levels on different lung sound classes with two degrees of freedom. Finally, a post hoc comparison was performed to evaluate the ability of medical students to recognise the lung sound classes (NAS, CAS and DAS) separately under the three different levels of GWN.

For human subjects naïve second- and third-year medical students were used. The data was collected from a proprietary website on which students were trained and assessed using MongoDB® (MongoDB, Inc., New York City, NY, USA) software along with entering it into a Microsoft® Excel® (Microsoft Corporation, Redmond, Washington, United States) spreadsheet for statistical analysis.

The ML model performance was recorded utilising Anaconda® (Austin, TX, USA) with Jupyter Notebook 6.4.7 utilising Python packages for machine learning training together with assessment and saved in comma-separated value (CSV) format.

For comparison of machine learning tools Friedman test was applied with post hoc pairwise comparison to compare the diagnostic accuracy of 24 different variations of ML models. To compare 12 spectrogram and 12 scalogram based ML models Wilcoxon signed-rank test was used.

Finally, Friedman test was applied with post hoc pairwise comparison to compare best ML model and medical students scores.

4.13. Bioethics of research

Permission for the study was obtained from Kaunas' Regional Bioethics committee (P1-BE-2-57/2021). The bioethics permission transcript has been attached in the appendix (Annex 2). The study was registered on the Clinical trials website (ID NCT05731193) and published on Good Clinical Practice Network.

4.14. Financing of the study

Partial research won sponsorship of €20,000 from joint funds of Kaunas University of Technology (Grant No. PP2023/39/4) and Lithuanian University of Health Sciences.

5. RESULTS

5.1. Descriptive statistics

Table 5.1.1. *Descriptives showing lung sounds population*

Adventitious lung sounds	Female	Female Age (SD)	Males	Male age (SD)	Overall	Overall age (SD)
NAS	26	69.5 (16.9)	26	56.5 (18.6)	52	63.0 (18.0)
CAS	10	75.5 (8.4)	13	66.0 (12.2)	23	70.1 (11.5)
DAS	12	78.7 (12.3)	21	69.0 (11.7)	33	72.5 (12.7)
Overall	48	73.1 (14.7)	60	62.9 (16.0)	108	67.4 (16.2)

SD – standard deviation.

Table 5.1.2. *Descriptive analysis by gender and age of medical students*

Female	Female Age (SD)	Males	Male age (SD)	Overall	Overall age (SD)
32	21.9 (2.4)	13	21.6 (3.1)	45	21.8 (2.6)

SD – standard deviation.

5.2. ML model performance

In total, 24 machine models' variations were tested with spectrogram and scalogram visualisations, under three levels of GWN noise (no added noise, GWN SNR-40 and GWN SNR-20). The impact of GWN was monitored on three main classes of lung sounds: NAS, CAS, DAS.

To display and comprehend performance of ML model, three main methods were used: confusion matrix, PR-AUC, ROC-AUC. The models were tested for overall impact of GWN on their performance via Friedman test.

From Fig. 5.2.1, the impact of Gaussian white noise is clearly observed in true positive (TP), true negative (TN), false positive (FP), and false negative (FP), as seen in the Confusion matrix of spectrogram based AdaBoost model. At no GWN added levels, 47/70 CAS are identified correctly, 25/50 of DAS class is correctly identified, and 96 of 130 are correctly identified. The confusion matrix at SNR-40 shows a sharp decrease in the performance of spectrogram-based AdaBoost models' performance with only 3/70 CAS correctly identified, the slightly more significant number of DAS correctly identified 27/50 and 110/130 of NAS correctly identified. The drop in values at the GWN SNR-20 level for DAS with 0/50 was identified correctly, and

CAS showed abysmal performance compared to no GWN at 6/70. Meanwhile, NAS class sounds improved, with 120/130 being correctly identified. Though the prior statement is true, we can see that the model has a tendency to label all three classes of lung sounds as NAS at GWN SNR-20, showing serious issues with the correct classification at higher levels of ambient noise.

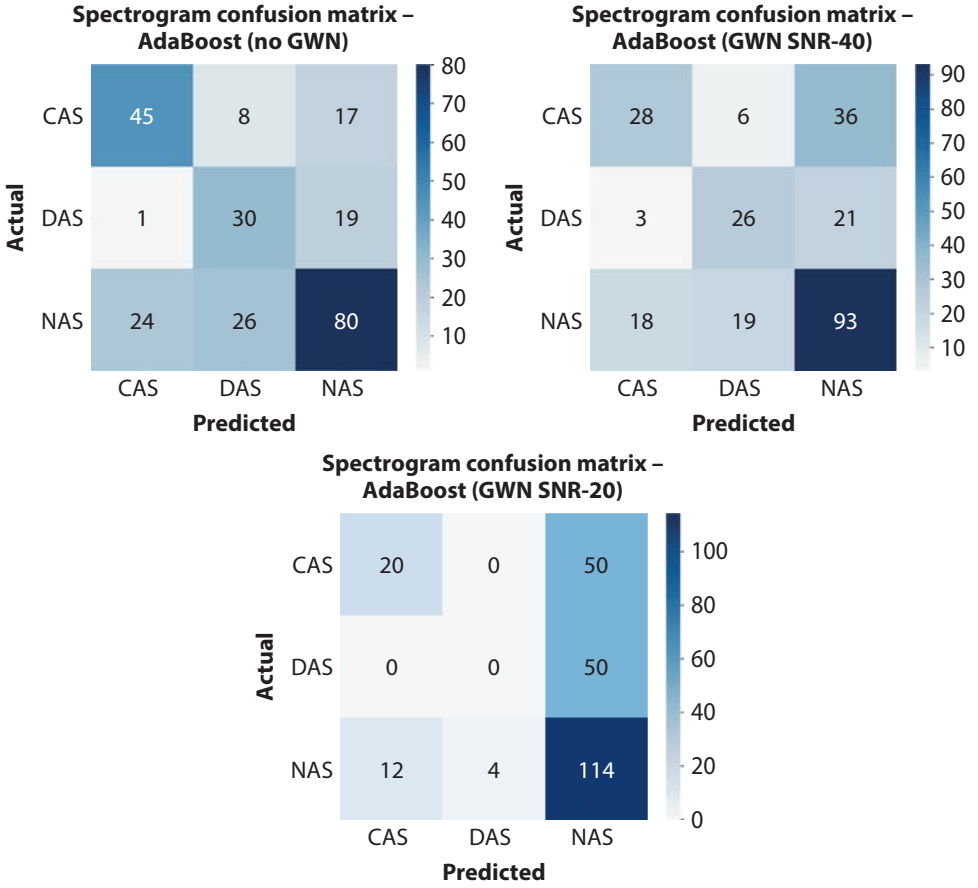


Fig. 5.2.1. *Confusion metrics showing significant impact (P = 0.000) of GWN on AdaBoost models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, AdaBoost – Adaptive Boosting.

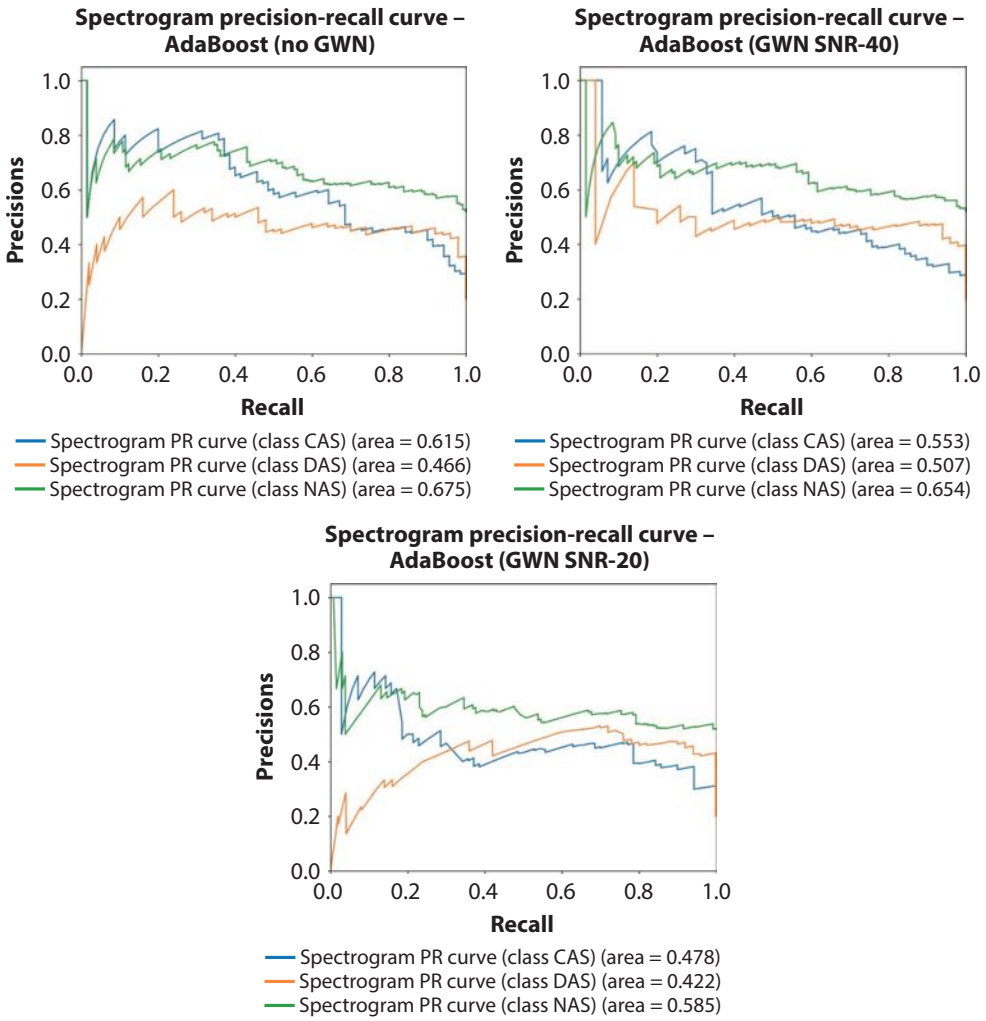


Fig. 5.2.2. *PR curve showing significant impact (P = 0.000) of GWN on AdaBoost models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, AdaBoost – Adaptive Boosting.

From Fig. 5.2.2, it can be observed that spectrogram-based AdaBoost struggles to achieve good precision to recall values for all classes, but the DAS class performance is significantly worse with much lower PR-AUC out of the three. At medium levels of GWN (SNR-40), the NAS and CAS lung sound classes identification are impacted, as exemplified by a drop in precision compared to recall and lower PR-AUC. However, DAS class has not been negatively impacted. Finally, once GWN is increased to SNR-20

levels, all three classes of lung sounds are heavily and negatively impacted. DAS class sound identification bears the biggest brunt of the impact, and precision compared to recall rates drops off into an abyss. Therefore, the model shows overall poor precision to recall performance with limited robustness even at medium levels of GWN.

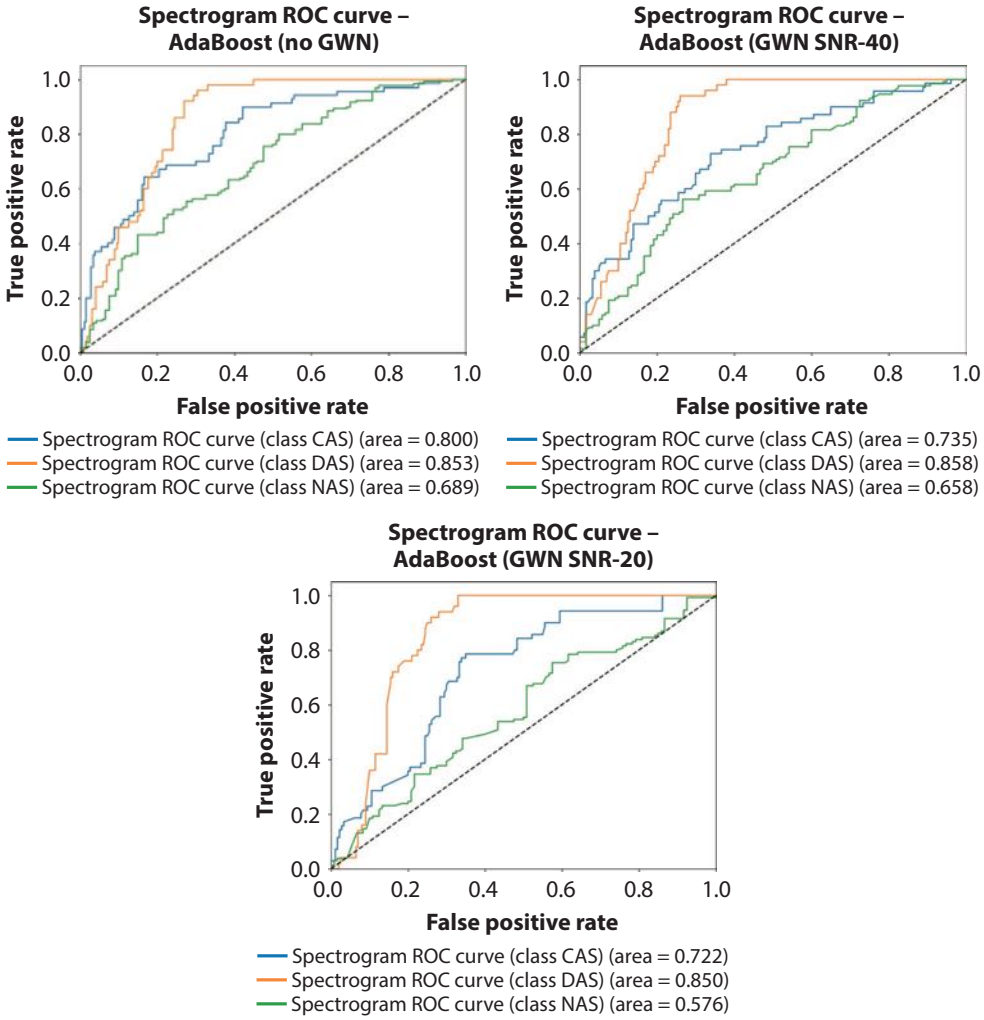


Fig. 5.2.3. *ROC curve showing significant impact (P = 0.000) of GWN on AdaBoost models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS- continuous auscultated sound, DAS – discontinuous auscultated sound, AdaBoost – Adaptive Boosting.

From Fig. 5.2.3, the impact of Gaussian white noise (GWN) levels on the actual positive rate (TPR) compared to the false positive rate (FPR) in the ROC curve can be observed for the spectrogram-based (Adaptive Boosting) AdaBoost model. At no GWN added level, the model shows strong performance, especially for CAS and DAS classes of sounds, with weaker performance for NAS class, as seen from the ROC area under the curve (AUC) scores. However, once the levels of GWN are increased to SNR-40, the lines for all three classes start to separate out as with CAS and NAS classes, TPR reducing significantly compared to FPR, but with DAS sounds maintaining relatively high ROC-AUC (area) score. Finally, once GWN at SNR-20 is added to all three classes of lung sounds, AdaBoost, a distinct separation between all three classes of lung sounds appears with DAS identification performance as the best, followed by DAS and then NAS and shown by ROC-AUC values.

From Fig. 5.2.4, the impact of Gaussian white noise is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the Categorical data Gradient Boosting (CatBoost) confusion matrix. At no GWN added levels, 47/70 CAS are identified correctly, 25/50 of DAS class is correctly identified, and 96 of 130 are correctly identified. The confusion matrix at SNR-40 shows a sharp decrease in the performance of spectrogram-based CatBoost ML model, with only 3/70 CAS correctly identified and a slightly more significant number of DAS correctly identified at 27/50 and 110/130 of NAS correctly identified. The drop in values at the GWN SNR-20 level for DAS with 0/50 was identified correctly, and CAS showed almost abysmal performance at 6/70. Meanwhile, NAS class sounds see even improvement, with 120/130 being correctly identified. Though the prior statement is true, we can see that the model has a tendency to label all three classes of lung sounds as NAS at GWN SNR-20, showing serious issues with the correct classification.

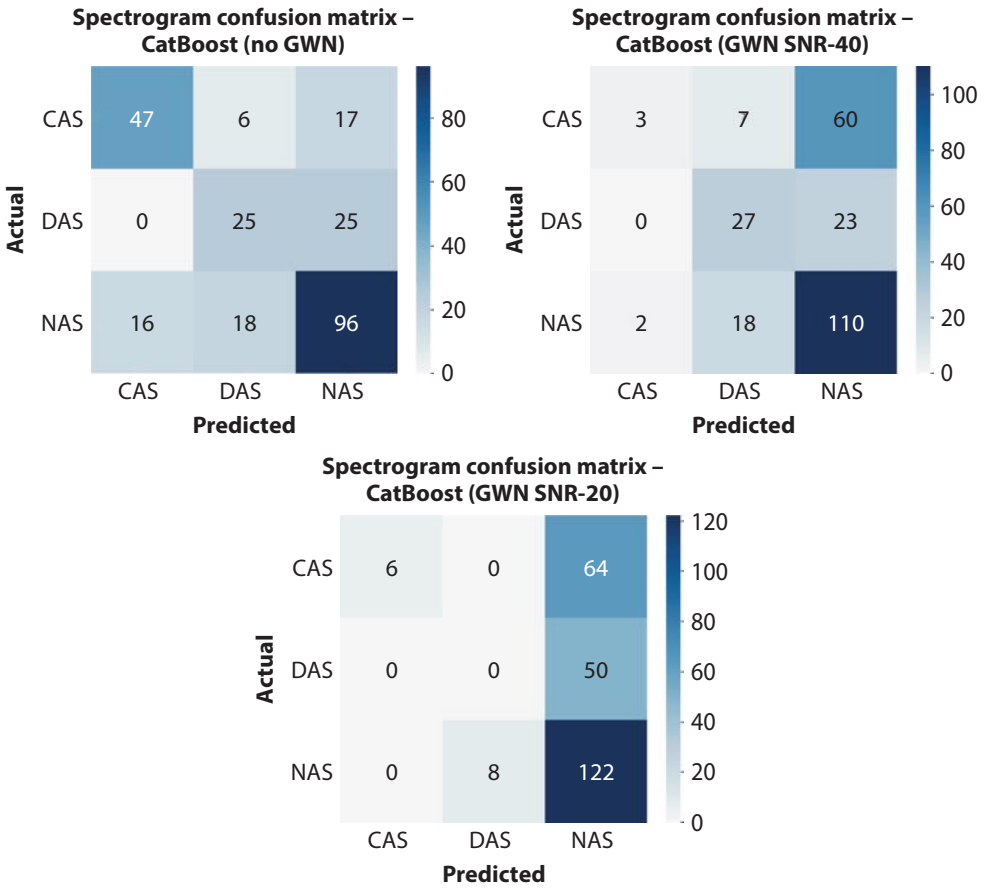


Fig. 5.2.4. *Confusion metrics showing significant impact (P = 0.000) of GWN on CatBoost models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, CatBoost – Categorical Boosting.

From Fig. 5.2.5, it can be observed that spectrogram-based Categorical Boosting (CatBoost) achieves relatively good levels of precision compared to recall for CAS, NAS class of sounds with slightly lower levels for DAS class as exemplified via precision to recall area under the curve (PR-AUC) values. At medium levels of GWN of SNR-40, the DAS and CAS lung sound identification are impacted by reduced PR-AUC, but the impact on NAS sounds is more limited. Finally, once GWN is increased to SNR-20 levels, all three classes of lung sounds are heavily and negatively impacted, with CAS and DAS class sound identification bearing the biggest brunt on the impact, precision compared to recall rates drop off into an abyss. The spectrogram-

based CatBoost model at no GWN added level shows relatively good performance, but with a caveat that this performance depends on the sound class, and shows limited robustness even to medium levels of GWN.

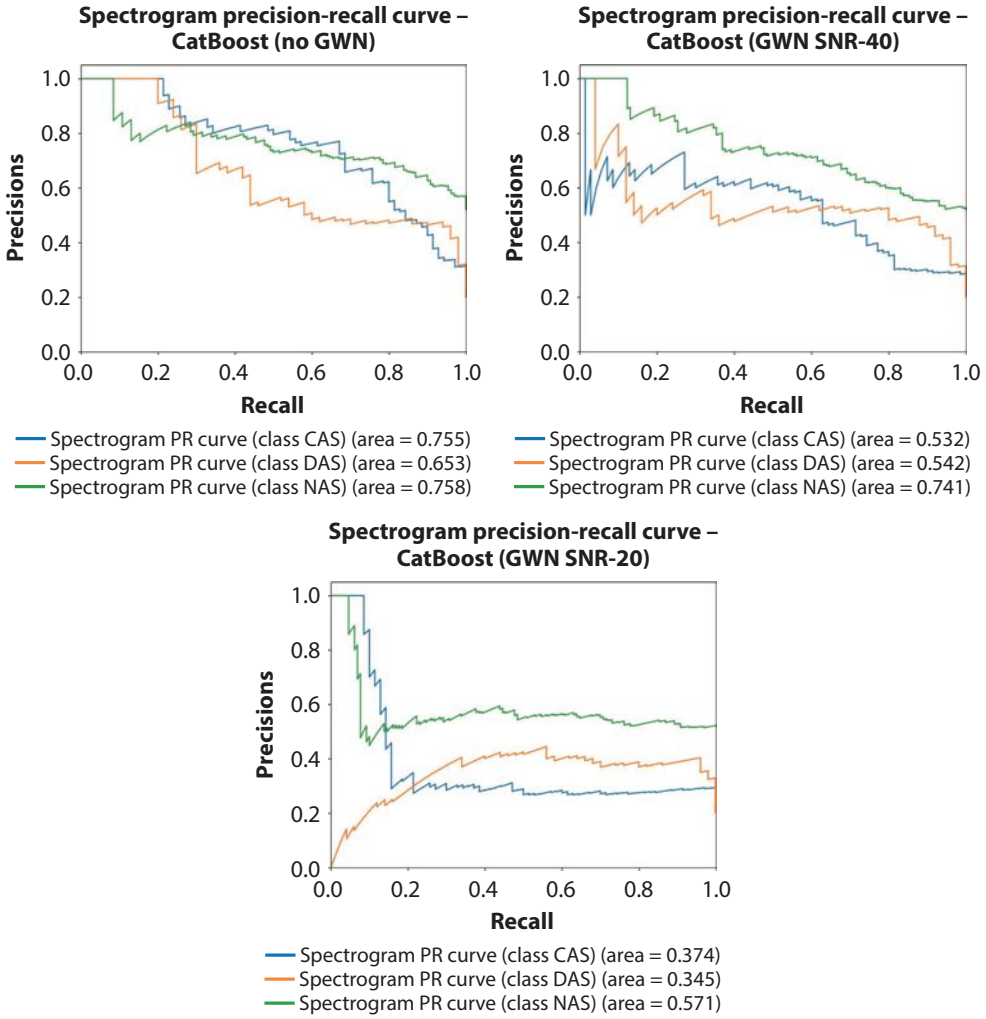


Fig. 5.2.5. *PR curve showing significant impact (P = 0.000) of GWN on CatBoost models' ability to identify lung sounds (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, CatBoost – Categorical Boosting.

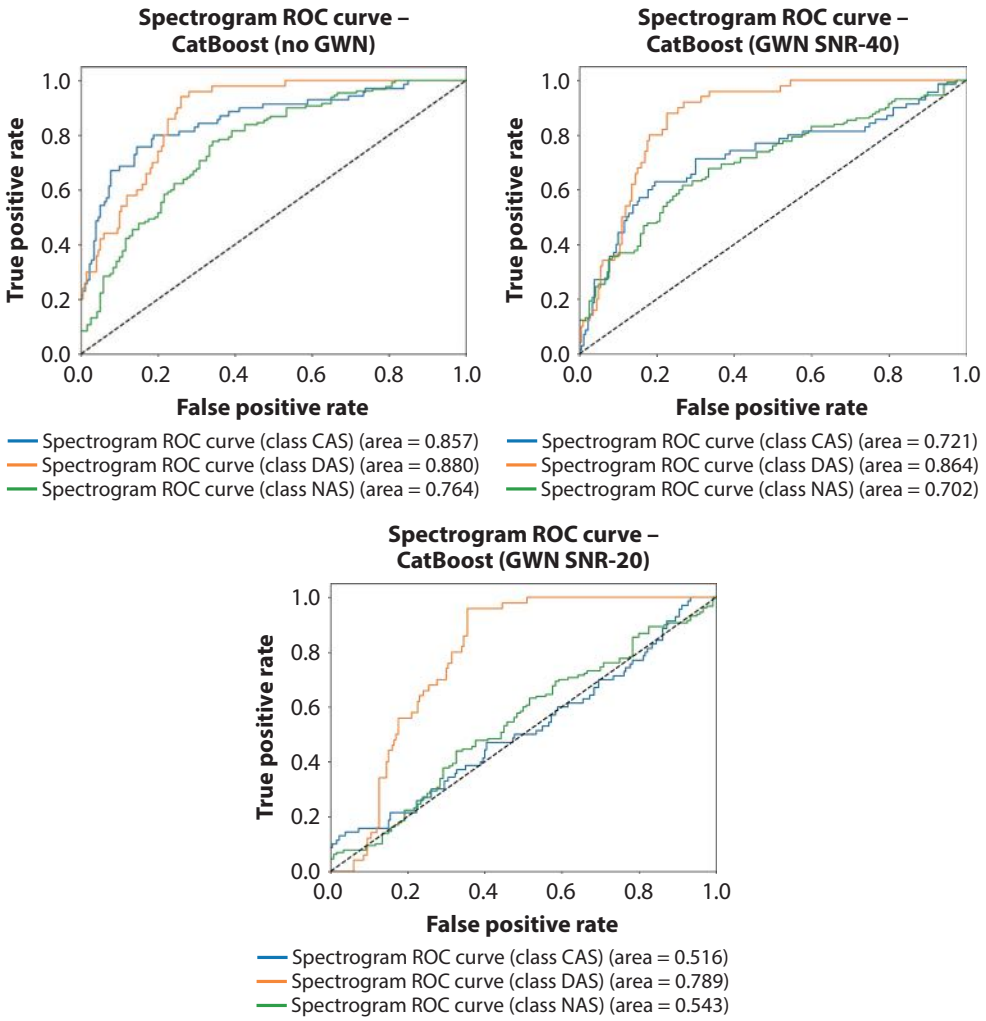


Fig. 5.2.6. *ROC curve showing significant impact (P = 0.000) of GWN on CatBoost models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, CatBoost – Categorical Boosting.

From Fig. 5.2.6, the impact of Gaussian white noise (GWN) levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) graph can be observed for the spectrogram-based Categorical Boosting (CatBoost) model. At no GWN added level, the model shows strong performance, especially for CAS and DAS classes of sounds, with slightly weaker performance for the NAS class, as seen from the ROC area under the curve (AUC) scores. However, once the

levels of GWN are increased to SNR-40, the lines for all three classes are impacted, with CAS and NAS class's TPR as compared to FPR being the lowest but with DAS sounds maintaining a relatively high ROC-AUC (area) score. Finally, once GWN at SNR-20 is added to all three classes of lung sounds, a distinct separation between all three classes of lung sounds appears. The CAS and NAS identification performance being the lowest, with the most excellent robustness shown by the DAS class, none of the less DAS and NAS classes are pretty much weaving around the random line (dashed line). Therefore, according to ROC-AUC values, the model is unfunctional as a diagnostic tool at the highest GWN levels.

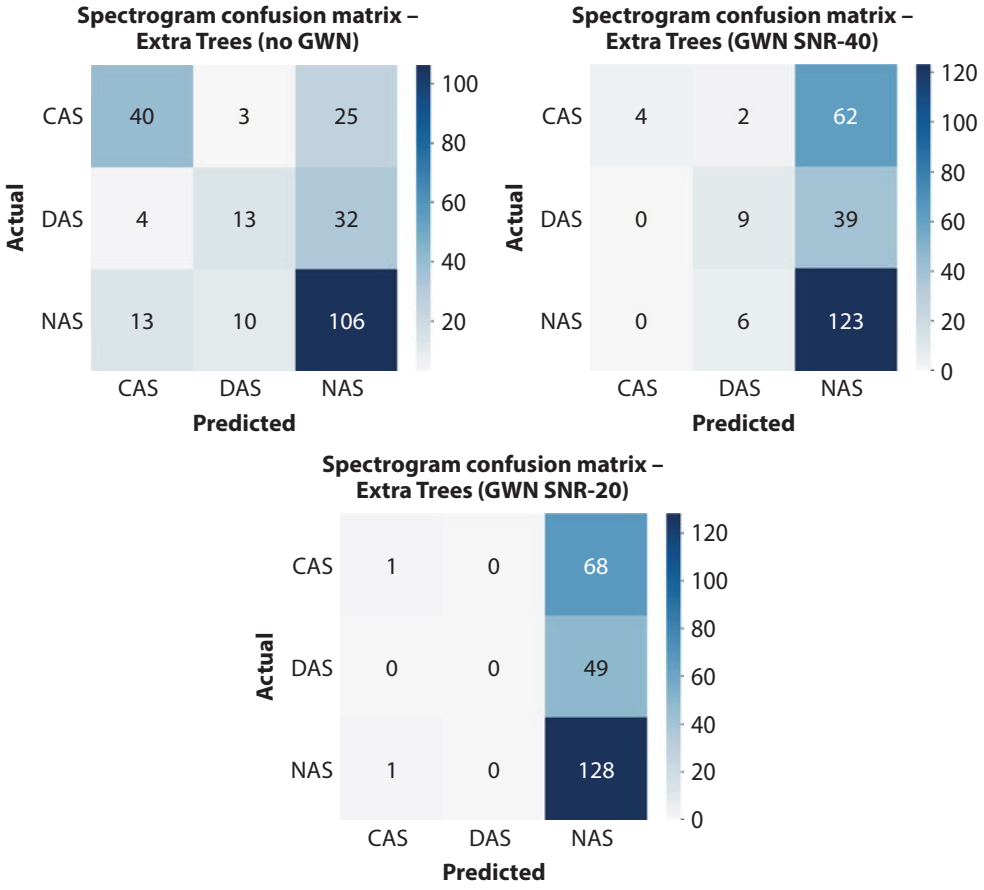


Fig. 5.2.7. *Confusion metrics showing significant impact (P = 0.000) of GWN on Extra Tree models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.7, the impact of Gaussian white noise is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the spectrogram-based Extra Trees' confusion matrix table. At no GWN added levels, 40/70 CAS was identified correctly, only 13/50 of the DAS class was correctly identified, and 103/130 of the NAS were correctly identified. The confusion matrix at SNR-40 shows a sharp decrease in the performance of the model performance with a sharp drop in TP for the CAS class, with only 4/70 correctly identified, the slightly more significant number of DAS correctly identified at 9/50, and 123/130 of NAS correctly identified. The values dropped at the GWN SNR-20 level, for DAS and CAS showed extremely poor performance, with 0/50 and 1/70 identified correctly, respectively. Meanwhile, NAS class sounds improved, with 128/130 correctly identified.

Though the prior statement is accurate, we can see that the model has a tendency to label all three classes of lung sounds as NAS at GWN SNR-20, showing serious issues with the correct classification. According to confusion matrix TP scores, the spectrogram-based extra tree model shows poor performance even at no GWN levels.

From Fig. 5.2.8, it can be observed that spectrogram-based Extra Trees struggles achieved overall sub-power levels of precision compared to recall. The NAS class performed followed by CAS and DAS as exemplified via precision to recall area under the curve (PR-AUC) values. At medium levels of GWN of SNR-40, all three classes were impacted, but the CAS class was the most significantly impacted, as seen with dropping precision compared to the recall curve and reduced PR-AUC values. Finally, once GWN is increased to SNR-20 levels, all three classes of lung sounds are heavily and negatively impacted, with DAS class sound identification bearing the biggest brunt of the impact, precision compared to recall rates drop off into an abyss for spectrogram-based Extra Trees ML model.

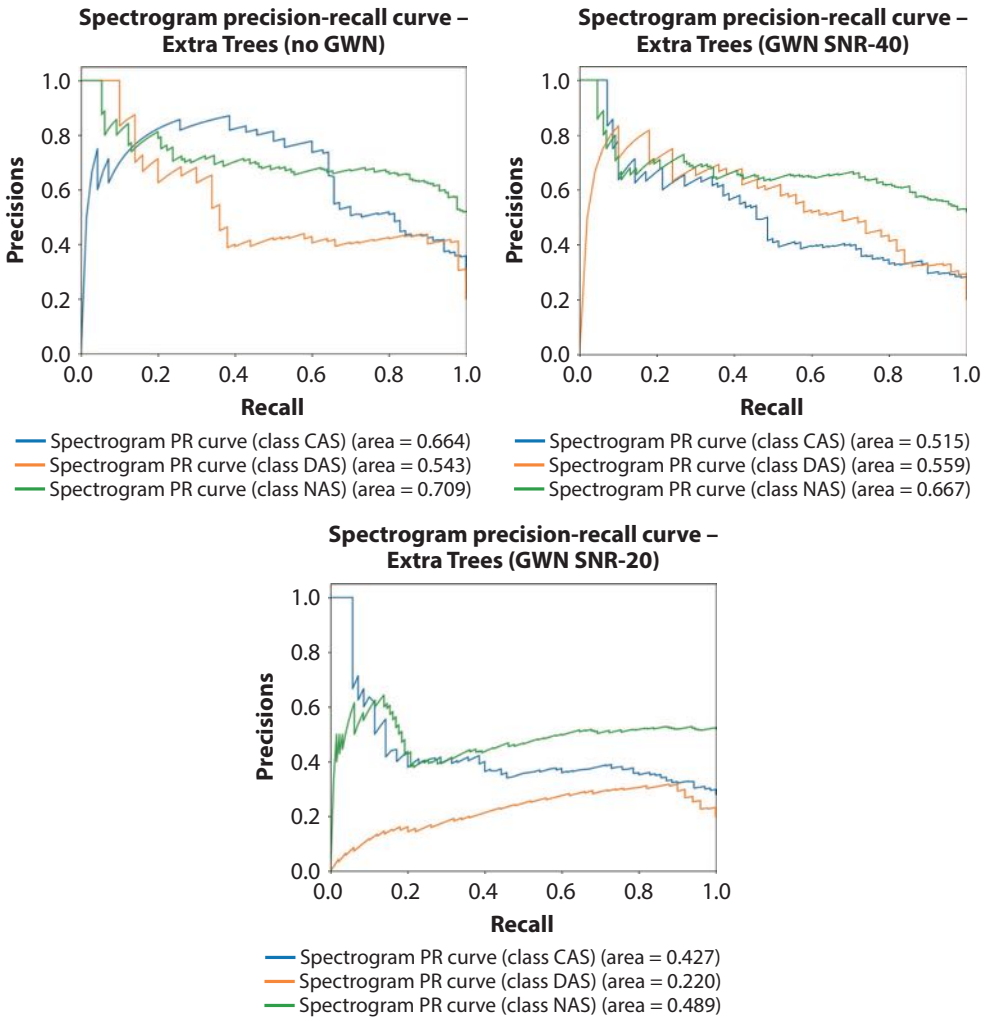


Fig. 5.2.8. *PR curve showing significant impact (P = 0.000) of GWN on Extra Tree models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

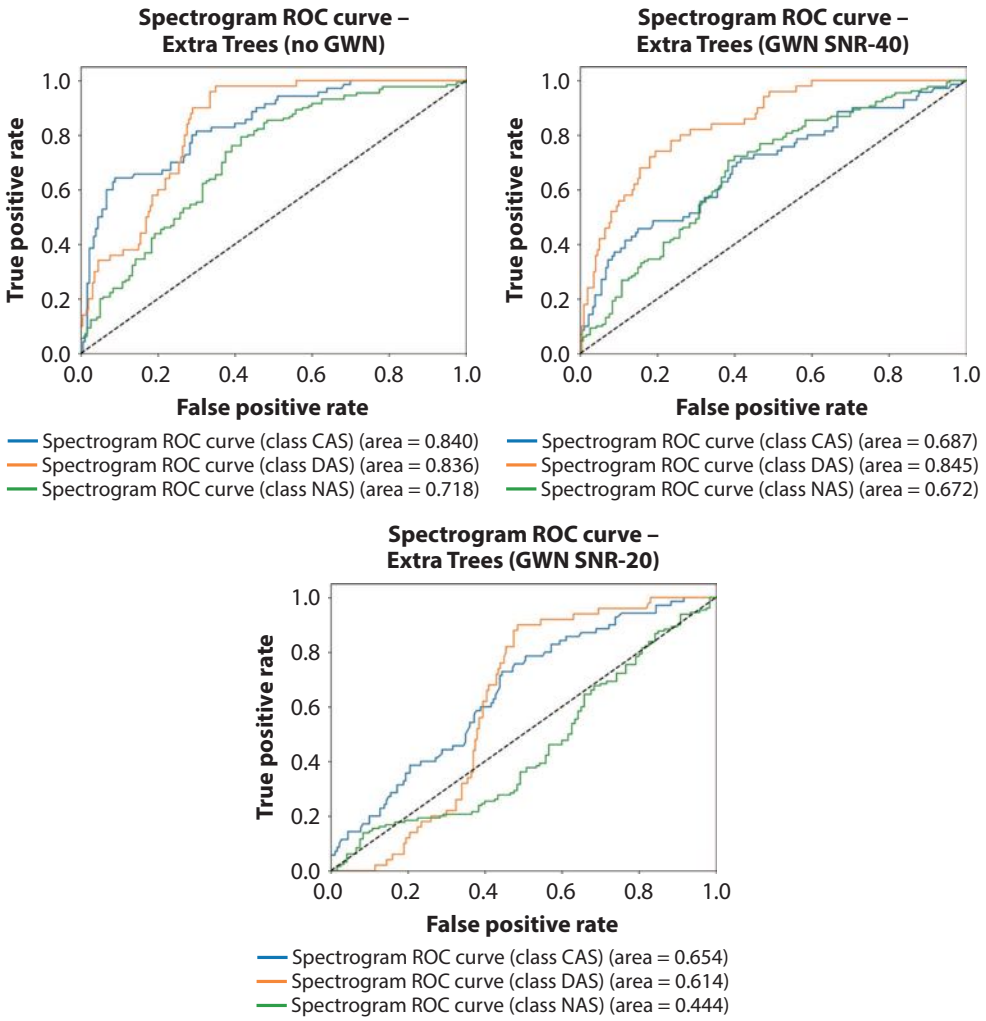


Fig. 5.2.9. *ROC curve showing significant impact (P = 0.000) of GWN on Extra Tree models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.9, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic graph can be observed for the spectrogram-based Extra Trees model. At no GWN added level, the model shows strong performance, especially for CAS and DAS classes of sounds, with slightly weaker performance for the NAS class as seen from ROC-AUC (area) scores. However, once the levels of GWN are increased to SNR-40, the TPR compared to FPR

for all three classes is negative. Especially for CAS and NAS classes, but with DAS sounds maintaining relatively as seen from the ROC area under the curve (ROC-AUC) scores. Finally, once GWN at SNR-20 is added to all three classes of lung sounds, a distinct separation between all three classes of lung sounds appears, with CAS and NAS identification performance being the lowest, with the greatest robustness shown by DAS class, none of the less DAS and NAS classes are pretty much weaving around the random line (dashed line). Therefore, according to ROC-AUC values, the model is unfunctional as a diagnostic tool at the highest GWN levels.

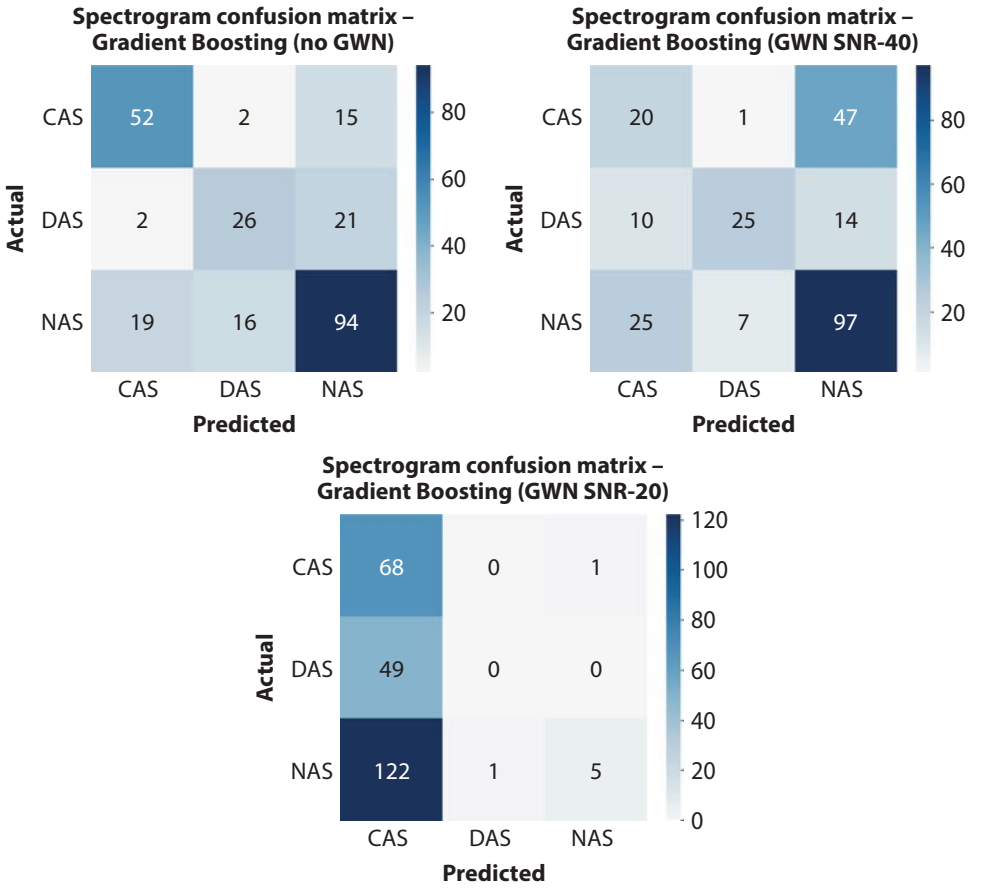


Fig. 5.2.10. *Confusion metrics showing significant impact (P = 0.000) of GWN on Gradient Boosting models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.10, the impact of Gaussian white noise (GWN) is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the spectrogram-based Gradient Boosting confusion matrix. At no GWN added levels, 52/70 CAS was identified correctly, only 26/50 of the DAS class was correctly identified, and 94/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium levels) shows a sharp decrease in performance of the model performance with a sharp drop in TP for the CAS class, with only 20/70 correctly identified, whilst maintaining performance for DAS and NAS with 25/50 97/130 scores, respectively. The TP scores dropped at the GWN SNR-20 level for NAS and DAS class identification, showing abysmal performance: 5/130 and 0/50 were identified correctly, respectively.

Meanwhile, CAS class sounds see even improvement, with 68/69 being correctly identified. Though the prior statement is accurate, we can see that the model tends to label all three classes of lung sounds as CAS at GWN SNR-20, showing serious issues with the correct classification. The Gradient Boosting ML model shows an overall reasonable performance according to confusion matrix scores, especially for CAS and NAS sounds, but with poor performance for the DAS class, especially with high levels of GWN.

From Fig. 5.2.11, it can be observed that spectrogram-based Gradient Boosting model achieves a reasonable precision performance compared to recall for CAS, NAS class of sounds with lower diagnostic levels for DAS classes exemplified via precision to recall area under the curve (PR-AUC) values. At medium levels of GWN of SNR-40, the NAS, DAS, and especially CAS lung sound identifications are impacted by reduced PR-AUC, but the impact on NAS sounds is more limited. Finally, once GWN is increased to SNR-20 levels, all three classes of lung sounds drop down as compared to no GWN-added levels, and this statement is especially true for CAS and DAS classes. Therefore, the spectrogram based gradient boosting models according to PR-AUC values reasonable performance for two classes, but with worse performing DAS class at no GWN-added levels and limited robustness to GWN even at GWN SNR-40 levels.

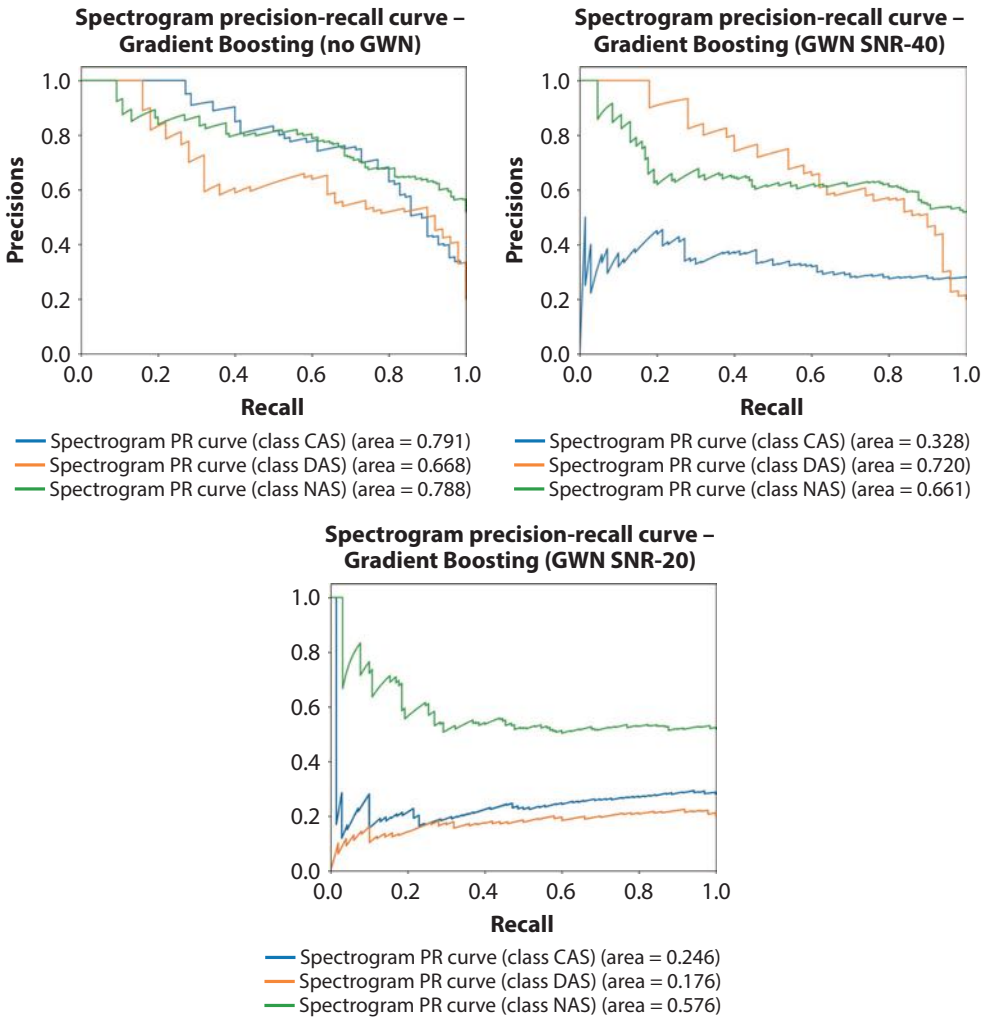


Fig. 5.2.11. *PR curve showing significant impact (P = 0.000) of GWN on gradient boosting models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

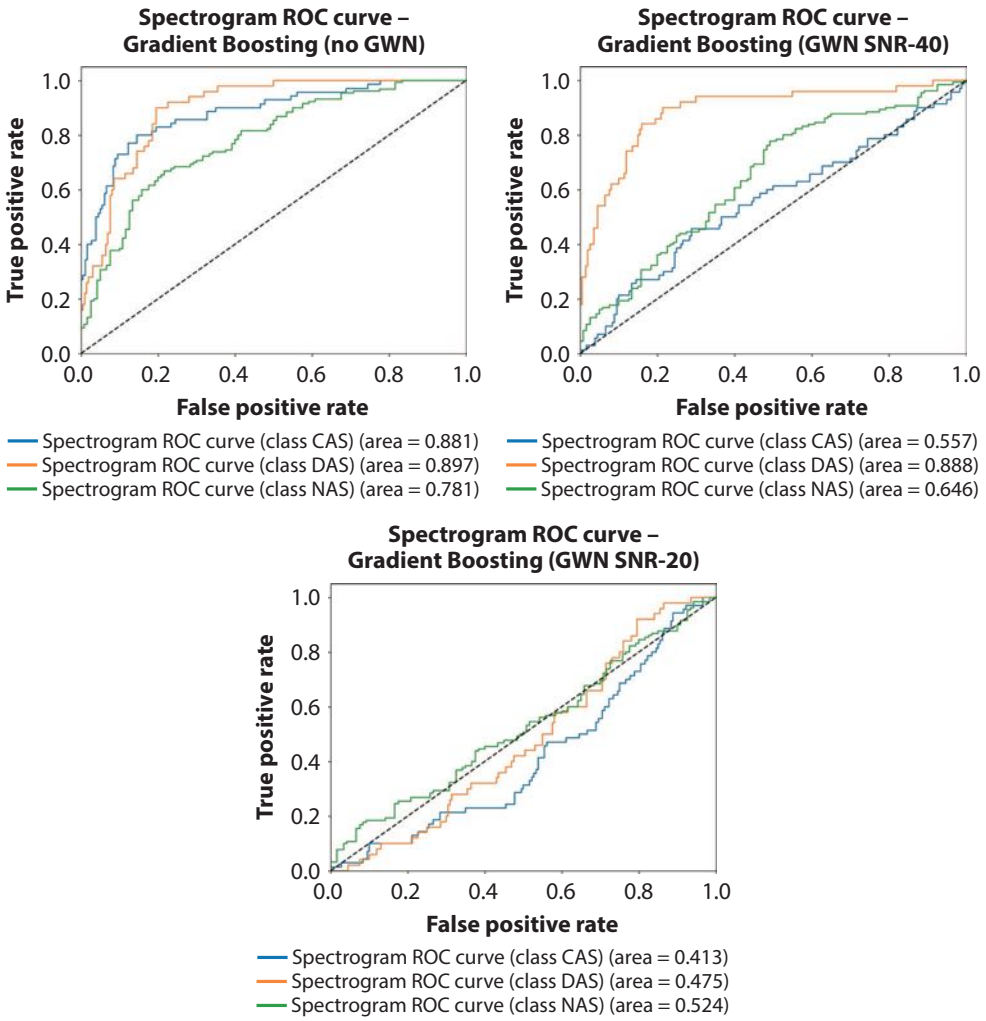


Fig. 5.2.12. *ROC curve showing significant impact (P = 0.000) of GWN on Gradient Boosting models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.12, the impact of Gaussian white noise (GWN) levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) graph can be observed for the spectrogram-based Gradient Boosting model. At no GWN added level, the model shows strong performance, especially for CAS and DAS classes of sounds, with slightly weaker performance for the NAS class as seen from the ROC area under the curve (ROC-AUC) scores. However, once the levels of

GWN are increased to SNR-40, the curves for all three classes are impacted negatively, with CAS and NAS classes TRP being the lowest, but with DAS sound class maintaining a relatively high ROC-AUC (area) score. Finally, once GWN at SNR-20 is added to all three classes the ML model performs poorly, especially for identification of CAS class sound. Therefore, the spectrogram-based Gradient Boosting model shows good performance at no GWN added levels, but lacks robustness with a great drop of performance at medium levels of GWN. The model's ability to identify true positives for CAS and NAS are extremely poor and this is especially true for all three sound classes at GWN SNR-20 levels.

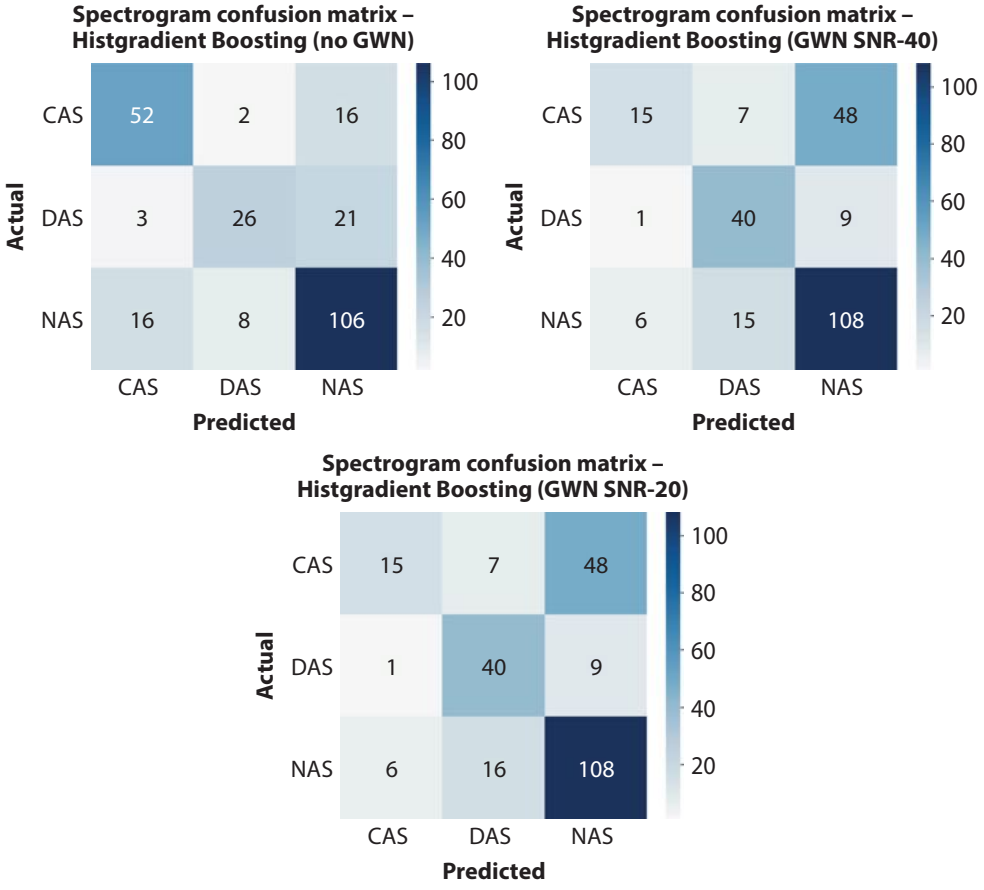


Fig. 5.2.13. *Confusion metrics showing significant impact (P = 0.000) of GWN on Histgradient models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.13, the impact of Gaussian white noise (GWN) is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the spectrogram-based Histgradient ML models' confusion matrix. At no GWN added levels, 52/70 CAS was identified correctly, only 26/50 of the DAS class was correctly identified, and 106/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium levels) shows a sharp decrease in performance with a drop in TP for the CAS class with only 15/70 correctly identified, whilst maintaining very strong performance for DAS and NAS with 40/50 and 108/130 correctly identified, respectively. This robustness continued with GWN level SNR-20, not impacting any of the three classes any further. The Histgradient model shows overall good performance according to TP scores in the confusion matrix table, especially for DAS and NAS sounds. Though, stable but poorer performance for CAS class with GWN at SNR-40 and SNR-20 levels. Additionally, the model showed reasonably good resilience to GWN.

Fig. 5.2.14 shows that spectrogram-based Histgradient ML model achieves good precision performance compared to recall for CAS, NAS class of sounds with lower diagnostic levels for DAS class as exemplified via precision to recall area under the curve (PR-AUC) values. At medium levels of GWN of SNR-40, the NAS, CAS, and out of the two, especially CAS lung sound identification, are impacted, as shown by reduced PR-AUC. However, the impact on DAS sounds shows robustness. Finally, once GWN is increased to SNR-20 levels, all three classes spread out, but the curves drop lower, indicating that the ambient noise impacts all classes of lung sounds. However, CAS is more negatively affected than DAS and NAS. The spectrogram-based Histgradient model shows good precision to recall scores with strong PR-AUC values at no GWN-added levels. Additionally, robustness of the model is resilient up to and including GWN SNR-40 levels with especially good performance for DAS class.

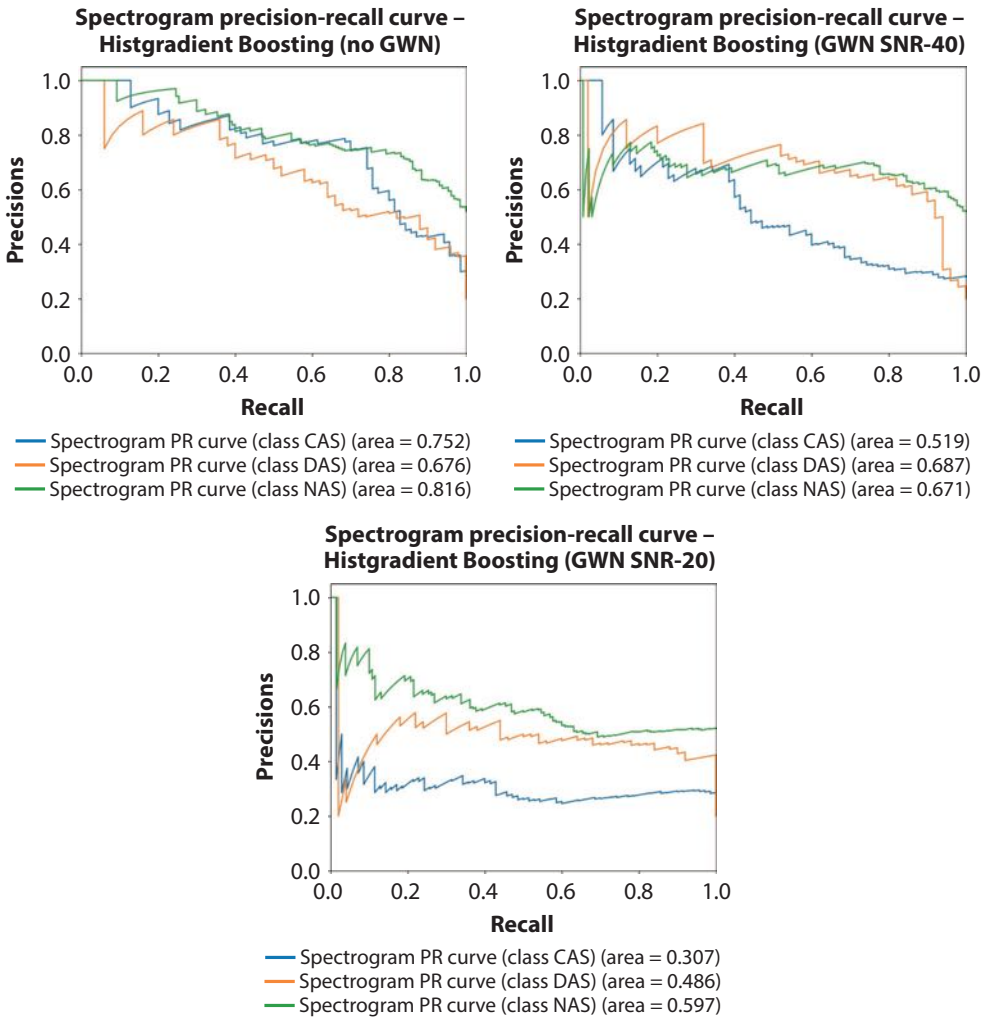


Fig. 5.2.14. *PR curve showing significant impact (P=0.000) of GWN on Histgradient models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

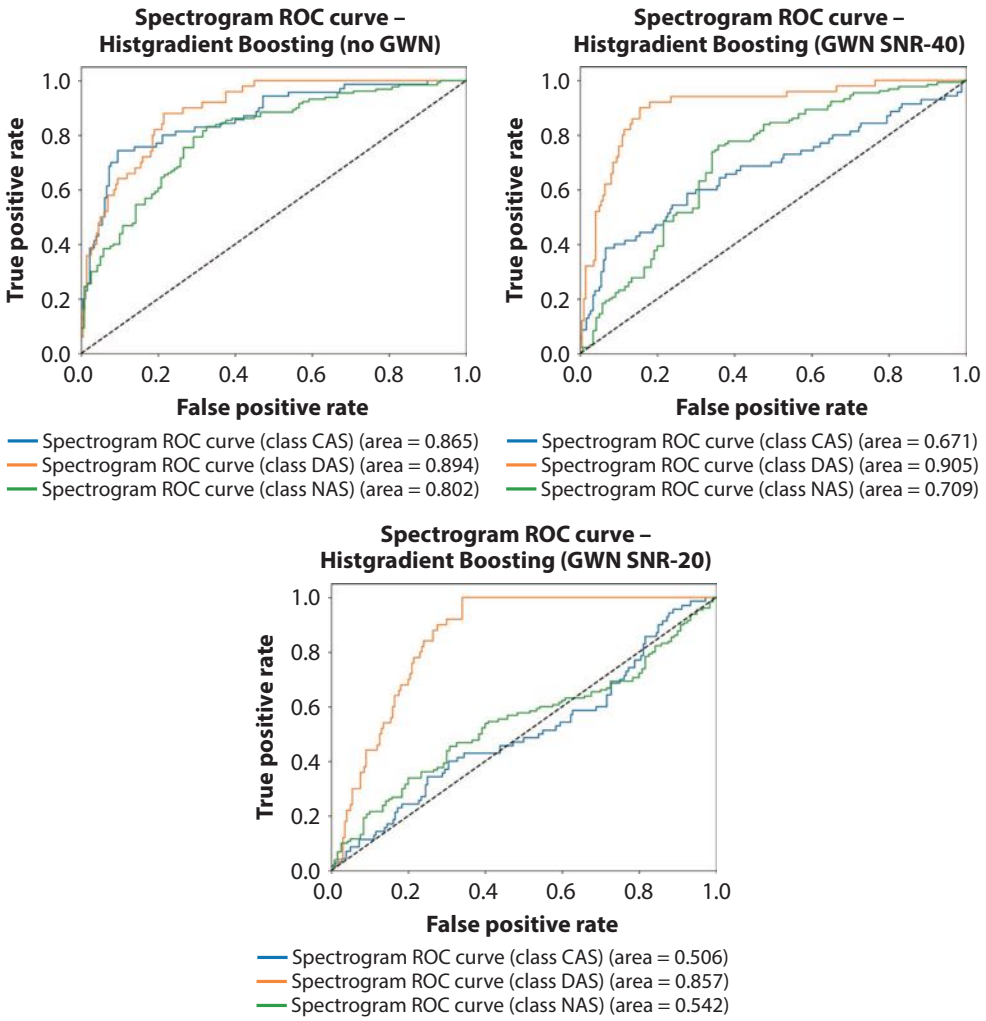


Fig. 5.2.15. *ROC curve showing significant impact (P=0.000) of GWN on Histgradient models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

Fig. 5.2.15 shows the impact of Gaussian white noise (GWN) levels on the true positive rate compared to the false positive rate in the receiver operating characteristic (ROC) curve for the spectrogram-based Histgradient ML model. At no GWN added level, the model showed a powerful performance, especially for DAS and CAS sounds, with slightly weaker (but still strong) performance for the NAS class, as seen from ROC area under the curve (ROC-AUC) scores. However, once the levels of GWN are increased to SNR-

40, the curves for two classes are impacted by the ambient noise: The CAS and NAS class's TPR scores drop. However, the rates are still quite good compared to other models. Whilst DAS sounds maintain a high ROC-AUC score. Finally, once GWN at SNR-20 is added to CAS and NAS classes, performance continues to worsen whilst DAS continues to show robustness. Nonetheless, at the highest levels of GWN due to the inability of the model to correctly predict two out of three classes, the model becomes ineffective. The diagnostic effectiveness of the Histgradient model is valid only at no GWN added and GWN SNR-40 levels.

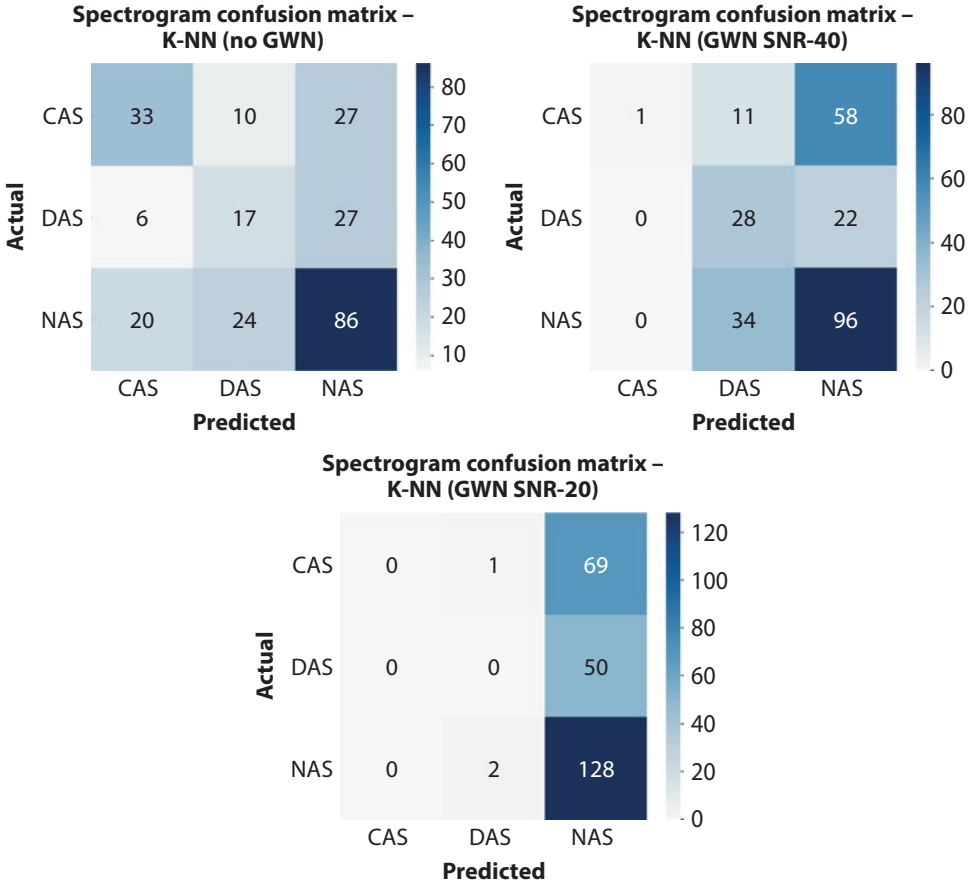


Fig. 5.2.16. *Confusion metrics showing significant impact (P = 0.000) of GWN on K-NN models' ability to identify lung sounds correctly (from top to bottom)*

Note extreme poor performance of this model. GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, K-NN – K-Nearest Neighbors.

Fig. 5.2.16 clearly shows the impact of Gaussian white noise on true positive (TP), true negative (TN), false positive (FP), and false negative (FP), as seen in the spectrogram-based K-Nearest Neighbors (K-NN) models' confusion matrix. At no GWN added levels, 33/70 CAS was identified correctly, with only 17/50 of DAS class correctly identified and 86/130 of NAS correctly identified. The confusion matrix at SNR-40 (medium levels) shows a sharp decrease in performance of the model performance with a sharp drop in TP for the CAS class with only 1/70 correctly identified, whilst maintaining very reasonable performance for DAS with 28/50 correctly and NAS with 96/130 correctly recognised. The increase in the GWN levels of SNR-20 saw a continued significant drop in the models' performance for DAS and CAS class sound identification, with both scoring 0. Only the NAS class maintaining a good TP score of 128/130. The spectrogram-based K-NN model overall showed very poor performance according to confusion matrix across for DAS and CAS sounds, poor robustness to GWN levels and bias towards mislabelling sound as NAS at higher GWN levels.

Fig. 5.2.17 shows spectrogram-based K-Nearest Neighbors (K-NN) ML model achieves poor precision performance compared to recall for CAS, NAS class of sounds with lower diagnostic levels for DAS class as precision to recall area under the curve (PR-AUC) values. At medium levels of GWN of SNR-40, the DAS, CAS, and out of the two, the diagnostic performance is worse for DAS lung sound class identification, which is impacted with reduced PR-AUC, but the impact of GWN of SNR-40 on NAS class is minimal. Finally, once GWN is increased to SNR-20 levels, we see an extreme impact on the DAS class with limited impact on the CAS and NAS sound classes. According to PR graph the spectrogram-based K-NN model was useless for predicting three classes under different levels of GWN.

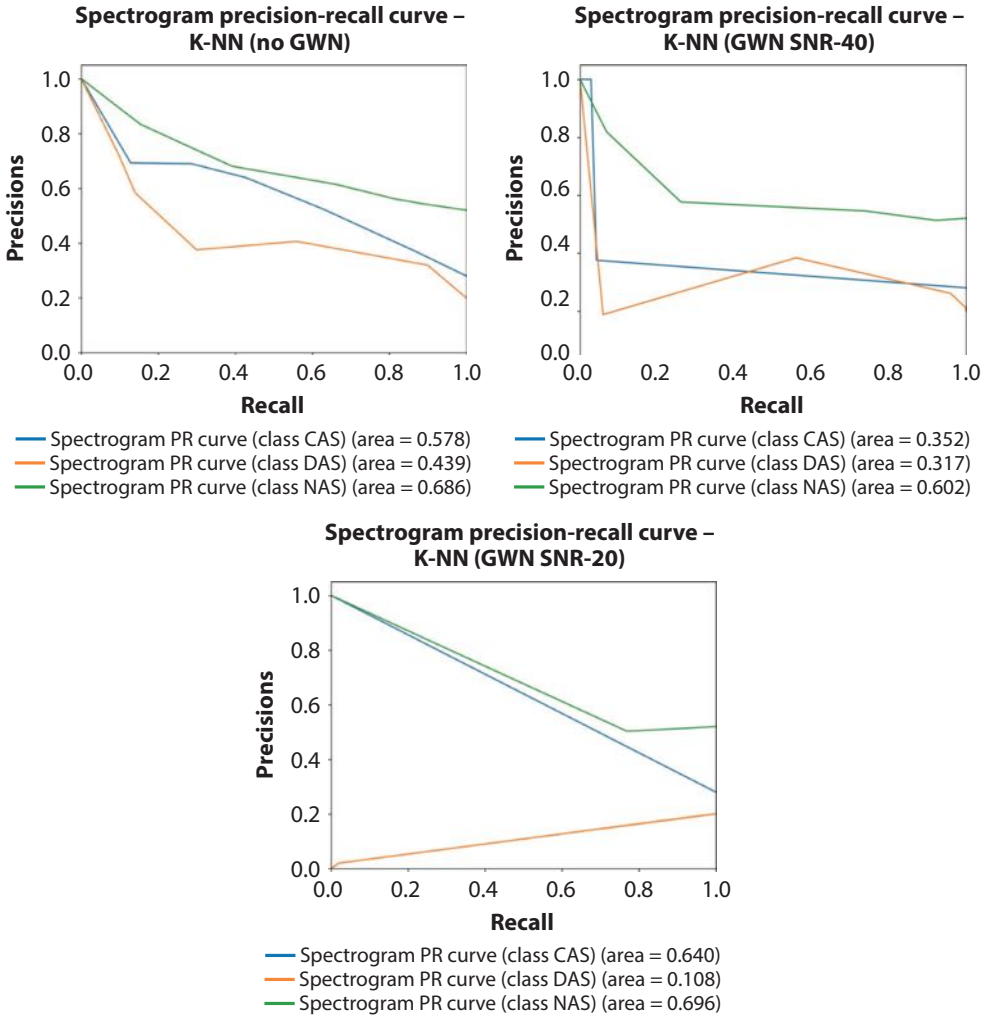


Fig. 5.2.17. *PR curve showing significant impact (P = 0.000) of GWN on K-NN models' ability to identify lung sounds correctly (from top to bottom)*

Note extreme performance of the model. GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, K-NN – K-Nearest Neighbors.

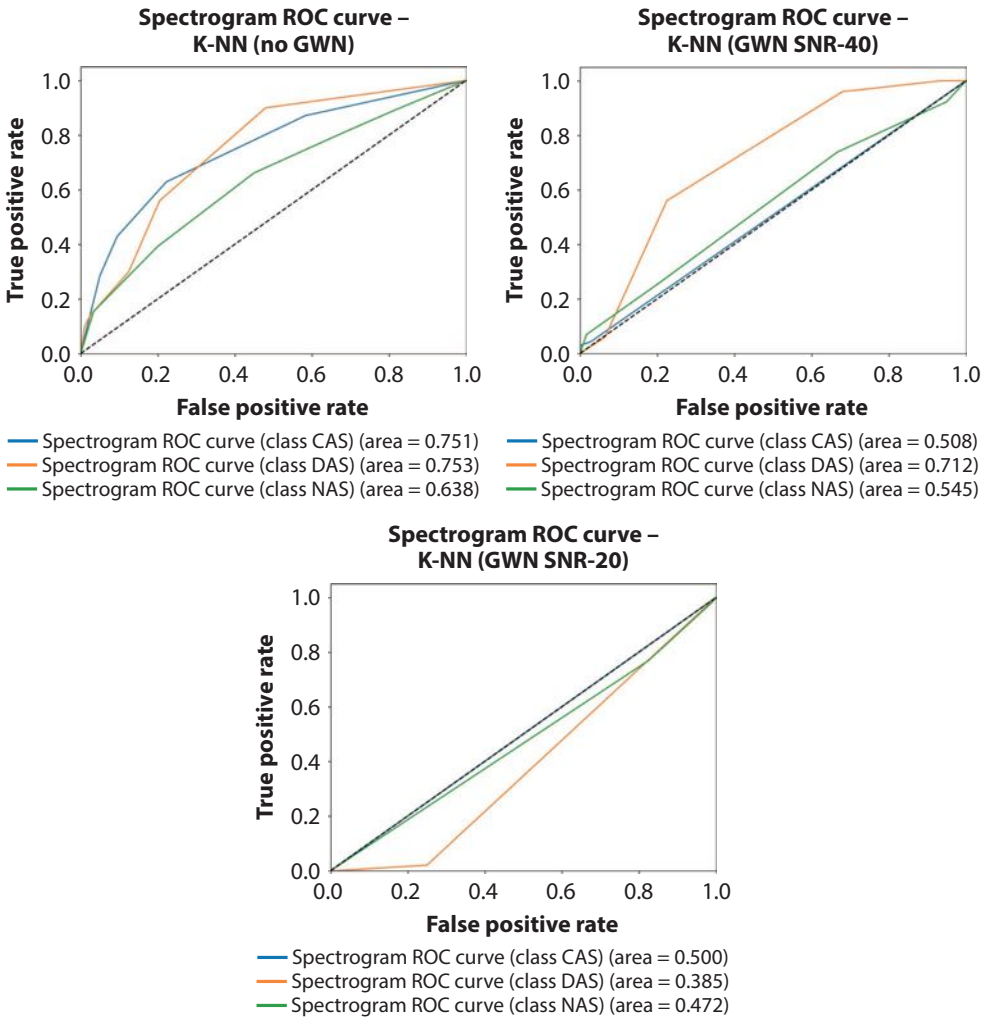


Fig. 5.2.18. *ROC curve showing significant impact (P = 0.000) of GWN on K-NN boosting models' ability to identify lung sounds correctly (from top to bottom).*

Note extreme performance of the model. GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, K-NN – K-Nearest Neighbors.

From Fig. 5.2.18, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based K-Nearest Neighbors (K-NN) model. At no Gaussian white noise (GWN) added level, the model shows a reasonable performance, especially for CAS and DAS classes of sounds, with weaker (but still firm) performance for the NAS

class, as seen from the ROC area under the curve (ROC-AUC) scores. However, once the levels of GWN are increased to SNR-40, the lines for all two classes of TPR are compared to the FPR drops for the CAS and NAS classes. As seen from ROC-AUC (area) scores, DAS sounds maintain a higher level. Finally, once GWN at SNR-20 is added, all three classes' true positive rates drop significantly, and the machine model becomes completely useless. Therefore, the spectrogram-based K-NN model only functions at no GWN added levels and shows a lack of robustness even to medium levels of GWN.

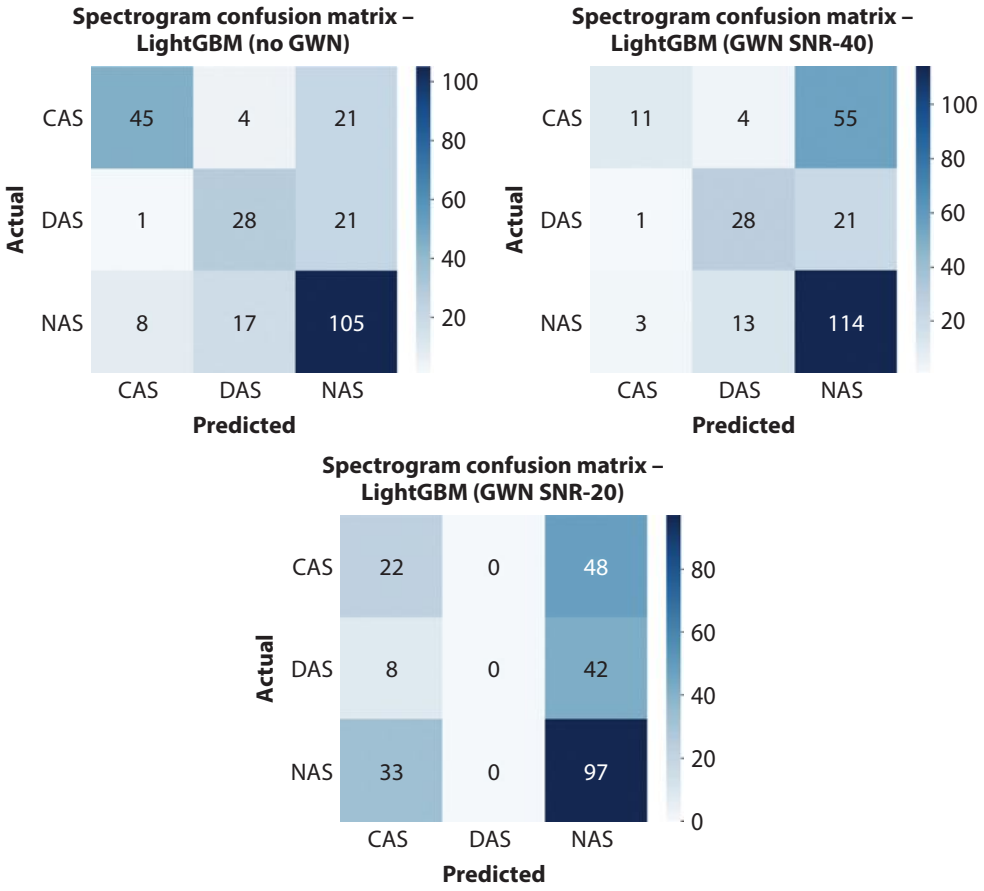


Fig. 5.2.19. *Confusion metrics showing significant impact (P = 0.000) of GWN on LightGBM models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, LightGBM – Light Gradient Boosting Machine.

From Fig. 5.2.19, the impact of Gaussian white noise is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the spectrogram-based Light Gradient Boosting Machine (LightGBM) model's confusion matrix. At no Gaussian white noise (GWN) added levels, TP rates were above average for CAS, DAS and NAS with scores of 45/70, 28/50 and 105/130, respectively. The confusion matrix at SNR-40 (medium levels of GWN) shows a significantly negative impact only on the model's ability to identify CAS class with a score of 11/70. However, once the GWN increased to SNR-20 level, it significantly worsened the model's performance in discriminating between the three classes, with DAS being impacted the most. The spectrogram-based LightGBM ML model lost its power to discriminate between classes with only 22/70 and 0/50 for CAS and DAS, respectively. The score for NAS was 97/130. The spectrogram-based model showed a reasonably good overall performance for all three classes at no GWN-added levels. However, the model lacked robustness and showed great TP score variability depending on GWN levels, as the score for CAS was at medium levels, whilst performance drastically worsened for recognition of the model at SNR-20 levels. The model's discrimination power at the highest levels of GWN got even further impacted, and sounds were classed mainly as NAS or CAS.

From Fig. 5.2.20, it can be observed that the spectrogram-based Light Gradient Boosting Machine (LightGBM) achieves good performance of precision to recall for CAS, NAS class of sounds with lower diagnostic levels for DAS class as exemplified by precision to recall area under the curve (PR-AUC). At medium levels of Gaussian white noise (GWN) of SNR-40, the CAS, NAS and out of the two, the diagnostic performance was worse for CAS class identification is impacted with reduced PR-AUC, but the DAS shows robustness. Finally, once GWN is increased to SNR-20 levels, we see an extreme impact on the DAS class with a lesser impact on CAS and NAS than no GWN levels. Therefore, the spectrogram-based LightGBM ML model shows good performance with intermediate robustness to GWM.

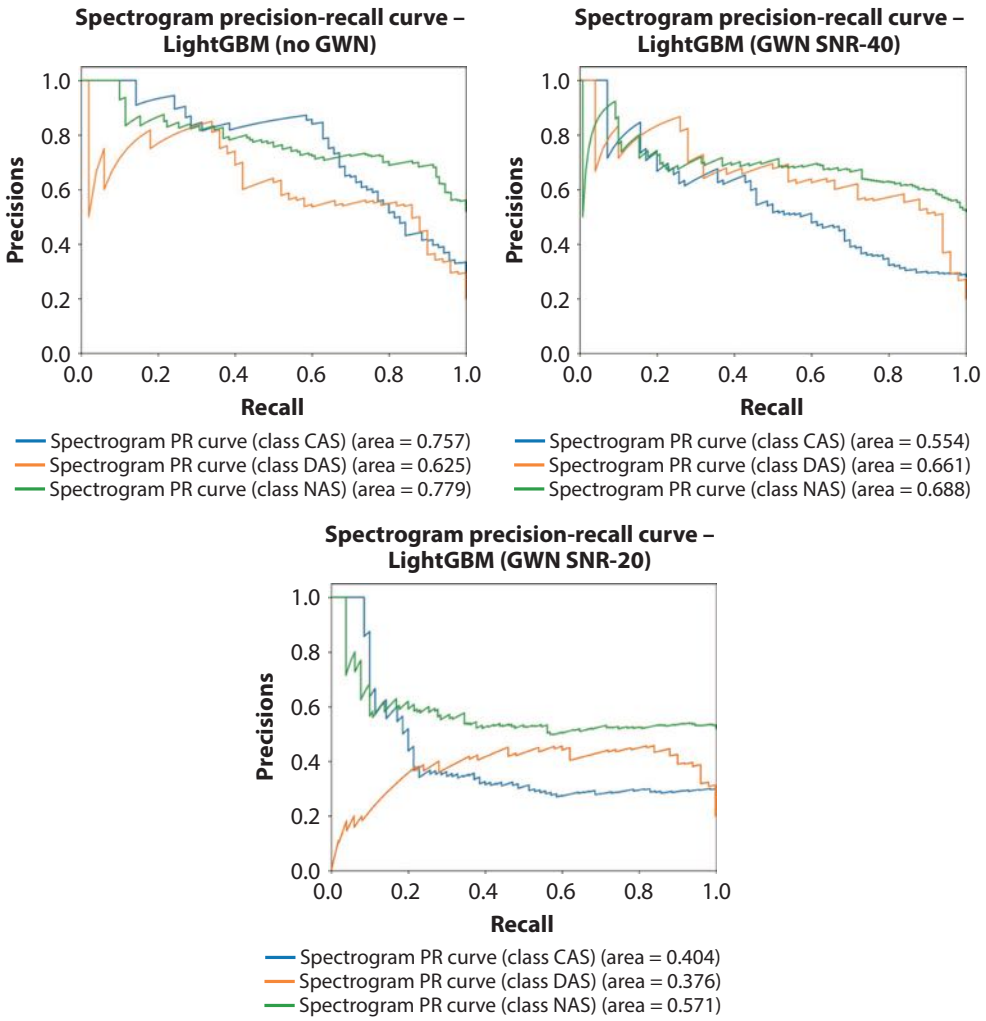


Fig. 5.2.20. *PR curve showing significant impact (P = 0.000) of GWN on LightGMB models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, LightGBM – Light Gradient Boosting Machine.

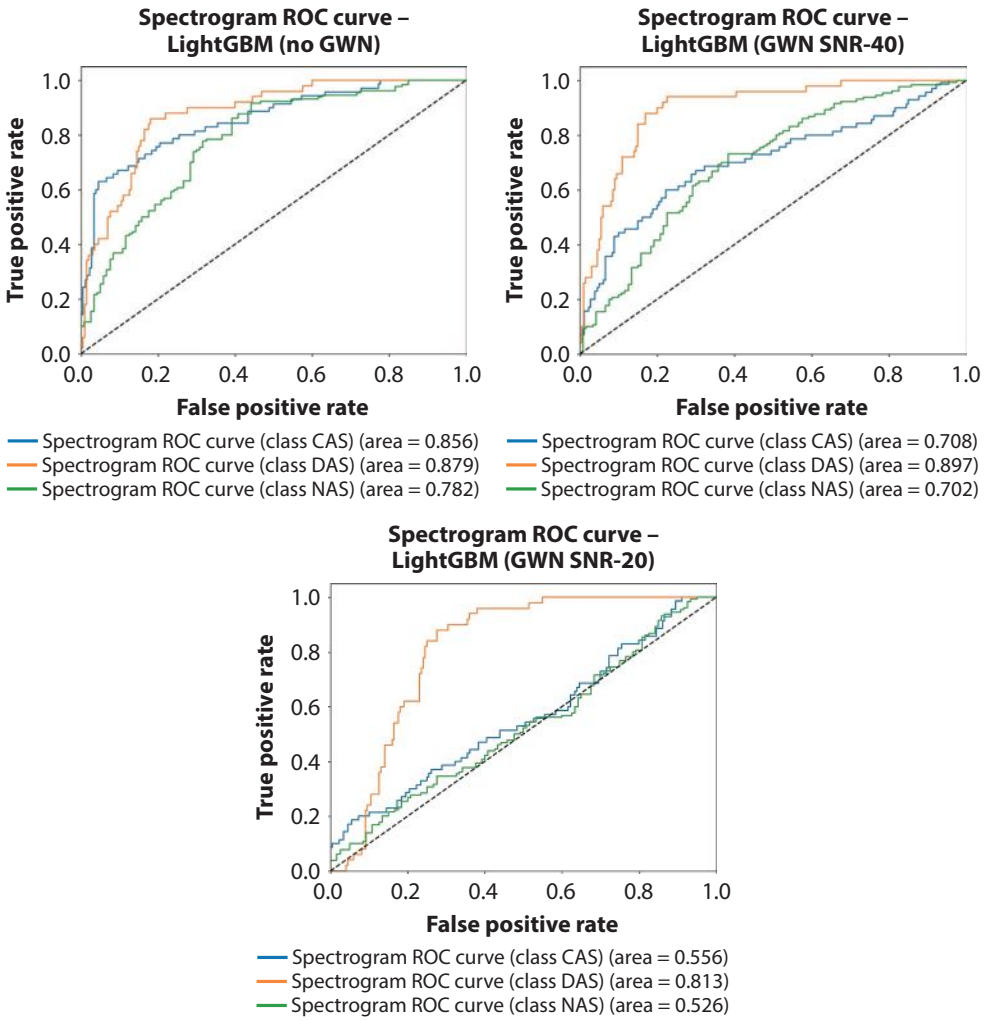


Fig. 5.2.21. *ROC curve showing significant impact (P = 0.000) of GWN on LightGBM models' ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, LightGBM – Light Gradient Boosting Machine.

From Fig. 5.2.21, the impact of Gaussian white noise (GWN) levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based Light Gradient Boosting Machine (LightGMB model). At no GWN-added level, the model shows a good performance, especially for the DAS class, with weaker (but still reasonable) performance for NAS and

CAS classes, as seen from ROC-AUC (area) scores. The model shockingly shows great resistance to medium levels of GWN at SNR-40. The two groups that are negatively affected are CAS and NAS. Finally, once GWN at SNR-20 is added, two groups, NAS and CAS TPR drop significantly, and the machine model becomes useless. The DAS is the only class that maintains strong true positive rates. Therefore, the spectrogram-based LightGBM ML model only functions at no GWN and GWN SNR-40 and shows robustness to medium levels of GWN but not high levels of GWN.

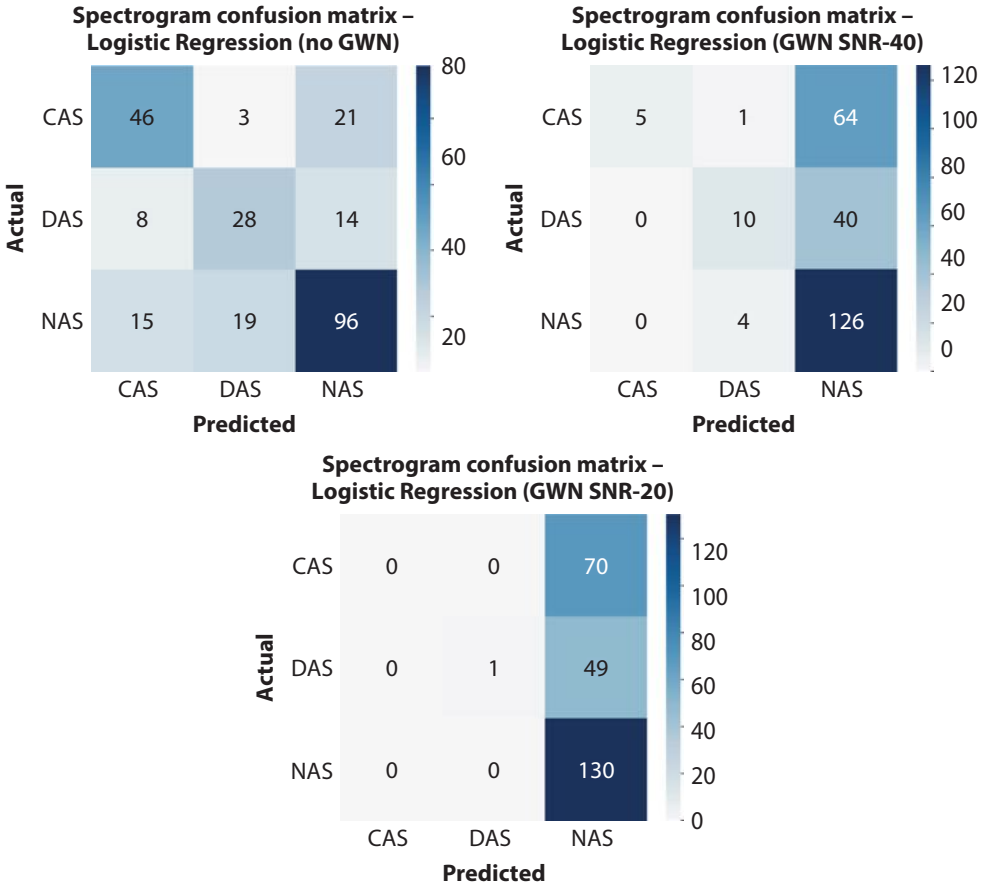


Fig. 5.2.22. *Confusion metrics showing significant impact (P = 0.000) of GWN on Logistic Regression model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.22, the impact of Gaussian white noise is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the spectrogram-based Logistic Regression model's confusion matrix. At no Gaussian white noise (GWN)-added levels, TP rates were above average for CAS, DAS and NAS with scores of 46/70, 28/50 and 96/130, respectively. The confusion matrix at SNR-40 (medium levels of GWN) significantly impacted the model, with scores for CAS and DAS at 5/70 and 10/50, respectively, with only 126/130 scores for NAS. However, once the GWN increased to SNR-20, it significantly worsened the model's performance in discriminating between the three classes, leading to the machine model losing its power to discriminate between classes with only 0/50 and 1/70 for CAS and DAS, respectively. The score for NAS was 130/130. The spectrogram-based Logistic Regression model showed a reasonably good overall performance for all three classes at no GWN-added levels. However, the model lacked any robustness as the score for CAS and DAS worsened significantly at GWN SNR-40 levels. The model's discrimination power at the highest levels of GWN got even further impacted, and sounds were classed mainly as NAS.

From Fig. 5.2.23, it can be observed that spectrogram-based Logistic Regression ML model achieves a reasonable performance of precision compared to recall for CAS and NAS sound classes but with significantly lower diagnostic levels for the DAS class, as exemplified by precision to recall area under the curve (PR-AUC). At medium levels of GWN of SNR-40, all three classes are impacted, with the CAS class being impacted the most, followed by the NAS class, but the DAS class shows robustness, as shown by PR-AUC values. Finally, once GWN is increased to SNR-20 levels, we see an extreme impact on the DAS class recognition with a lesser impact on CAS and NAS classes, as compared to no GWN levels.

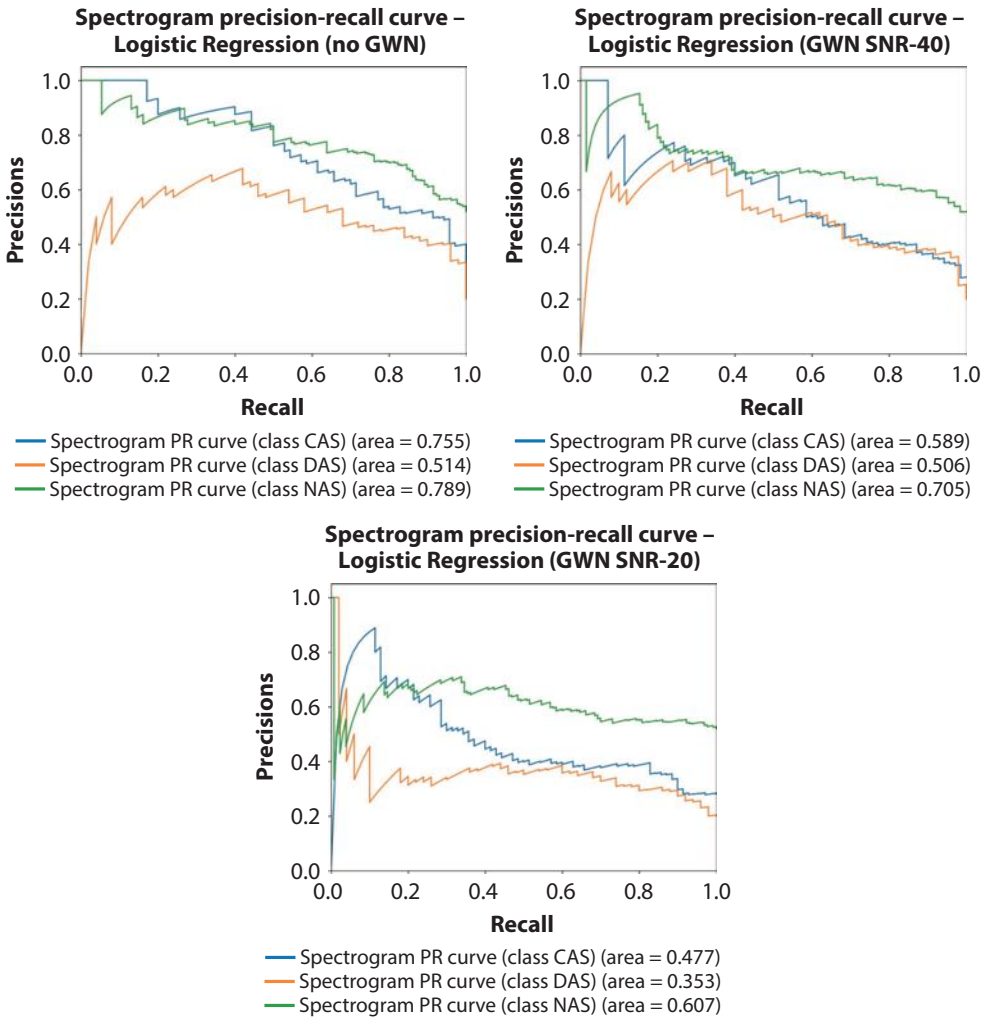


Fig. 5.2.23. *PR curve showing significant impact (P = 0.000) of GWN on Logistic Regression model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

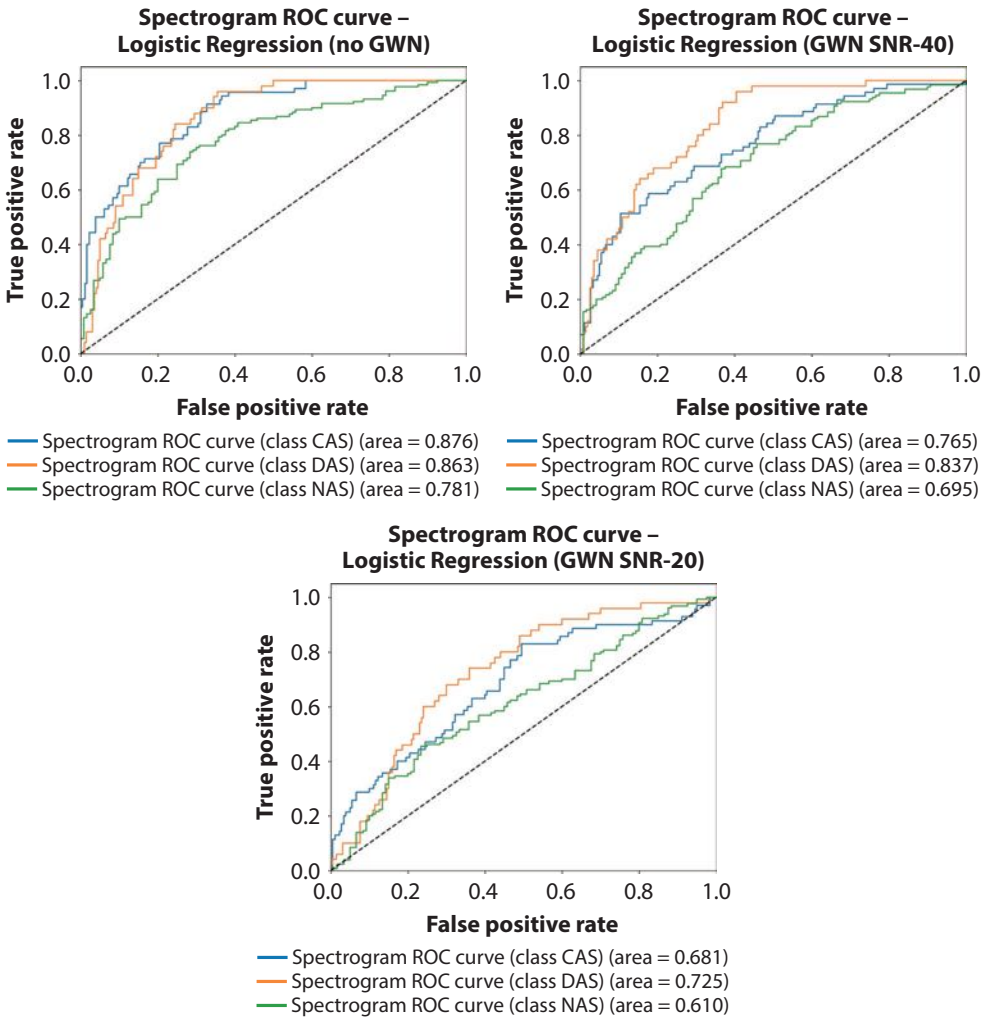


Fig. 5.2.24. *ROC curve showing significant impact (P = 0.000) of GWN on Logistic Regression model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.24, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based Logistic Regression model. At no GWN added level, the model shows a very good performance, especially for CAS and DAS classes, with weaker (but still good) performance for the NAS class recognition, as seen from the ROC area under the curve (AUC) scores. The model performance drops is visible at

medium levels of GWN at SNR-40 for two sound groups: CAS and special NAS. It maintains good robustness for the DAS class. Finally, once GWN at SNR-20 is added, all three classes are impacted, with DAS maintaining the best ROC-AUC scores. The class's TPR drops significantly for all three classes, but nonetheless, the performance is of a reasonable standard, and all three lines for all three classes maintain similar curvature and ROC-AUC values above 0.600, showing strong performance of the Logistic Regression ML model with ambient noise and strong robustness to even the highest GWN levels.

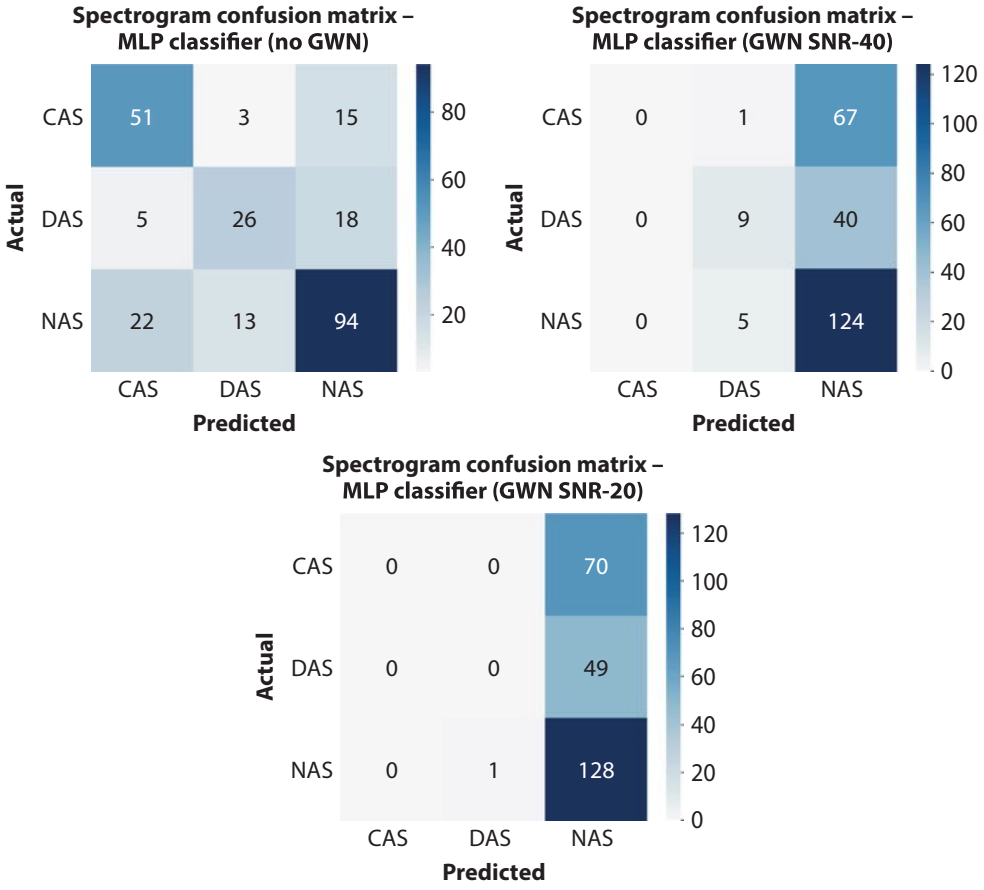


Fig. 5.2.25. *Confusion metrics showing significant impact (P = 0.000) of GWN on MLP model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, MLP – Multilayer Perceptron.

From Fig. 5.2.25, the impact of Gaussian white noise (GWN) is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in spectrogram-based Multilayer Perceptron (MLP) ML model's confusion matrix. At no GWN-added levels, TP values were above average for CAS, DAS and NAS, with scores of 51/70, 26/50 and 94/130, respectively. The confusion matrix at SNR-40 (medium levels) showed a significant impact of ambient noise on the model with scores for CAS and DAS at 0/70 and 9/50, respectively, with only 124/130 scores for NAS class identification being strong. However, once the GWN increased to SNR-20, it significantly worsened the model's performance in discriminating between the three classes, leading to the machine model losing its power to discriminate between classes with only 0/50 and 0/70 for CAS and DAS, respectively. The score for NAS sound class was 128/130. The MLP model performed reasonably well for all three classes at no GWN-added levels. However, the model lacked robustness as the CAS and DAS score worsened significantly at GWN SNR-40 levels. The model's discrimination power got even further impacted, and sounds were classed mainly as NAS.

From Fig. 5.2.26, it can be observed that the spectrogram-based Multilayer Perceptron (MLP) achieves a reasonable performance of precision compared to recall (PR) for the CAS and NAS classes, but with significantly lower diagnostic levels for the DAS class as exemplified via precision to recall area under the curve (PR-AUC). At medium levels of GWN of SNR-40, all three classes, especially CAS and NAS, are being impacted, but the DAS class is showing robustness, as exemplified by PR-AUC. However, DAS is still the worst-performing class at medium levels of GWN for this model. Finally, once GWN is increased to SNR-20 levels, a drop in precision to recall is observed with reduced PR-AUC values, impacting particularly the DAS class, with a lesser impact on CAS and very little impact on NAS, indicating how high levels of noise impact all three classes of lung sounds at very different levels but all negatively.

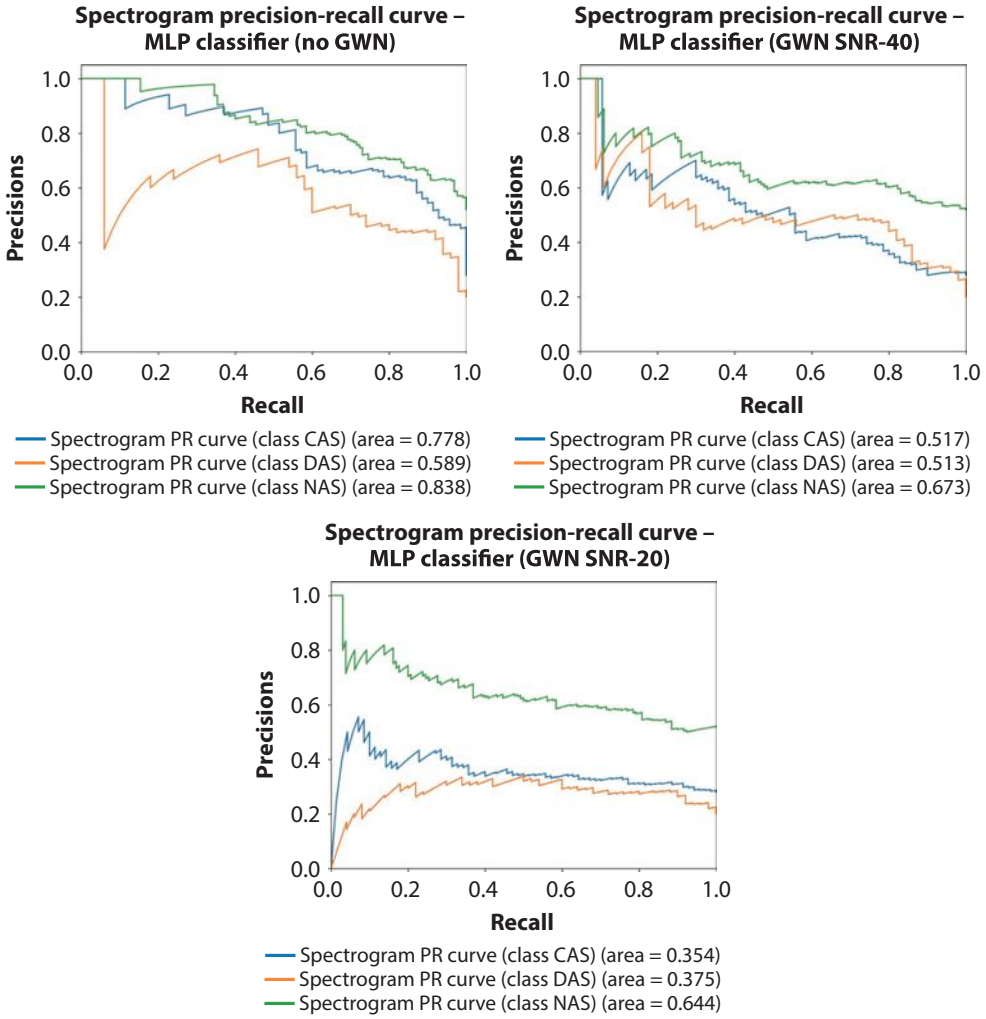


Fig. 5.2.26. *PR curve showing significant impact (P = 0.000) of GWN on MLP model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, MLP – Multilayer Perceptron.

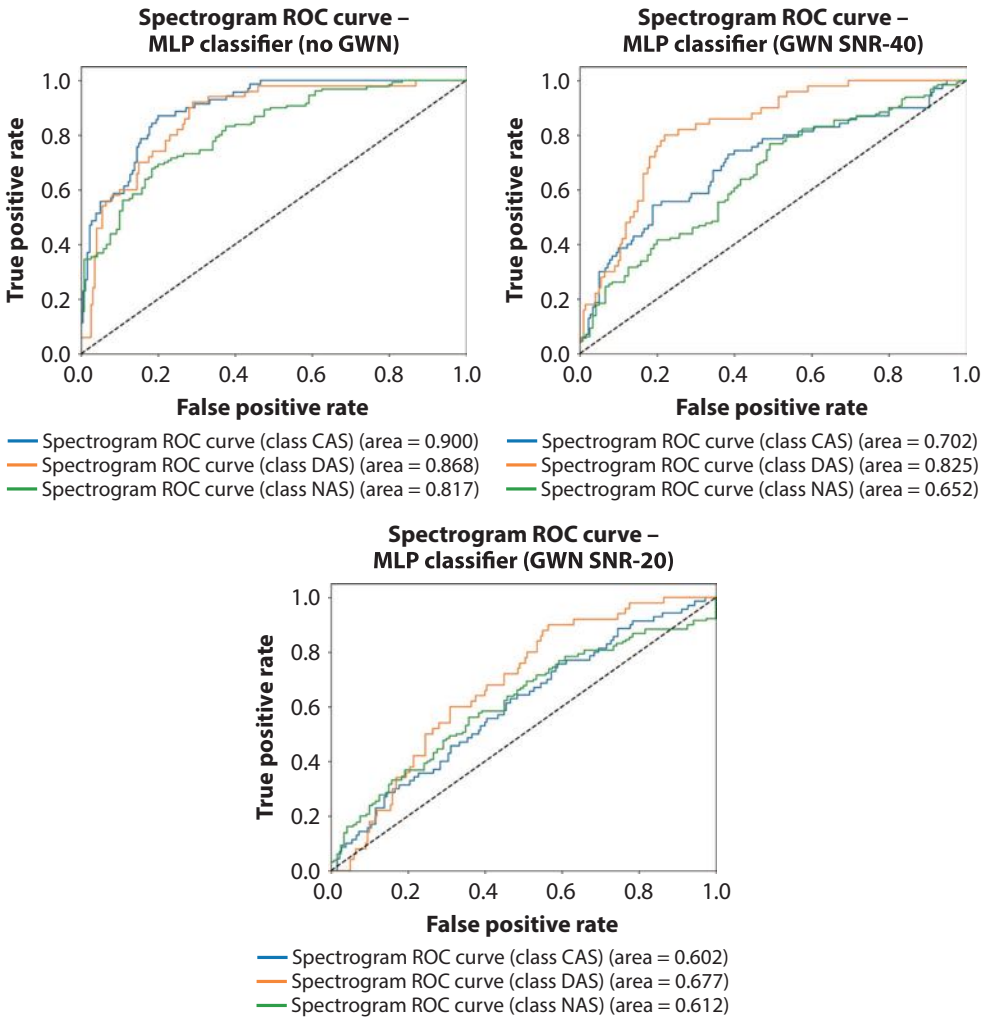


Fig. 5.2.27. *ROC curve showing significant impact (P = 0.000) of GWN on MLP model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, MLP – Multilayer Perceptron.

From Fig. 5.2.27, the impact of Gaussian white noise (GWN) levels on true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based Multilayer Perceptron (MLP) ML model. At no GWN added level, the model shows a very good performance, especially for CAS, followed by DAS classes with weaker (but still very good) performance for the NAS class as observed from the ROC area under the curve (ROC-AUC)

scores. The model performance drops at medium levels of GWN (SNR-40) for two groups: CAS and particularly NAS. DAS class maintains a very good performance at medium Gaussian white noise (GWN) levels. Finally, once GWN at SNR-20 is added to all three classes, all models' performance becomes very average. Nonetheless, all classes maintain a good ratio of TPR to FPR with a reasonable ROC-AUC throughout all three levels of GWN, with the DAS class performing the best. The spectrogram-based MLP model does not have the highest values for no GWN-added levels for all sound classes. However, it maintains robustness at all three levels of GWN, which is one of the few models with this property in ROC graph.

From Fig. 5.2.28, the impact of Gaussian white noise was observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the spectrogram-based Random Forest ML model's confusion matrix. At no Gaussian white noise (GWN) added level, TP scores were reasonable for CAS and NAS recognition, with scores of 45/70 and 101/130, respectively. The model struggled with the DAS class, the TP rate standing only at 14/50. The confusion matrix at SNR-40 (medium levels) showed no significant impact on the model, with scores for CAS and DAS at 8/70 and 3/50, respectively, with only 123/130 scores for NAS. However, once the GWN increased to SNR-20, it significantly worsened the model's performance in discriminating between the three sound classes. The machine model lost its ability to discriminate between classes, with only 0/50, 19/70, and 93/130 scores for DAS, CAS, and NAS, respectively.

The spectrogram-based Random Forest ML model showed overall poor performance as it had problems identifying DAS class sounds even at no GWN added levels. The model was not robust as the CAS and DAS scores worsened significantly at GWN SNR-40 levels. The model's discrimination power got even further reduced, and sounds were classed either as CAS or as NAS at GWN SNR-20 level.

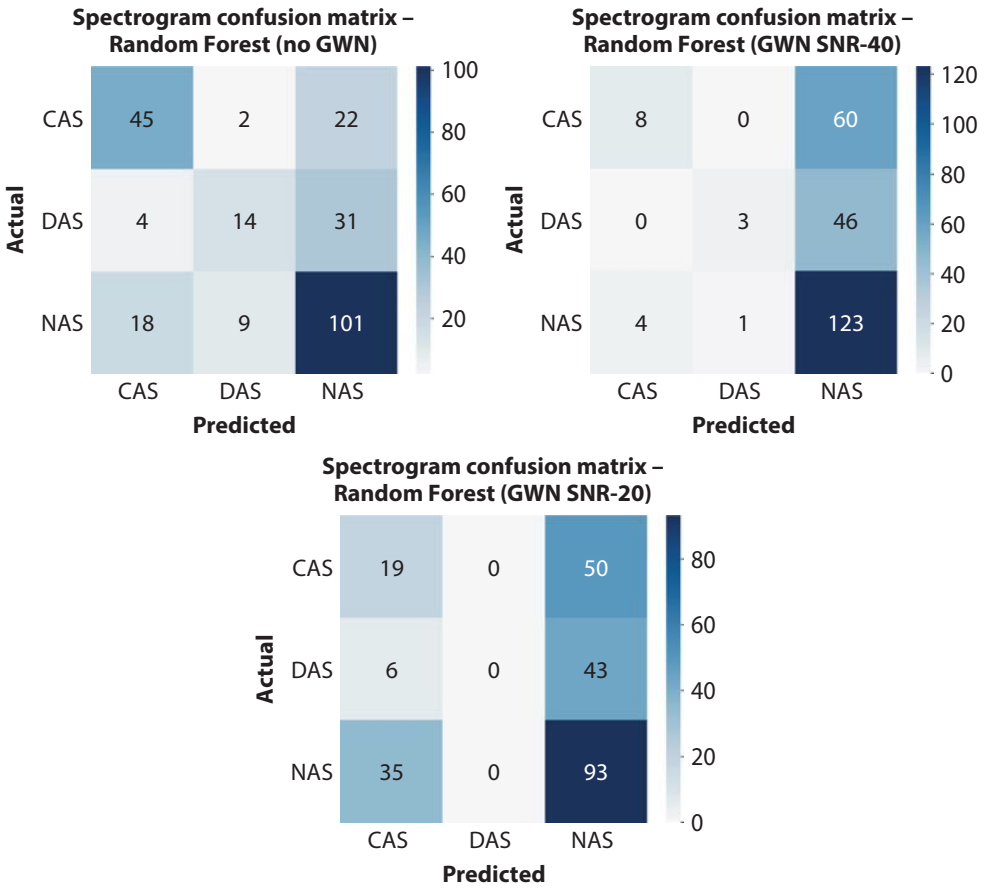


Fig. 5.2.28. *Confusion metrics showing significant impact (P = 0.000) of GWN on Random Forest model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

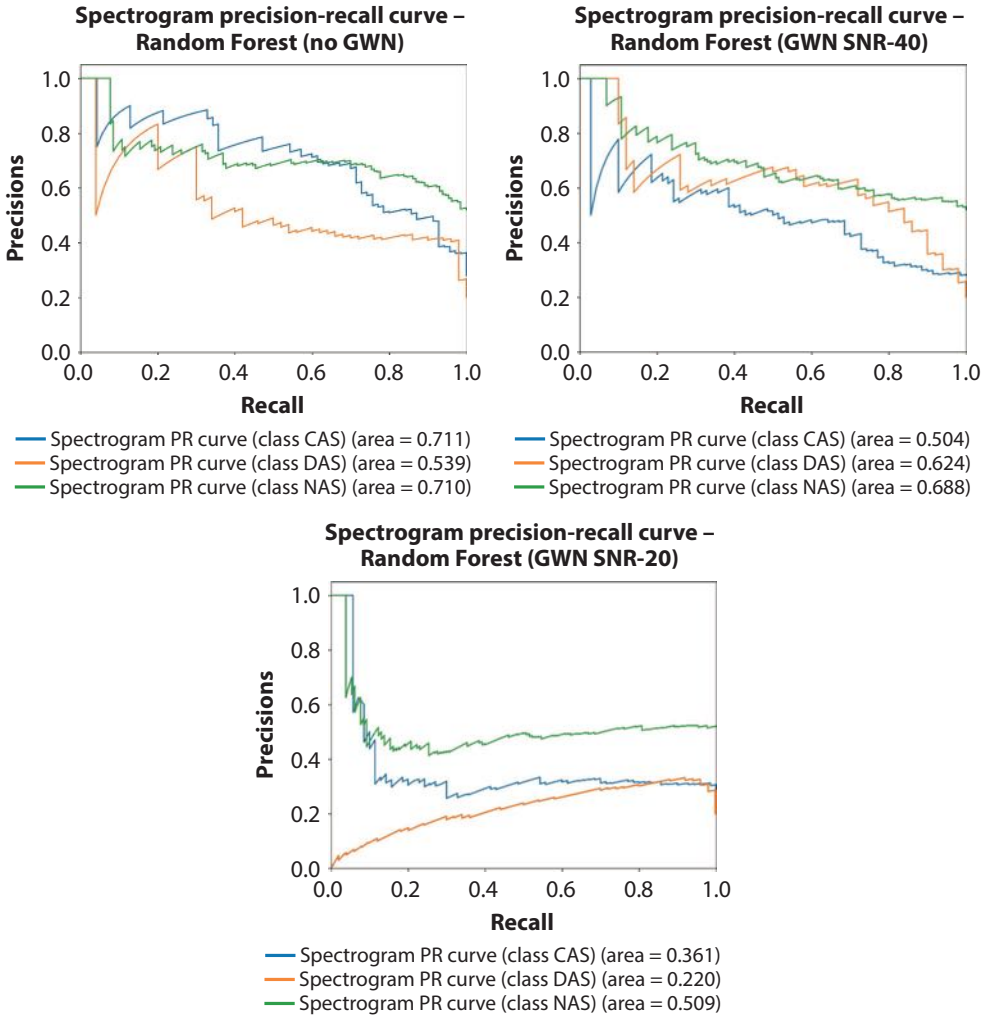


Fig. 5.2.29. *PR curve showing significant impact (P = 0.000) of GWN on Random Forest model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.29, it can be observed that spectrogram-based Random Forest ML model achieves reasonable precision performance compared to recall (PR) for CAS and NAS classes but with significantly lower diagnostic levels for the DAS class, as exemplified via precision to recall area under the curve (PR-AUC). At medium levels of Gaussian white noise (GWN) of SNR-40, two classes recognition was affected: CAS and NAS. However, the DAS class showed robustness, as exemplified by PR-AUC, and higher precision

than recall levels. Finally, once GWN is increased to SNR-20 levels, we see an extreme impact on the DAS class with a lesser impact on CAS and NAS, indicating how high levels of ambient noise impact all three classes of lung sounds identification at very different levels, all negatively.

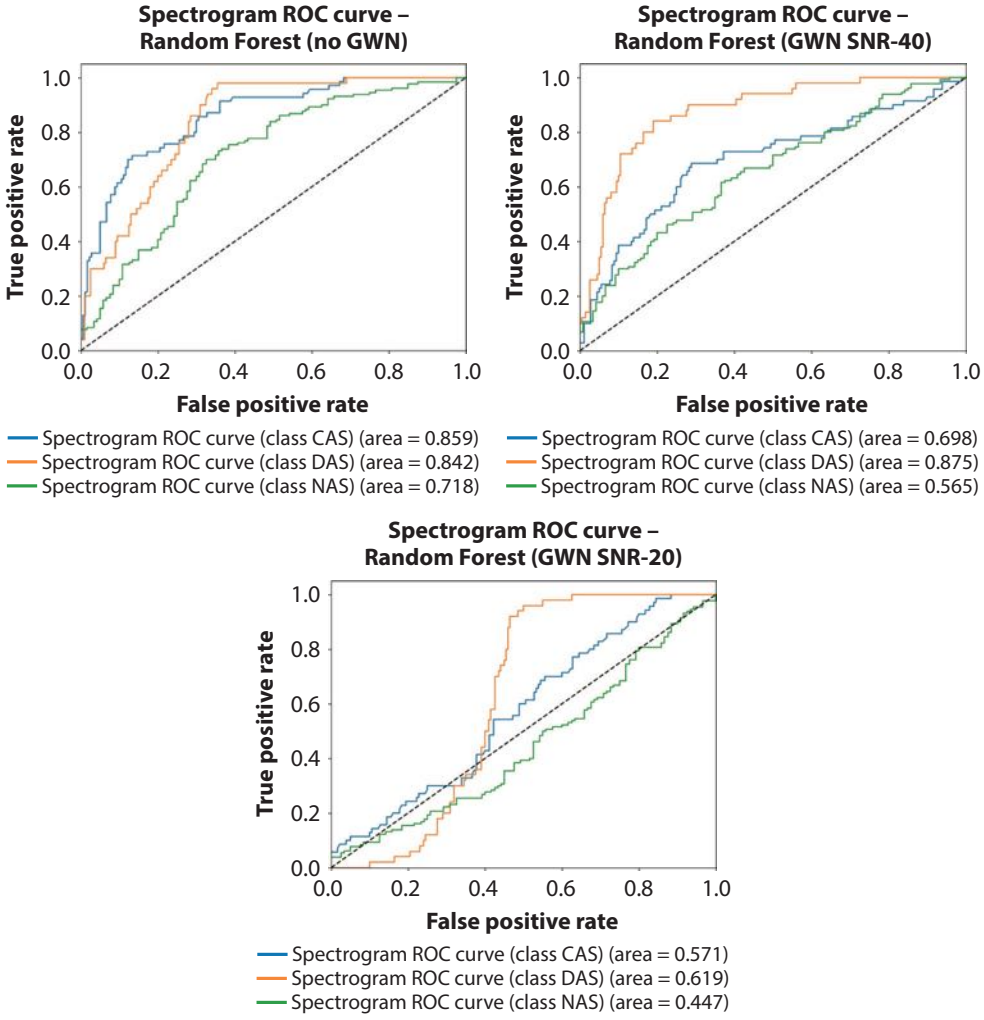


Fig. 5.2.30. *ROC curve showing significant impact (P = 0.000) of GWN on Random Forest model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.30, the impact of (Gaussian white noise) GWN levels on the true positive rate (TPR) compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based Random Forest ML model. At no GWN-added level, the model shows a very good performance, especially for CAS, followed by DAS classes, with weaker (but still of good standard) performance for NAS class as seen from the ROC area under the curve (AUC) scores. The model performance drops at medium levels of GWN at SNR-40 for two groups: DAS and particularly NAS. The DAS class maintains a very good performance at medium levels of GWN. Finally, once GWN at SNR-20 was added, all three classes were heavily impacted. The ML models' performance became very poor at the highest levels of GWN, showing that the model held robustness to ambient noise up to SNR-40 but, at SNR-20, lost its power for all classes with reduced TPR compared to FPR and a drop in ROC-AUC values.

From Fig. 5.2.31, the impact of Gaussian white noise is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) scores as seen in the spectrogram-based SVM model's confusion matrix. At no Gaussian white noise (GWN) level, TP rates were low for CAS and DAS, with scores of 13/70 and 2/50, respectively. The only class that correctly identified was NAS with 122/130 score. The confusion matrix at SNR-40 (medium noise levels) showed no significant impact on the model, with scores for CAS, DAS and NAS at 13/70, 2/50 and 122/130. However, once the GWN increased to SNR-20, it worsened the performance significantly of the model to discriminate between the three, leading to the machine model losing its power to discriminate between classes with only 5/50 and 1/70 for CAS and DAS correctly identified, respectively, with only NAS having a good score of 130/130. The spectrogram-based SVM ML model performed poorly even at the no GWN-added levels for CAS and DAS sound identification. Yet, some robustness was shown by the model, as the scores did not change at GWN SNR-40 levels. However, at GWN SNR-20, the model's discrimination power completely collapsed, and all classes were identified as NAS.

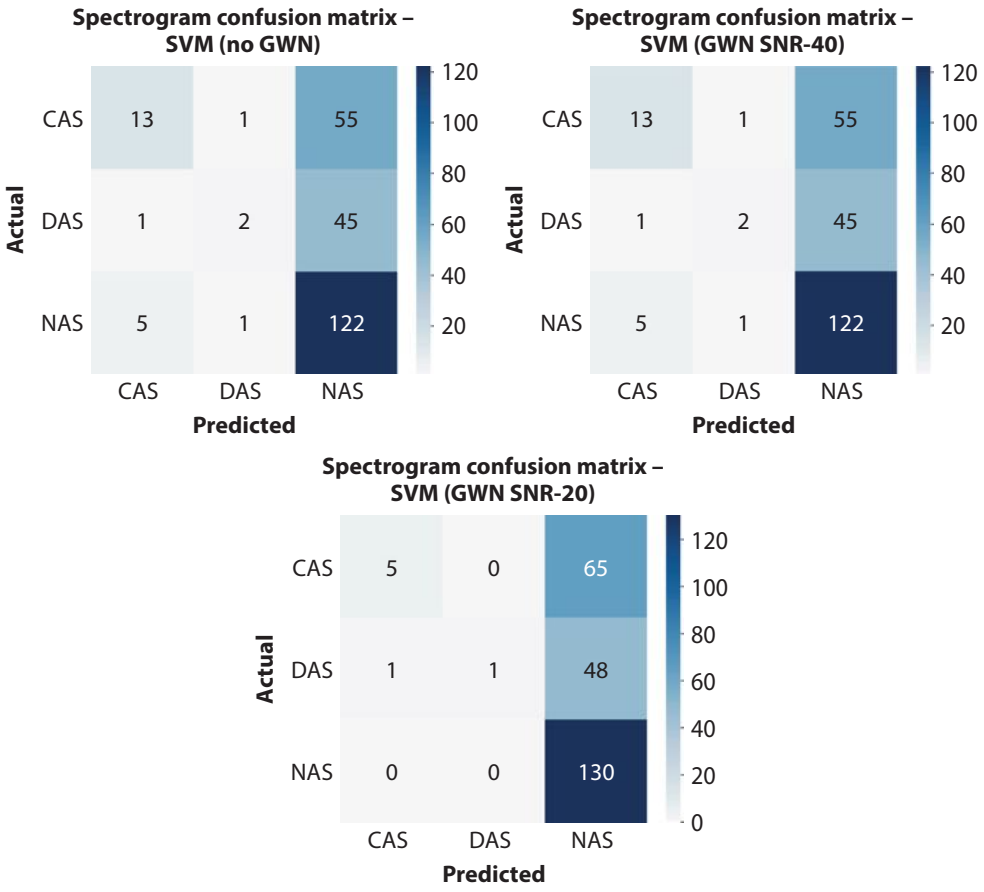


Fig. 5.2.31. *Confusion metrics showing significant impact (P = 0.000) of GWN on SVM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound, SVM – Support Vector Machines.

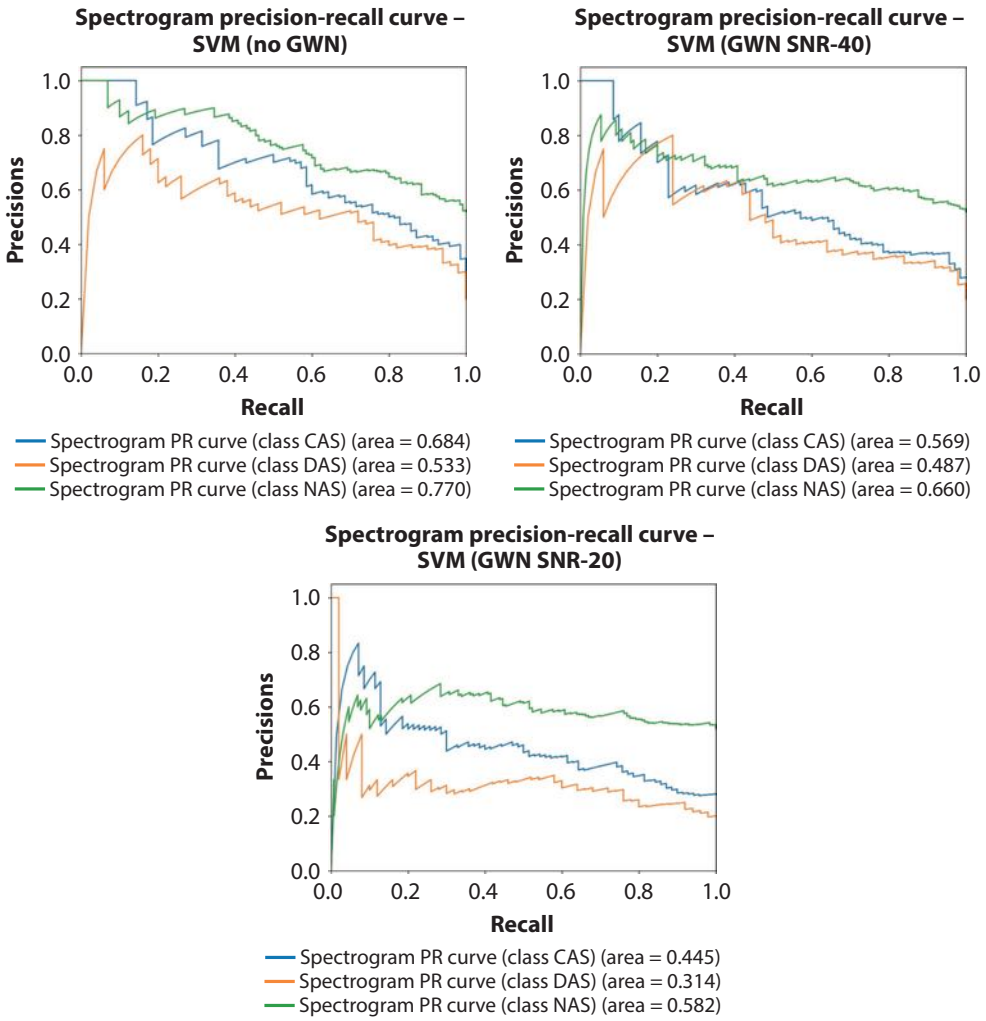


Fig. 5.2.32. *PR curve showing significant impact (P = 0.000) of GWN on SVM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, SVM – Support Vector Machines.

From Fig. 5.2.32, it can be observed that the spectrogram-based Support Vector Machines (SVM) model achieves a reasonable performance of precision compared to recall for the CAS and NAS class of sounds but with significantly lower diagnostic levels for the DAS class, as exemplified via precision-recall area under the curve (PR-AUC) values. At medium levels of Gaussian white noise (SNR-40), all three classes are affected. Finally, once GWN is increased to SNR-20 level, we see an extreme impact on the DAS class with a lesser impact on CAS and NAS, indicating how high levels of

noise impact all three classes of lung sounds at very different levels, but all negatively. Though the spectrogram-based SVM ML model achieved reasonable performance overall at no GWN-added, it came with a caveat of varied performance between groups, with DAS sound class identification performance being very poor. Yet, the SVM model showed little robustness to GWN at medium and even more so at high GWN levels.

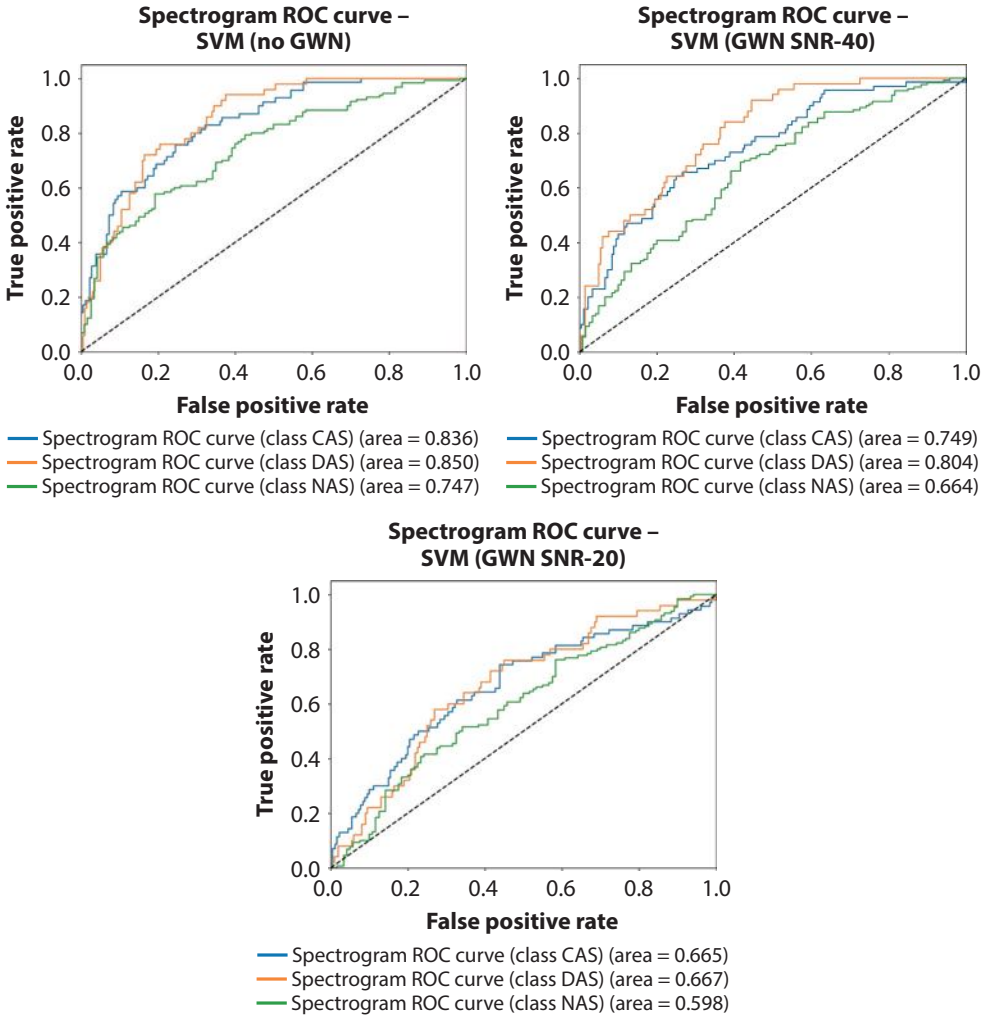


Fig. 5.2.33. *ROC curve showing significant impact (P = 0.000) of GWN on SVM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound, SVM – Support Vector Machines.

From Fig. 5.2.33, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based Support Vector Machines (SVM) model. At no GWN added level, this ML model shows a good performance, especially for DAS, followed closely by CAS classes with weaker (but still of good standard) performance for NAS class as seen from ROC area under the curve (ROC-AUC) scores. The model performance drops a bit at medium levels of Gaussian white noise (SNR-40) for two sound groups recognition: CAS and NAS classes. Meanwhile, the DAS class maintained excellent performance at medium levels of GWN. Finally, once GWN at SNR-20 was added, all three classes were heavily impacted. The performance became poorer at the highest levels of GWN, especially for the NAS class, showing that the SVM model holds robustness to ambient noise up to SNR-40 but, at SNR-20, loses its power for all classes with a drop of TPR compared to FPR and decreased ROC-AUC values.

From Fig. 5.2.34, the impact of Gaussian white noise is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the spectrogram-based Extreme Gradient Boosting classifier (XGBoost) ML model's confusion matrix. At no Gaussian white noise added level, TP values for CAS, DAS, and NAS sound classes were 50/70, 15/50 and 96/130, respectively. The confusion matrix at SNR-40 (medium levels of GWN) showed ML model's decreased ability to identify CAS classes correctly with a TP score of 17/70. DAS and NAS scores were at 26/50 and 111/130, respectively. However, once the GWN increased to SNR-20, it significantly worsened the model's performance to discriminate ability between the three sound classes, leading to the XGBoost machine learning model to lose its power to discriminate between classes with only 3/50 and 0/70 for CAS and DAS correctly identified. The NAS class at the highest GWN was identified. The DAS classes were hugely impacted, with 0/50 identified correctly, whilst CAS identification was at 31/70 and 62/130 for NAS. The XGBoost showed a good performance at no GWN added levels, especially for CAS and DAS classes, and reasonable robustness at GWN SNR-40 levels. However, once the levels increased, the ability of the model to discriminate between three classes was reduced to two: CAS and NAS.

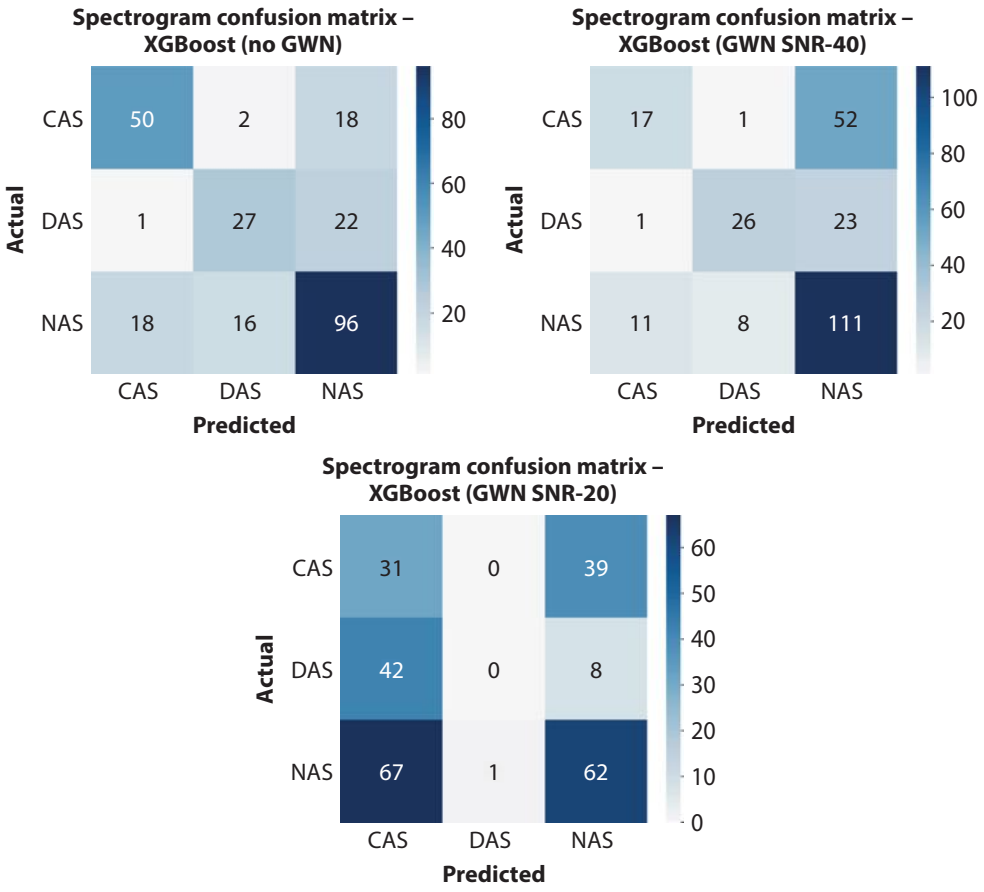


Fig. 5.2.34. *Confusion metrics showing significant impact (P = 0.000) of GWN on XGBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound, XGBoost – Extreme Gradient Boosting classifier.

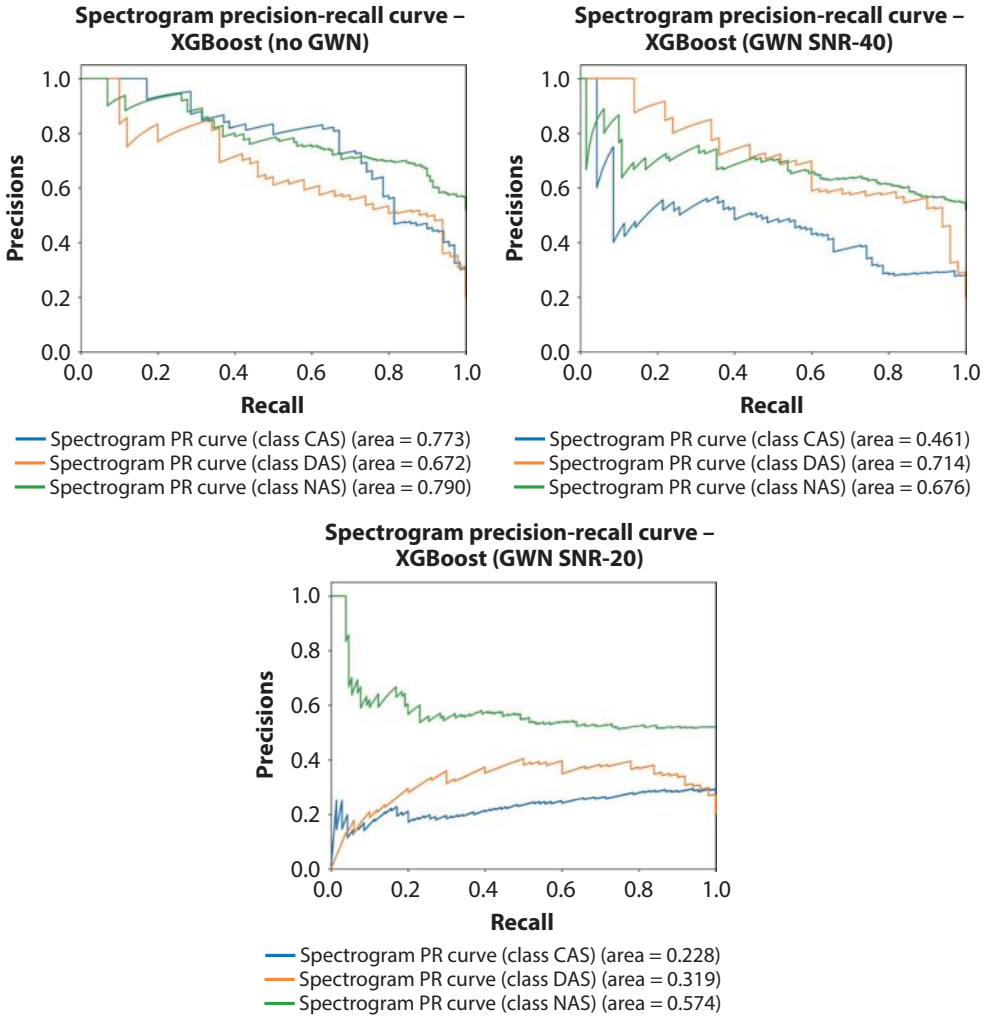


Fig. 5.2.35. *PR curve showing significant impact (P = 0.000) of GWN on XGBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, XGBoost – Extreme Gradient Boosting classifier.

From Fig. 5.2.35, it can be observed that spectrogram-based Extreme Gradient Boosting classifier (XGBoost) ML model achieved a good performance of precision compared to recall for the CAS and NAS classes but with slightly lower diagnostic levels for the DAS class, as exemplified via precision to recall area under the curve (PR-AUC). At medium levels of Gaussian white noise (SNR-40), the CAS class was impacted the most, with

a significant drop in precision compared to recall. Meanwhile, NAS sound class identification showed reasonable robustness to medium noise impact. Whilst DAS showed a very strong performance. Finally, once GWN is increased to SNR-20 levels, we see an extremely negative impact on NAS and DAS classes recognition with a lesser impact on NAS sound class.

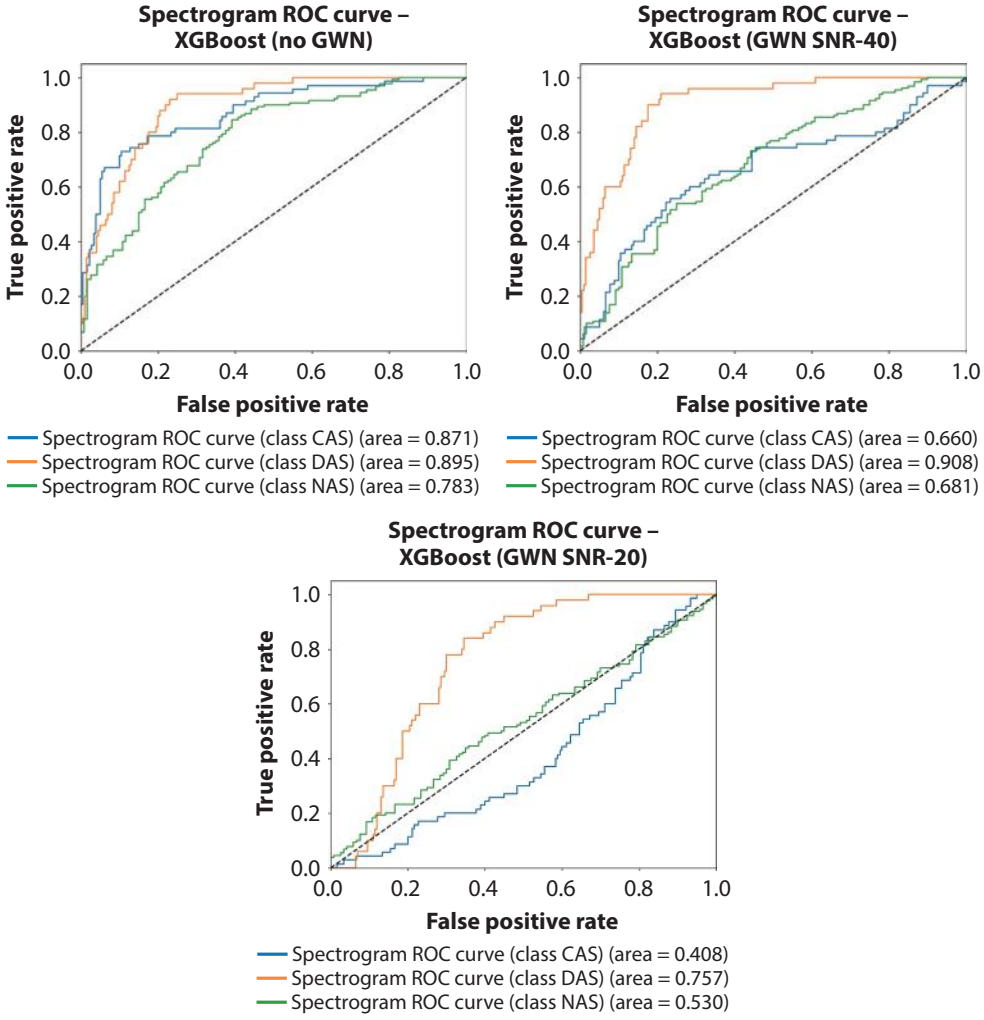


Fig. 5.2.36. *ROC curve showing significant impact (P = 0.000) of GWN on XGBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound, XGBoost – Extreme Gradient Boosting classifier.

From Fig. 5.2.36, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based Extreme Gradient Boosting classifier (XGBoost) model. At no GWN-added level, the model shows a very good performance, especially for DAS, followed closely by the CAS class with slightly weaker (but still of good standard) performance for the NAS class recognition, as seen from the ROC area under the curve (ROC-AUC) scores. The model performance drops significantly at medium levels of GWN (SNR-40) for two groups: CAS and NAS. The DAS class maintains a very good performance at medium levels of GWN. Finally, once GWN at SNR-20 is added, all three classes are very heavily impacted. The ML model's performance becomes poorer at the highest levels of GWN, especially for CAS and NAS sound identification, with only good performance for DAS. Therefore, the XGBoost model holds robustness to ambient noise up to SNR-40 but, at SNR-20, loses its power for two classes, making it unviable at the highest levels of GWN.

From Fig. 5.2.37, the impact of Gaussian white noise (GWN) is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the scalogram-based AdaBoost model's confusion matrix. At no GWN added levels, only 4/70 CAS was identified correctly, 15/50 of the DAS class was correctly identified, and 112/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium GWN levels) performs better by determining TP for CAS, DAS and NAS at 16/70, 30/50 and 101, respectively. However, once the GWN increased to SNR-20, it significantly worsened the model's performance in discriminating between the three classes, leading to the machine model losing its power to discriminate between classes with only 3/50 and 0/70 for CAS and DAS correctly identified, respectively. The NAS class at the highest GWN was identified at 121/130 score. The scalogram-based AdaBoost ML model showed poor performance at no GWN levels for CAS and DAS classes, showing bias towards NAS classification. This changed significantly at GWN SNR-40 levels, with the model performing better predictions for all three classes, especially CAS and DAS. Though once the highest levels of GWN were introduced at SNR-20 levels, the scalogram-based AdaBoost model lost its discrimination power completely and classed all lung sounds as NAS.

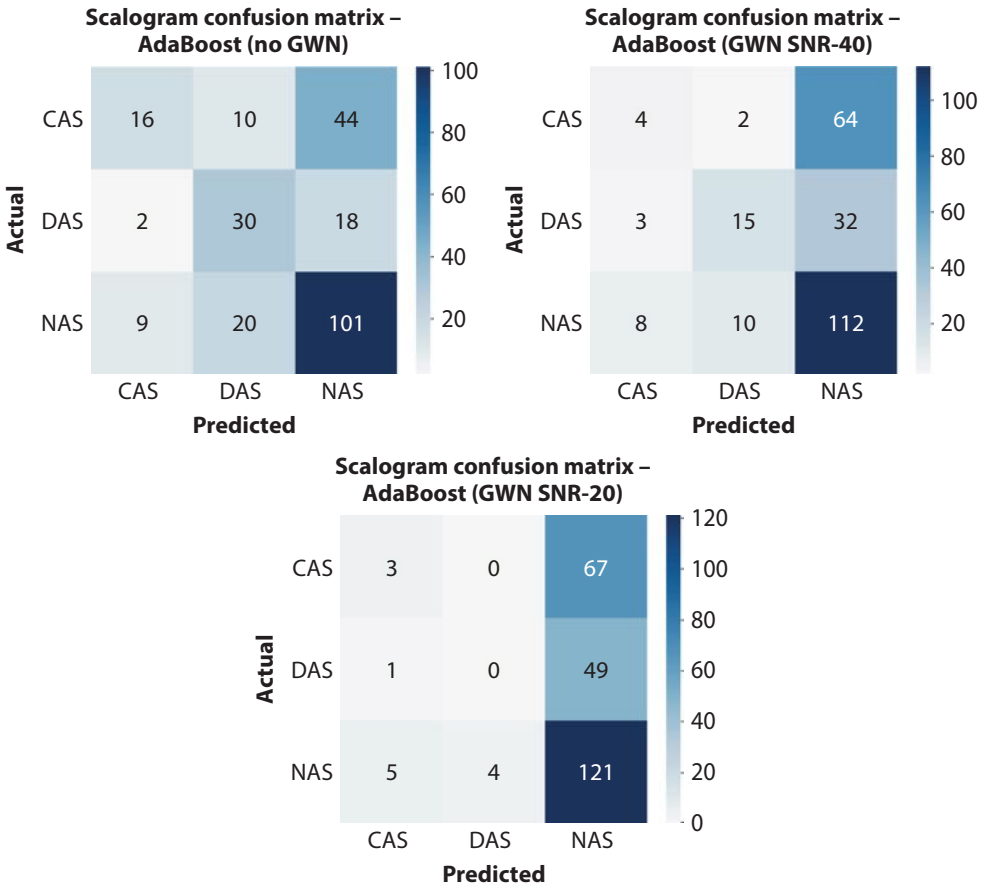


Fig. 5.2.37. *Confusion metrics showing significant impact (P = 0.000) of GWN on AdaBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, AdaBoost – Adaptive Boosting.

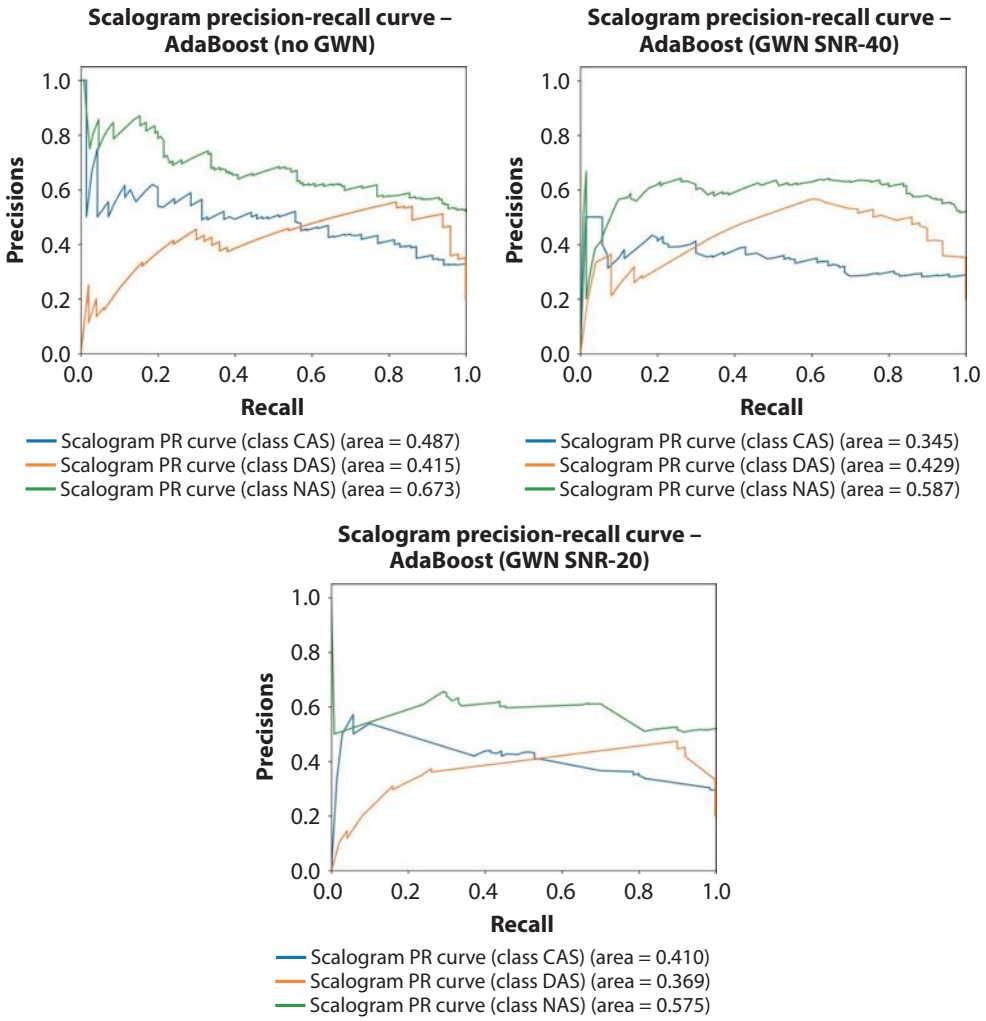


Fig. 5.2.38. *PR curve showing significant impact (P = 0.000) of GWN on AdaBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound, AdaBoost – Adaptive Boosting.

From Fig. 5.2.38, it can be observed that scalogram-based AdaBoost achieved a poor precision performance compared to recall for CAS, DAS classes but with slightly better performance diagnostic levels for the NAS class as exemplified via precision-recall area under the curve (PR-AUC) values. At medium levels of Gaussian white noise (SNR-40), CAS and NAS classes recognition was impacted the most with a significant drop in precision compared to recall, whilst DAS class identification showed reasonable

robustness, but it has to be remembered that its scores were the worst at no GWN levels. Finally, once GWN is increased to SNR-20 levels, we see an extremely negative impact on DAS class with a lesser impact on NAS and CAS classes compared to no GWN levels.

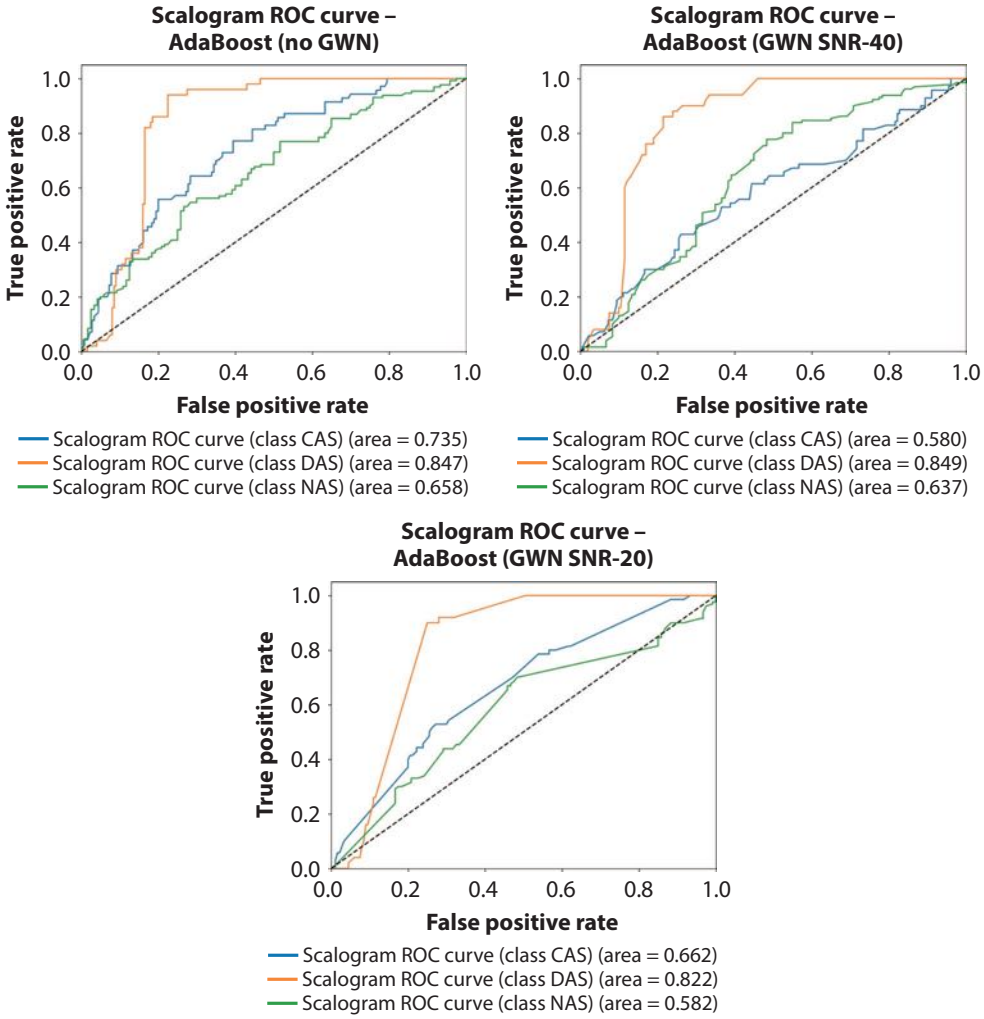


Fig. 5.2.39. *ROC curve showing significant impact (P = 0.000) of GWN on AdaBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, AdaBoost – Adaptive Boosting.

From Fig. 5.2.39, the impact of Gaussian white noise (GWN) levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the scalogram-based Adaptive Boosting (AdaBoost) model. At no GWN added level, the model shows a good performance, especially for DAS, followed by CAS classes with weaker (but only average standard) performance for NAS class as seen from ROC area under the curve (ROC-AUC) scores. The model performance drops significantly at medium levels of GWN at SNR-40 for two groups: CAS and NAS. The DAS class maintains a very good performance at medium levels of GWN. Finally, once GWN at SNR-20 is added, all three classes will be impacted. The performance becomes poorer at the highest levels of GWN, especially for CAS and NAS, with only very good performance for DAS. Therefore, the model holds robustness to ambient noise up to SNR-40 but, at SNR-20, loses its power for two sound classes recognition, making the AdaBoost model useless at the highest levels of GWN.

From Fig. 5.2.40, the impact of Gaussian white noise (GWN) is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the scalogram-based Categorical data Gradient Boosting (CatBoost) model's confusion matrix. At no GWN added levels, 33/70 CAS was identified correctly, with 28/50 of DAS class correctly identified and 91/130 of NAS correctly identified. The confusion matrix at SNR-40 (medium levels) shows a significant decrease in the model's performance, with a drop in TP for CAS and DAS at 15/70 and 20/50 respectively. However, it maintained reasonable performance for NAS 97/130. The increase in the GWN levels to SNR-20 significantly worsens the model's performance in discriminating between the three classes, leading to the machine model to lose its power to discriminate between classes, with only 0/50 and 11/130 for DAS and NAS sounds correctly identified. The CAS class at the highest GWN was identified at 60/70. This showed that even though the scalogram-based CatBoost model showed a reasonable performance at no GWN level for all three classes. The model showed some resistance to ambient noise, and it still discriminated all three classes with varied TP values and ever-increasing FP values. Once GWN was increased to SNR-20, all classes were primarily identified as CAS and became useless at discriminating between all three classes.

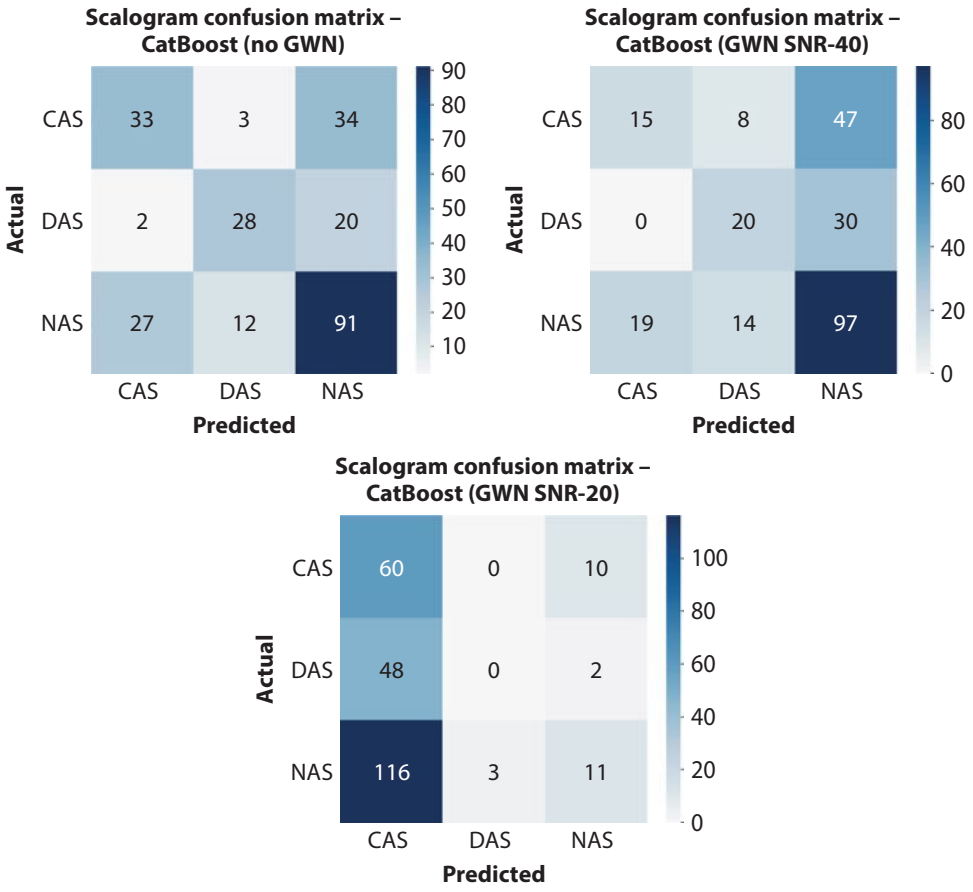


Fig. 5.2.40. *Confusion metrics showing significant impact (P = 0.000) of GWN on CatBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, CatBoost – Categorical data Gradient Boosting.

From Fig. 5.2.41, it can be observed that the scalogram-based categorical boosting (CatBoost) model achieved a weaker performance of precision compared to recall for the CAS class of sounds but with slightly better performance diagnostic levels for DAS and NAS classes, as exemplified via area under the curve (AUC) for PR curve. At medium levels of GWN of SNR-40, CAS and DAS, classes were most heavily impacted, with a lesser impact on NAS. Finally, once GWN is increased to SNR-20 levels, we see an extremely negative impact on DAS classes with a lesser impact on NAS and CAS classes than on no GWN levels.

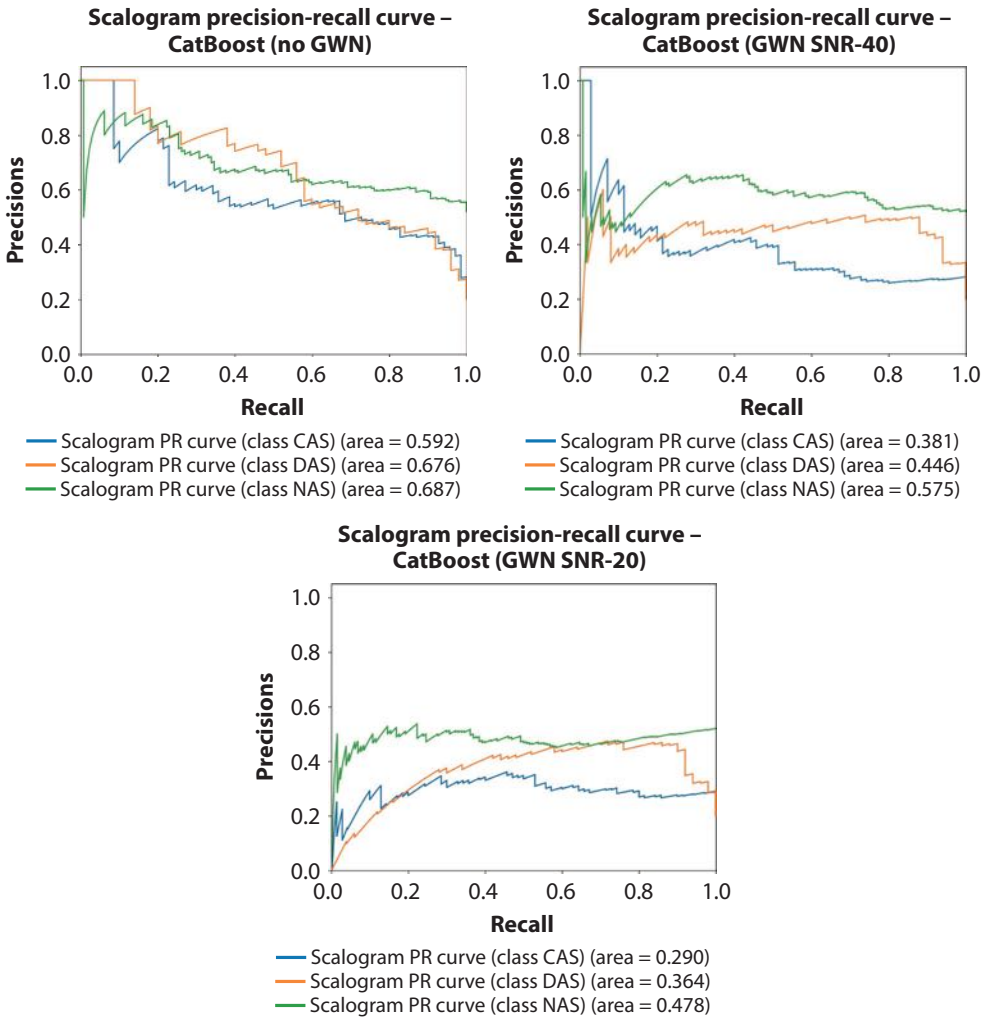


Fig. 5.2.41. *PR curve showing significant impact (P = 0.000) of GWN on CatBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, CatBoost – Categorical data Gradient Boosting.

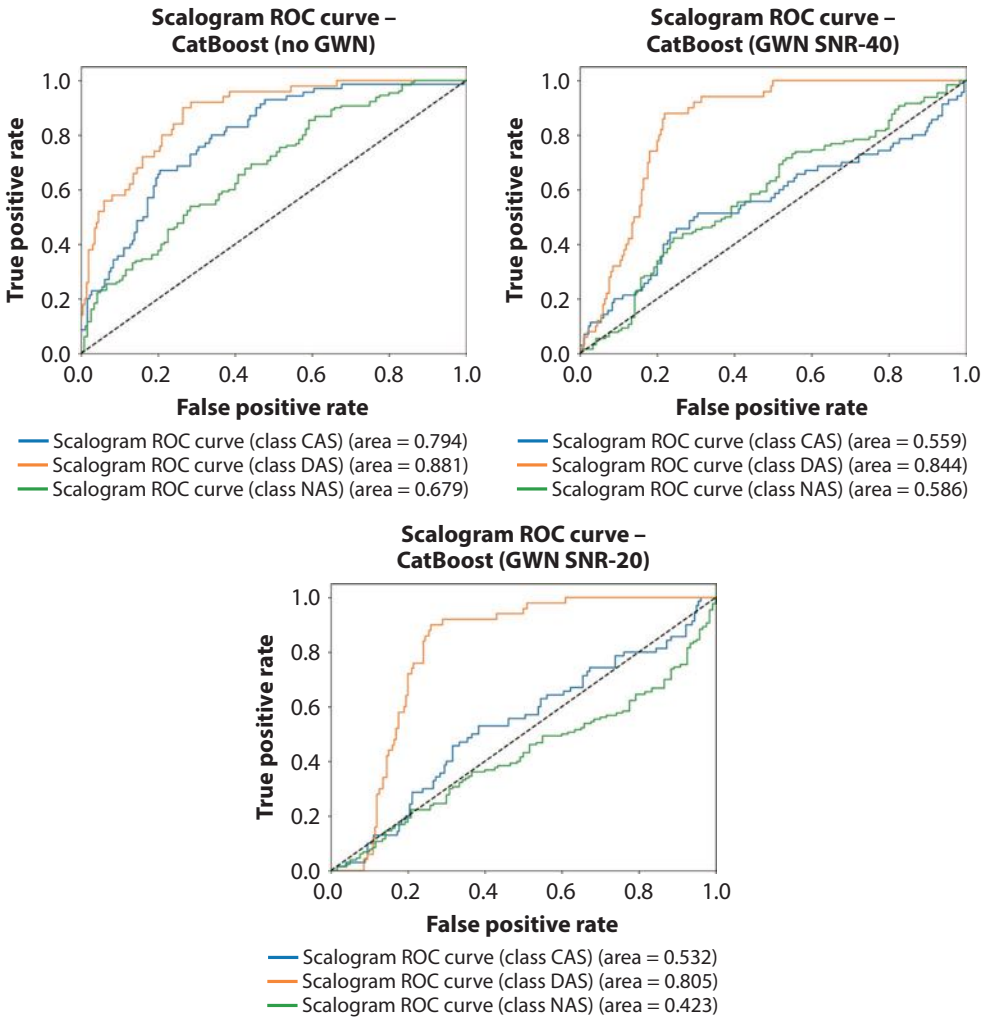


Fig. 5.2.42. *ROC curve showing significant impact (P = 0.000) of GWN on CatBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, CatBoost – Categorical data Gradient Boosting.

From Fig. 5.2.42, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the scalogram-based categorical boosting (CatBoost) model. At no GWN added level, the model shows a good performance for the DAS class, followed by CAS classes with the weakest performance for the NAS as seen from the ROC area under the curve

(ROC-AUC). The model performance drops significantly at medium levels of GWN (SNR-40) for two groups: CAS and NAS. Meanwhile, the DAS class performs well at medium levels of GWN. Finally, once GWN increased to SNR-20 level, all two classes will be very heavily impacted. The CatBoost model's performance becomes poorer at the highest levels of GWN, especially for CAS and NAS sound recognition, and true favourable rates decrease to an inferior level with only good performance for DAS sound group. Therefore, the CatBoost model shows reasonable performance at no GWN levels, but already at SNR-40, the model starts having problems with CAS and NAS classes identification. Therefore, this model is only a valuable as a diagnostic tool at no GWN level.

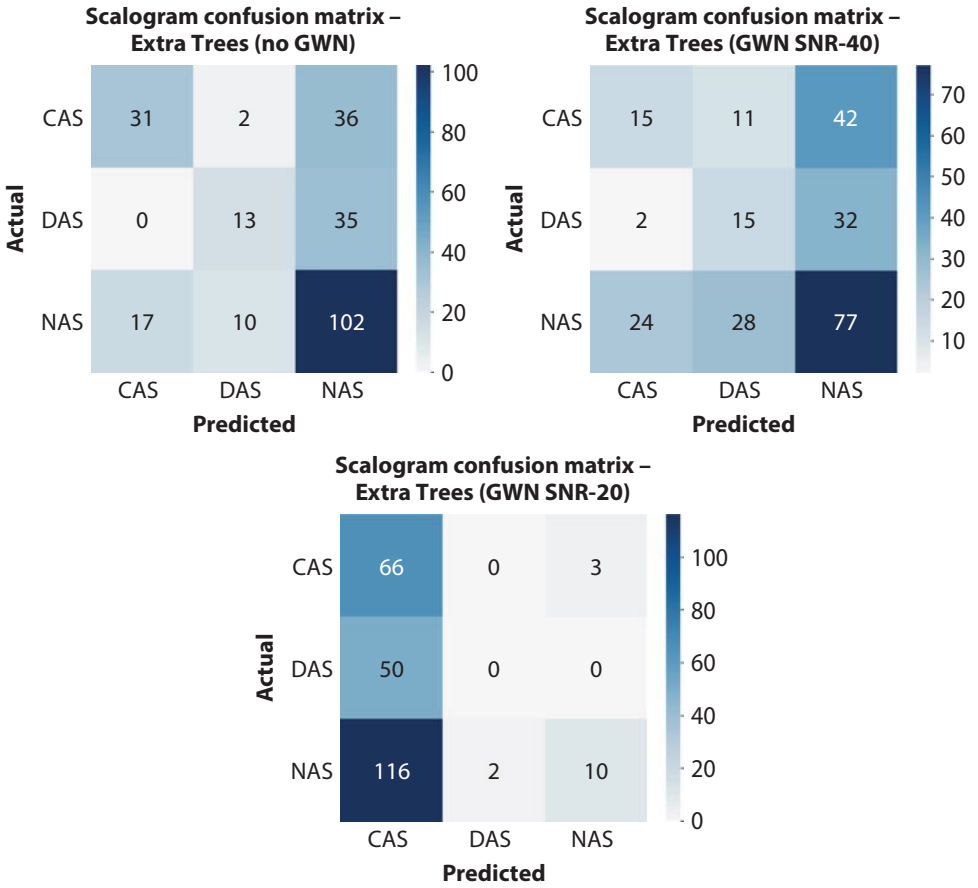


Fig. 5.2.43. *Confusion metrics showing significant impact (P = 0.000) of GWN on Extra Trees model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.43, the impact of Gaussian white noise is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the scalogram-based Extra Trees model's confusion matrix. At no GWN added levels, 31/70 CAS was identified correctly, 13/50 of the DAS class was correctly identified, and 102/130 of the NAS was correctly identified. The confusion matrix at SNR-40 (medium levels) shows a decrease in model performance with a drop in TP for CAS and NAS at 15/70 and NAS at 77/130, respectively, but maintained performance (although still poor) for DAS at 15/50. The increase in the GWN levels to SNR-20 significantly worsens the model's performance in discriminating between the three sound classes, leading to the machine model losing its power to discriminate between classes, with only 0/50 for DAS and 10/130 for NAS correctly identified. The CAS class at the highest GWN was identified at 66/70. This showed that the scalogram-based Extra Trees model had a poor diagnostic accuracy for the DAS class even at no GWN level and showed a lack of robustness when GWN level was increased to SNR-40; class discrimination significantly worsened with GWN SNR-20 level where all classes were mostly identified as CAS.

From Fig. 5.2.44, it can be observed that scalogram-based Extra Trees ML model achieved an overall reasonable performance of precision compared to recall for CAS and DAS class of sounds but with significantly worse diagnostic performance for the DAS class as exemplified via precision to recall area under the curve (PR-AUC). At medium levels of GWN (SNR-40) in all three classes, the DAS class felt the most significant impact. Finally, once GWN is increased to SNR-20 level, we see an extremely negative impact on DAS class with a lesser impact on CAS. The NAS class shows robustness, but the results are worse than no GWN level, as the PR-AUC shows.

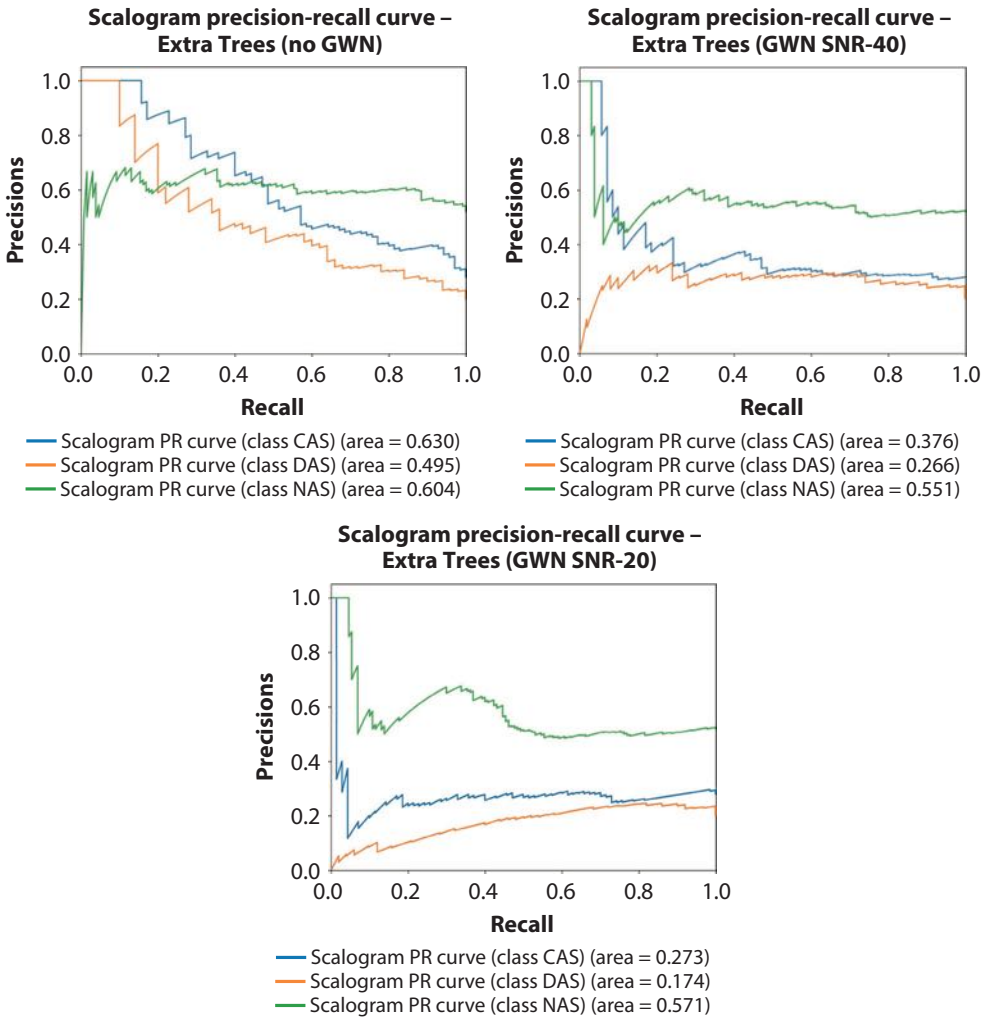


Fig. 5.2.44. *PR curve showing significant impact (P = 0.000) of GWN on Extra Trees model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

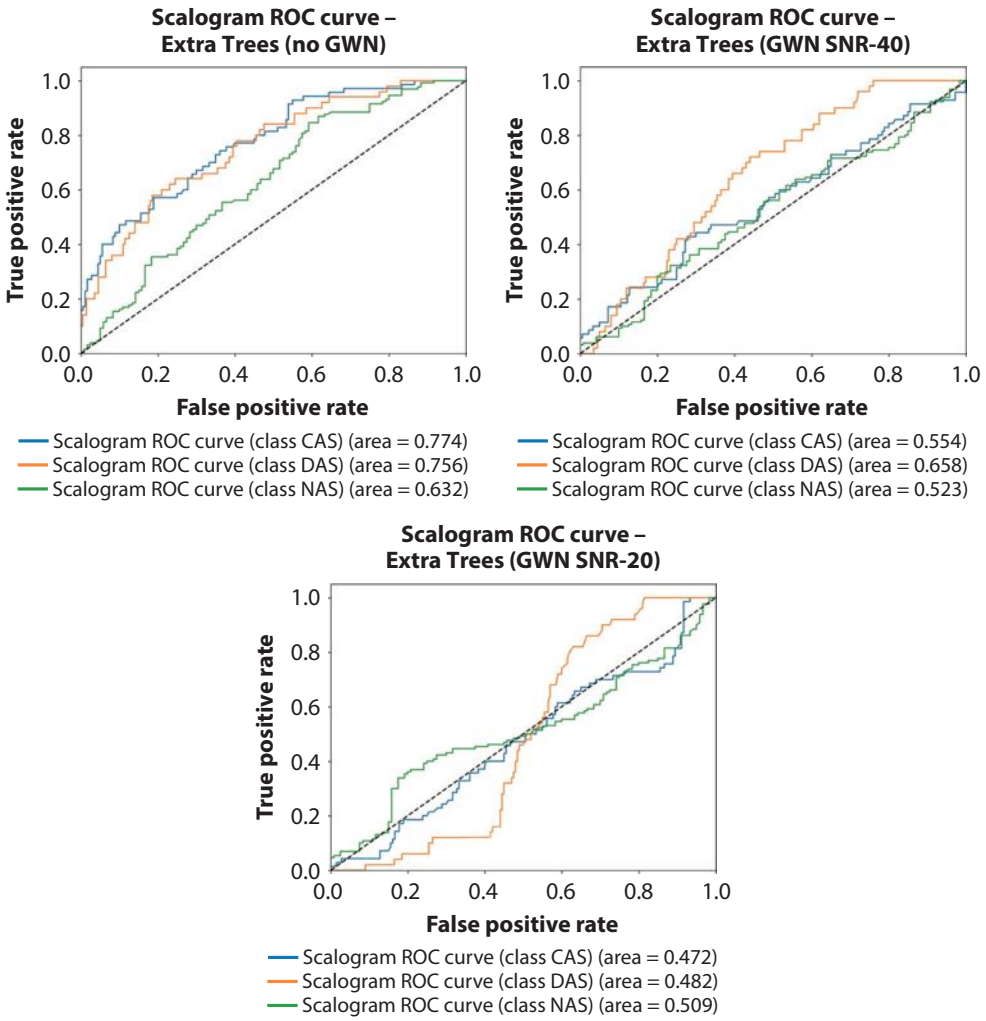


Fig. 5.2.45. *ROC curve showing significant impact (P = 0.000) of GWN on Extra Trees model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.45, the spectrogram-based Extra Trees model can observe the impact of GWN levels on the true positive rate (TPR) compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve. At no GWN added to the level, the model shows a reasonable performance, especially for CAS, followed by DAS classes with weaker (but only of average standard) performance for the NAS class, as seen from the ROC graph's area under the curve (AUC). The model performance drops significantly at

medium levels of GWN (SNR-40) for all three groups, especially CAS and NAS. Finally, once GWN at SNR-20 is added, all three classes will be heavily impacted. The spectrogram-based Extra Trees ML model shows reasonable performance at no GWN level. However, once GWN is added even at medium level (SNR-40), the model loses its power with a drop of TPR compared to FPR, especially for CAS and NAS sound classes identification. Therefore, the Extra Trees model is only useful, according to the ROC curve, without GWN, because it lacks robustness to noise.

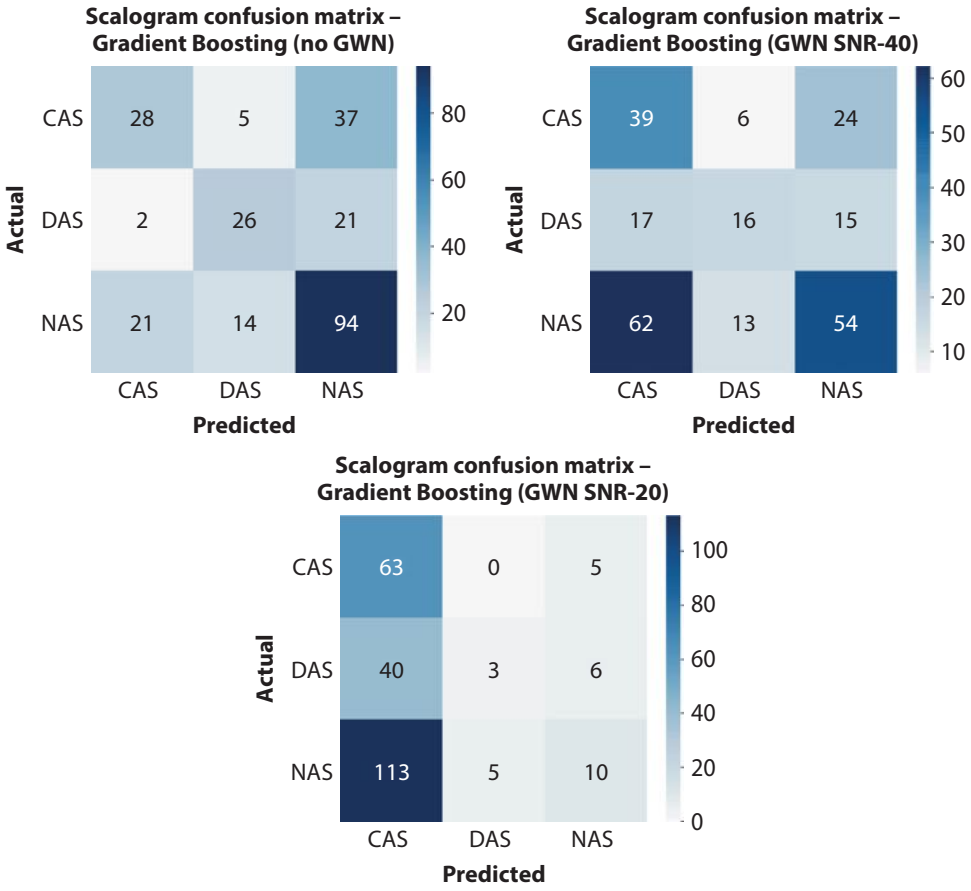


Fig. 5.2.46. *Confusion metrics showing significant impact (P = 0.000) of GWN on gradient boosting model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.46, the impact of Gaussian white noise (GWN) is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the scalogram-based Gradient Boosting ML model's confusion matrix. At no GWN level, 39/70 CAS was identified correctly, 16/50 of the DAS class was correctly identified, and 54/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium levels) shows a decrease in model performance with a drop-in TP for CAS 28/70 but maintained a good performance for DAS and NAS classes with 26/50 and 94/130, respectively. The increase in the GWN level to SNR-20 worsens the model's performance in discriminating between the three, leading to the machine model losing its power to discriminate between sound classes with only 3/50 and 10/130 for DAS and NAS correctly identified. The CAS class at the highest GWN was identified at 63/70 score. This showed that even though the model had a reasonable diagnostic accuracy for all three sound classes at no GWN level, also it showed robustness when GWN levels were increased to medium level (SNR-40), yet, class discrimination significantly worsened with GWN SNR-20 level, where all classes were mostly identified as NAS.

From Fig. 5.2.47, it can be observed that scalogram-based Gradient Boosting ML model achieved an overall reasonable to good performance of precision to recall (PR) for NAS and DAS, with the worst performance observed for CAS as exemplified via PR-AUC. At medium levels of GWN of SNR-40 in all three sound classes, the DAS class recognition was the most significant impacted, followed closely by worst results in CAS class identification. Finally, once GWN is increased to SNR-20 levels, it negatively impacts CAS class with a lesser impact on DAS class sound recognition. The Gradient Boosting ML model shows some robustness in NAS class identification; nonetheless, all three classes' PR-AUC values drop significantly at high levels of GWN.

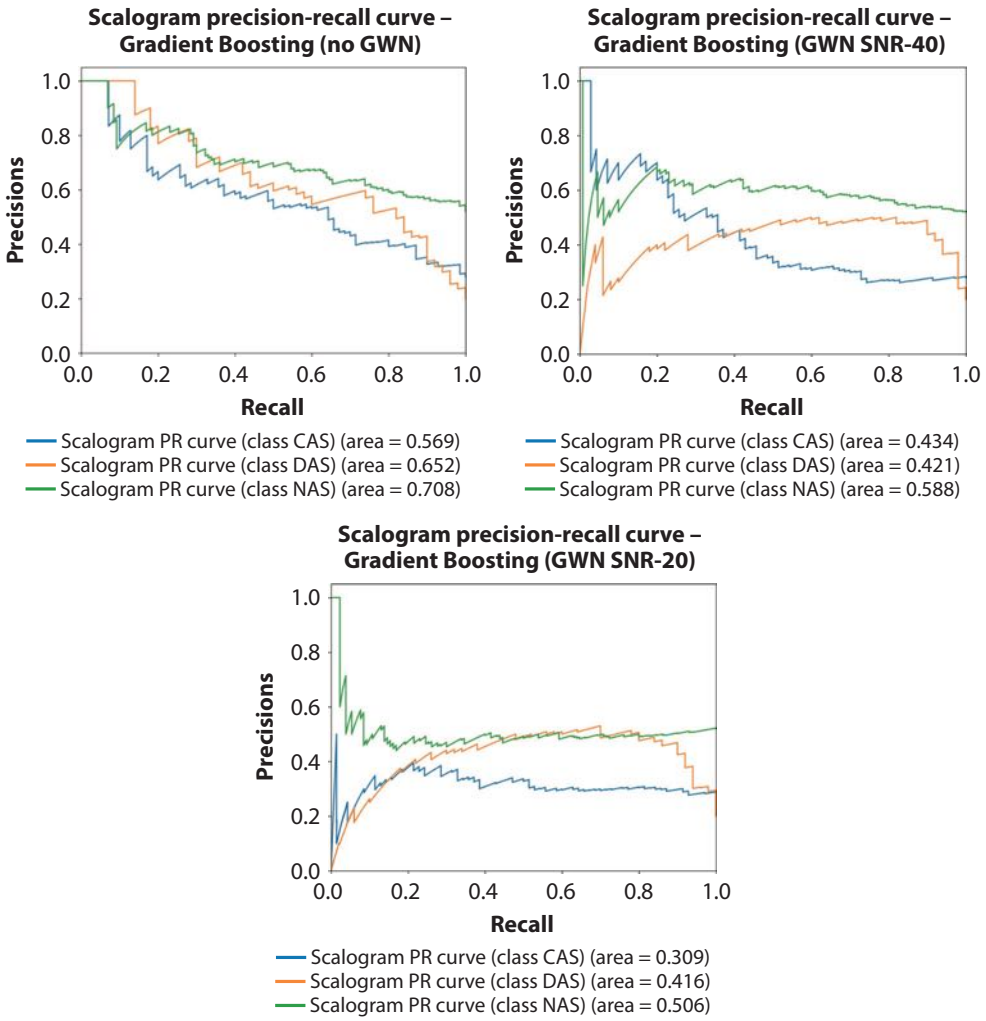


Fig. 5.2.47. *PR curve showing significant impact (P = 0.000) of GWN on Gradient Boosting model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

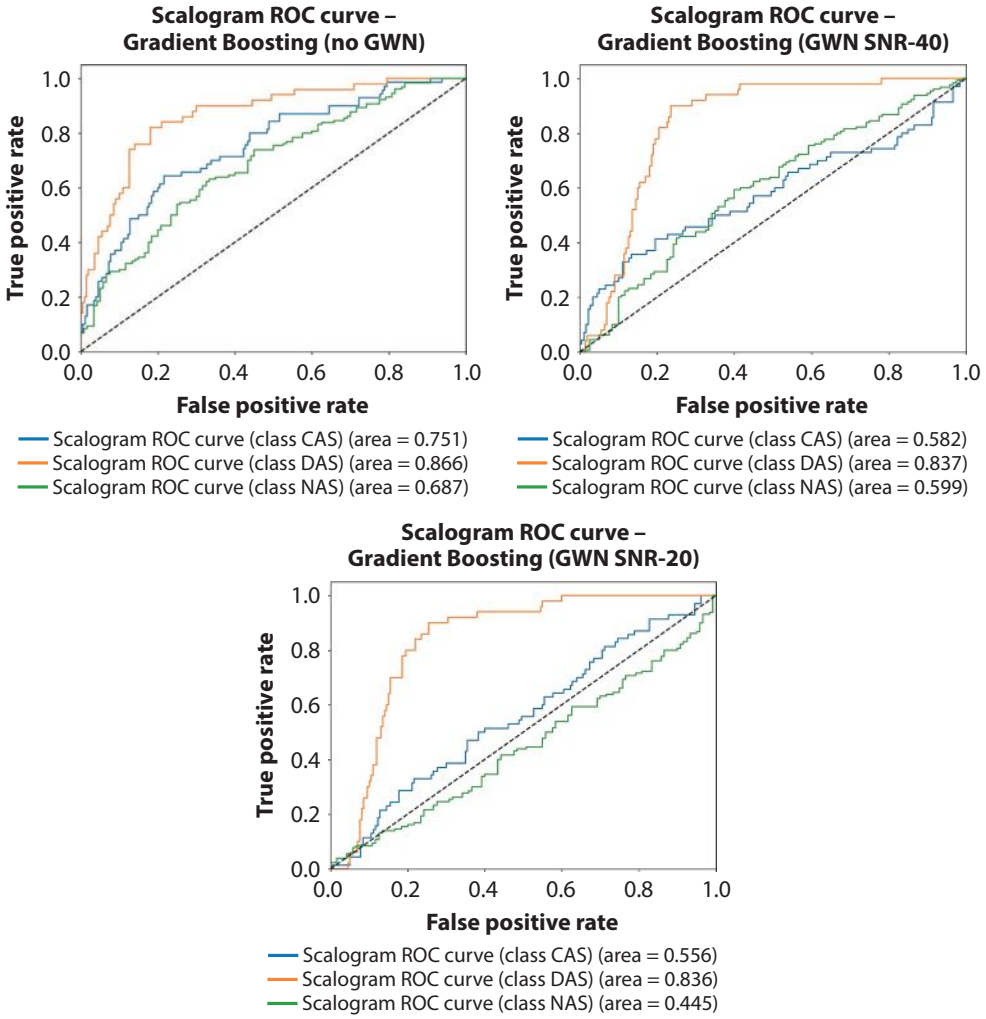


Fig. 5.2.48. *ROC curve showing significant impact (P = 0.000) of GWN on Gradient Boosting model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.48, the impact of Gaussian white noise (GWN) levels on the true positive rate (TPR) compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the spectrogram-based Gradient Boosting ML model. At no GWN level, the model shows a good performance for DAS and a reasonable performance for CAS sound classes identification, but only a very average performance for NAS class, as seen from the area under the ROC graph's curve (ROC-AUC) values. The model performance drops significantly at medium levels of GWN (SNR-40) for two sound groups: CAS and NAS. Finally, once GWN is increased to SNR-20, the performance of ML module to identify CAS, especially NAS, becomes abysmal, with DAS class identification showing robustness. The spectrogram-based Gradient Boosting shows good to reasonable performance at no GWN level, depending on the sound class. However, at GWN SNR-40 level, the TPR, compared to FPR, dropped off for CAS and NAS identification quite significantly and it worsened at SNR-20 level. Therefore, even though the Gradient Boosting ML model shows average performance without ambient noise added, the performance varies between sound classes, and this model lacks robustness even at medium GWN levels.

From Fig. 5.2.49, the impact of Gaussian white noise (GWN) is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the scalogram-based Histgradient ML model's confusion matrix. At no GWN added levels, 34/70 CAS was identified correctly, with only a paltry 26/50 DAS class and 88/130 of NAS identified. The confusion matrix at SNR-40 (medium GWN levels) shows a decrease in the model's performance with a drop in TP for CAS and DAS classes, only 25/70 and 12/50 correctly identified for both. The DAS sound class identification showed robustness with a score of 41/50. The increase in the GWN level to SNR-20 worsens the model's performance in discriminating between the three sound classes, with DAS and NAS identification standing at 5/50 and 14/130, respectively. The DAS class at the highest GWN was identified at 62/70 score. This showed that the scalogram-based Histgradient ML model had a reasonable performance at no GWN, but then, once GWN was added, the model's power to distinguish between sound groups was completely lost at the model classified all three classes mainly, falsely. This model lacks any robustness to GWN.

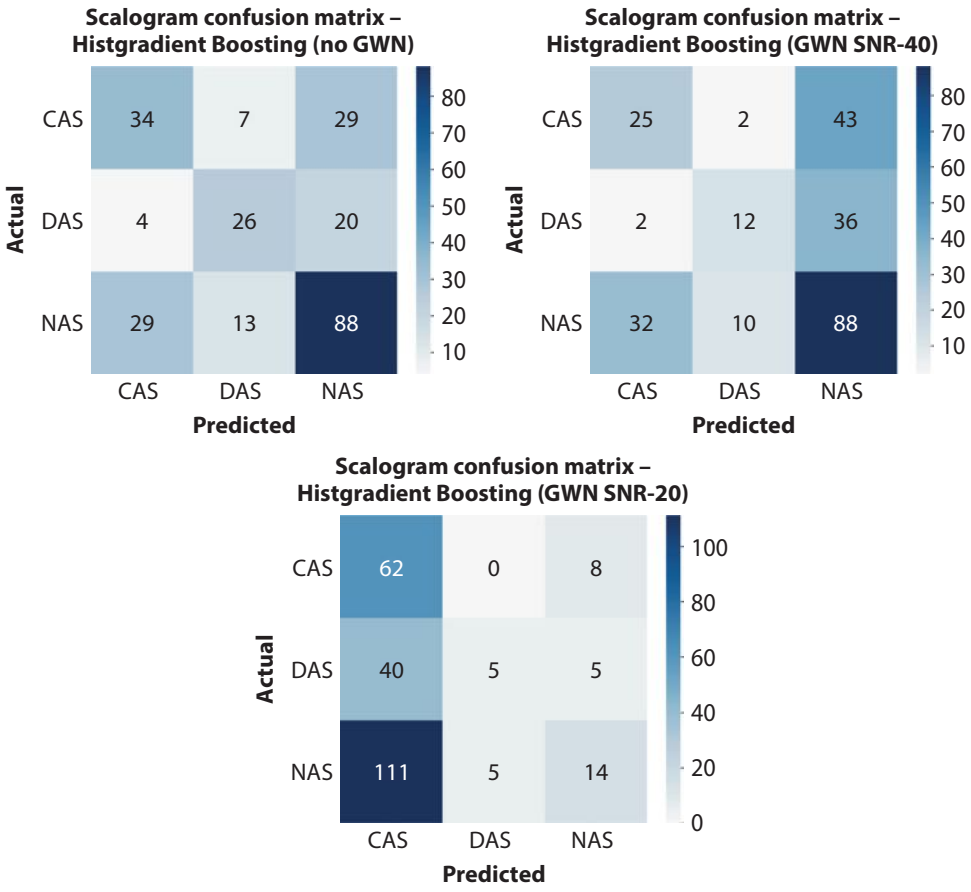


Fig. 5.2.49. *Confusion metrics showing significant impact (P = 0.000) of GWN on Histgradient model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.50, it can be observed that scalogram-based Histgradient ML model achieved an overall reasonable performance for precision to recall (PR) for NAS and DAS, with the worst performance observed for CAS as exemplified via area under the curve of PR graph (PR-AUC). At medium levels of GWN (SNR-40), all three sound classes are impacted more or less equally negatively. Finally, once GWN is increased to SNR-20 level, we see a significantly negative impact on CAS classes' identification with a lesser impact on NAS class. The DAS class recognition shows some robustness from GWN SNR-40 to SNR-20 levels, but the PR-AUC values are quite poor.

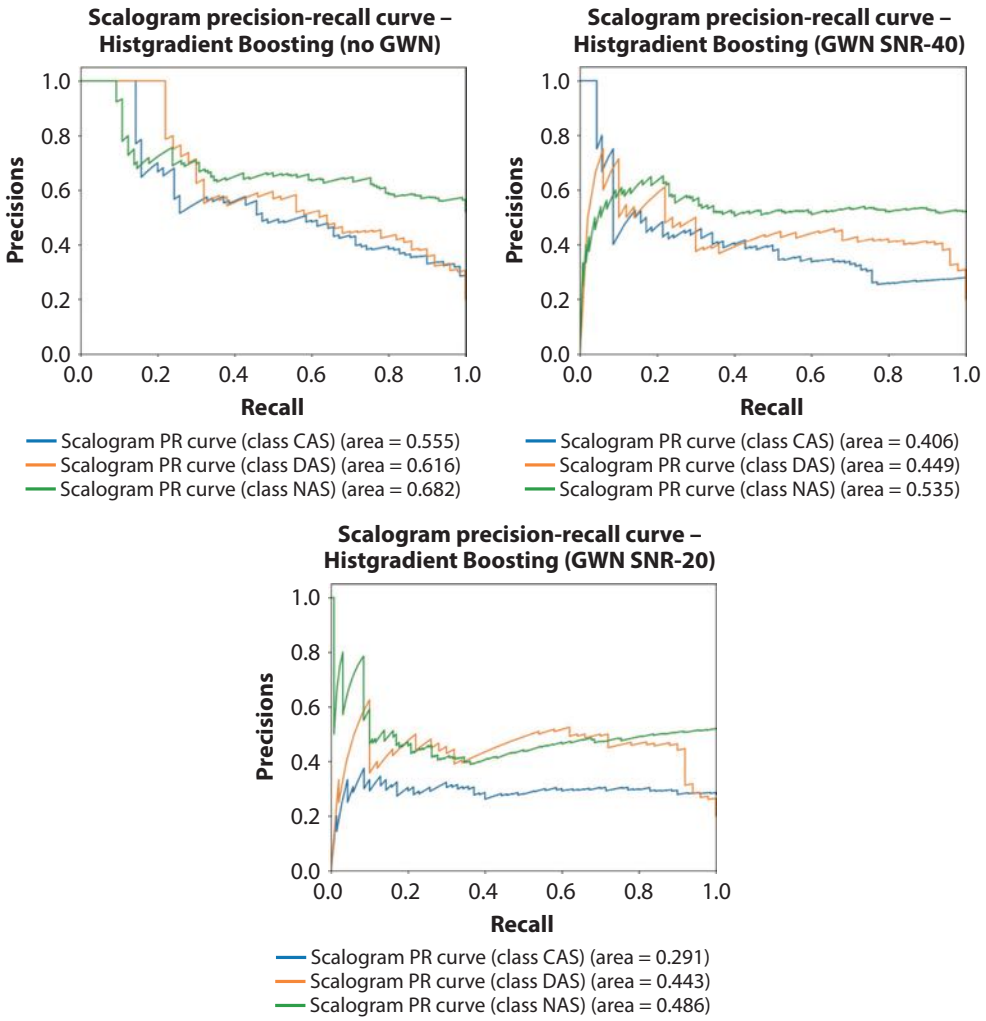


Fig. 5.2.50. *PR curve showing significant impact (P = 0.000) of GWN on Histgradient model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

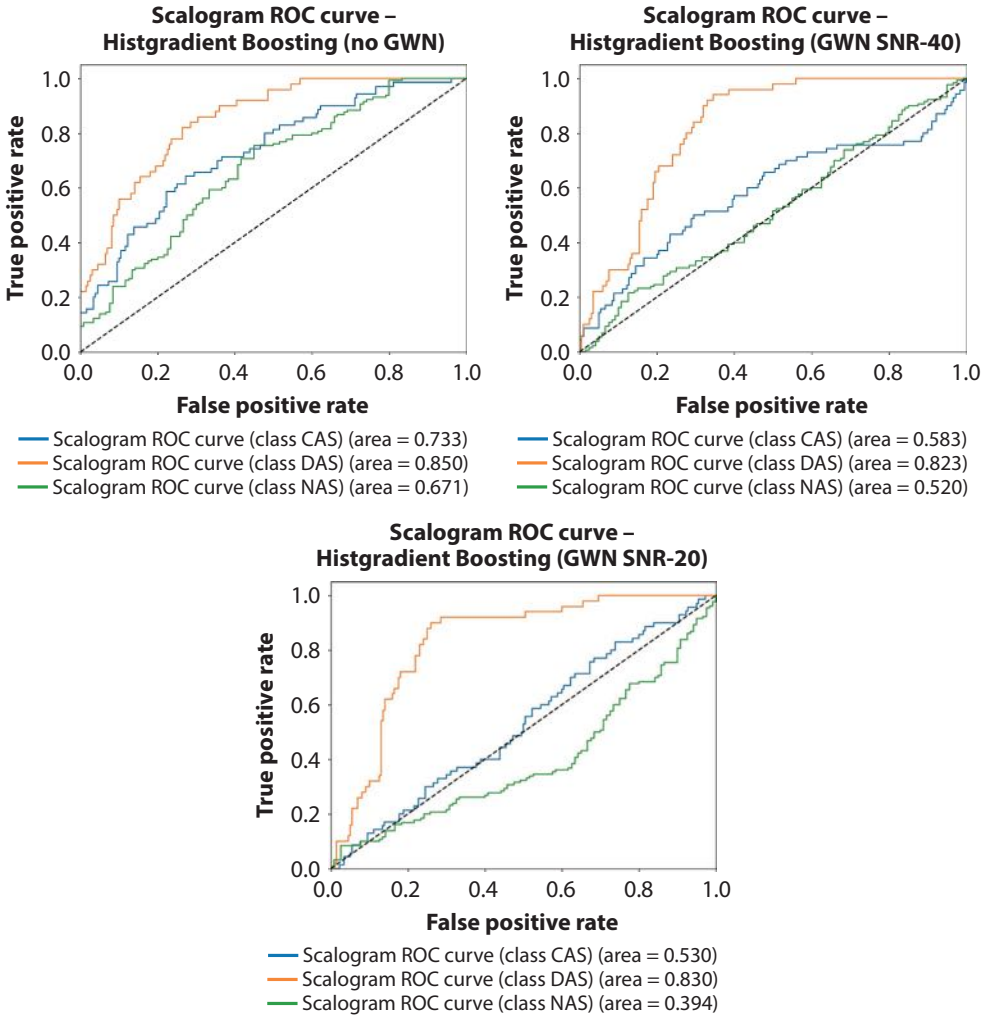


Fig. 5.2.51. *ROC curve showing significant impact (P = 0.000) of GWN on Histgradient model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.51, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the scalogram-based Histgradient ML model. At no GWN level, the model shows a good performance for DAS sound class recognition and a reasonable performance for CAS class, but only a very average performance for NAS class, as seen from the area under the ROC graph (ROC-AUC) curve. The model performance drops

significantly at medium levels of GWN (SNR-40) for two sound class groups: CAS and NAS. Finally, once GWN at SNR-20 is added, the model's diagnostic accuracy performance of CAS, especially NAS, classes becomes abysmal, with DAS class identification showing robustness. Depending on the class, the scalogram-based Histgradient model shows good to reasonable performance at no GWN levels. However, at GWN SNR-40 level, the TPR drop off for CAS and NAS identification significantly and worsens at SNR-20 level. Therefore, even though the model shows on average performance without ambient noise added, the performance varies between classes and lacks robustness even at medium GWN levels.

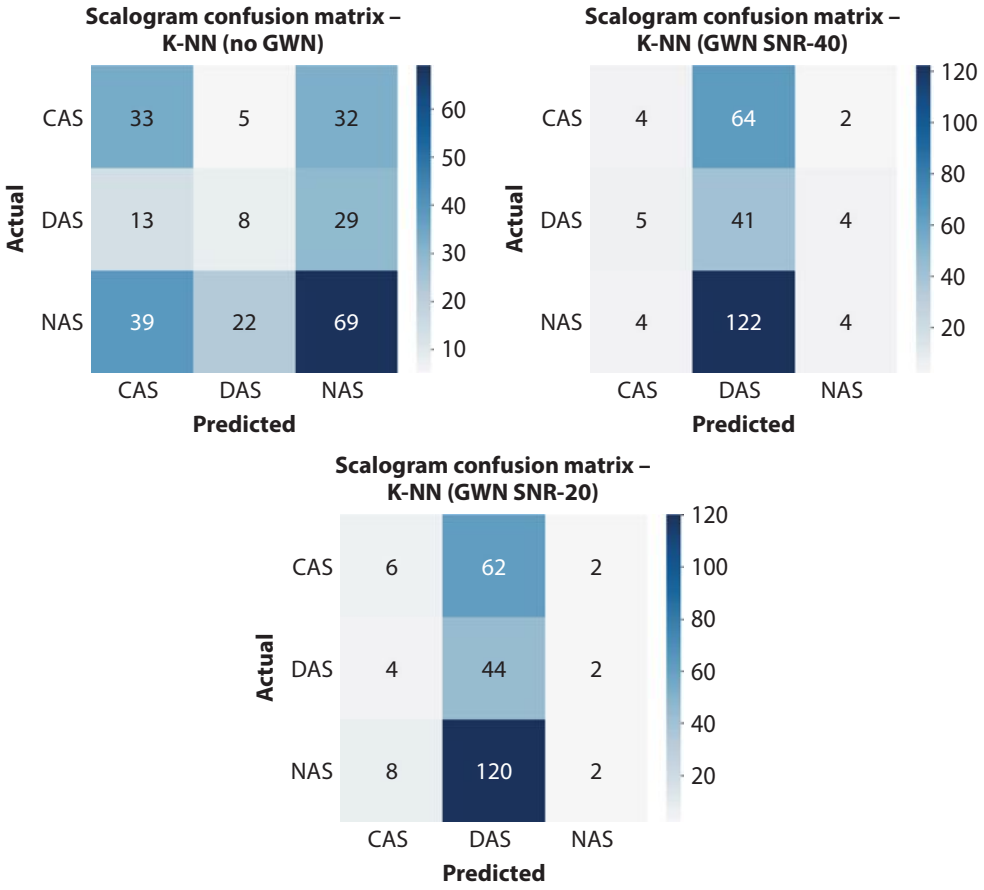


Fig. 5.2.52. *Confusion metrics showing significant impact (P = 0.000) of GWN on K-NN model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, K-NN – K-Nearest Neighbors.

From Fig. 5.2.52, the impact of Gaussian white noise is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the scalogram-based K-Nearest Neighbors (K-NN) ML model's confusion matrix. At no GWN level, 33/70 CAS was identified correctly, with only a paltry 8/50 of DAS class correctly identified, and 69/130 of NAS correctly identified. The confusion matrix at SNR-40 (medium GWN level) shows a decrease in the model's performance with a drop in TP for CAS and NAS classes, with only 4/70 and 4/130, respectively, correctly identified for both. Yet, the DAS class identification showed robustness with a score of 41/50. The increase in the GWN levels to SNR-20 more or less maintained this contrast between class identification with CAS and NAS identification standing at 6/70 and 2/130, respectively. The DAS class at the highest GWN was identified at 62/70 score. This showed that the scalogram-based K-NN model had a poor TP performance at no GWN, but then once GWN was added, the model's power to distinguish between classes was utterly lost as the model classified all three sound classes mostly belonging to DAS class.

From Fig. 5.2.53, it can be observed that scalogram-based K-Nearest Neighbors (K-NN) model achieved an overall poor performance of precision to recall (PR) for NAS, CAS, with the worst performance observed for DAS sound class, as exemplified via area under the curve (AUC) for PR graph (PR-AUC). Medium levels of GWN (SNR-40) negatively affected all three sound classes' recognition, as seen in the drop of the PR curve lines; this is especially true for DAS and CAS classes, with only a slight drop being observed in the NAS class. Finally, once GWN was increased to SNR-20 level, an extremely negative its impact on all three sound classes' identification is observed. This is particularly true for the DAS class. The overall performance of the K-NN model from no GWN to highest levels of GWN (SNR-20) is very poor.

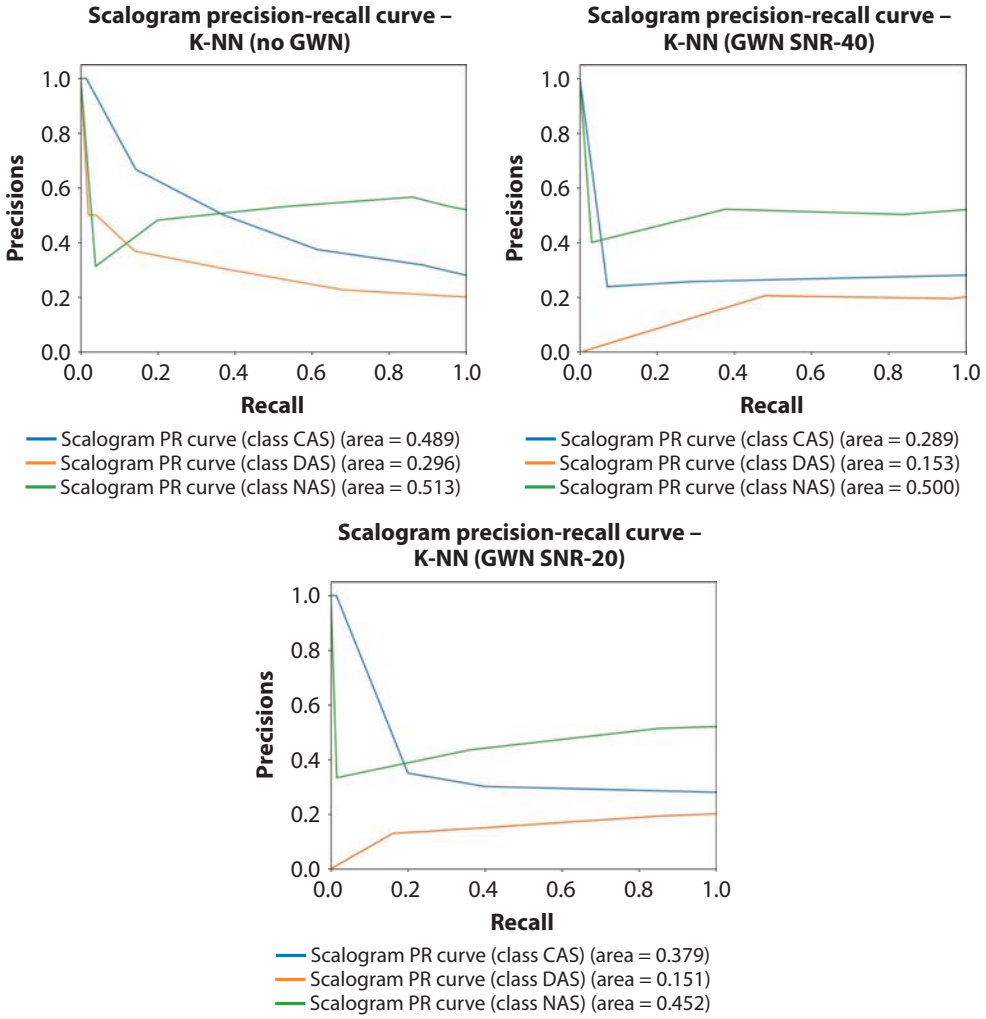


Fig. 5.2.53. *PR curve showing significant impact (P = 0.000) of GWN on K-NN model's ability to identify lung sounds correctly (from top to bottom).*

Note this model's performance was extremely poor. GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound, K-NN – K-Nearest Neighbors.

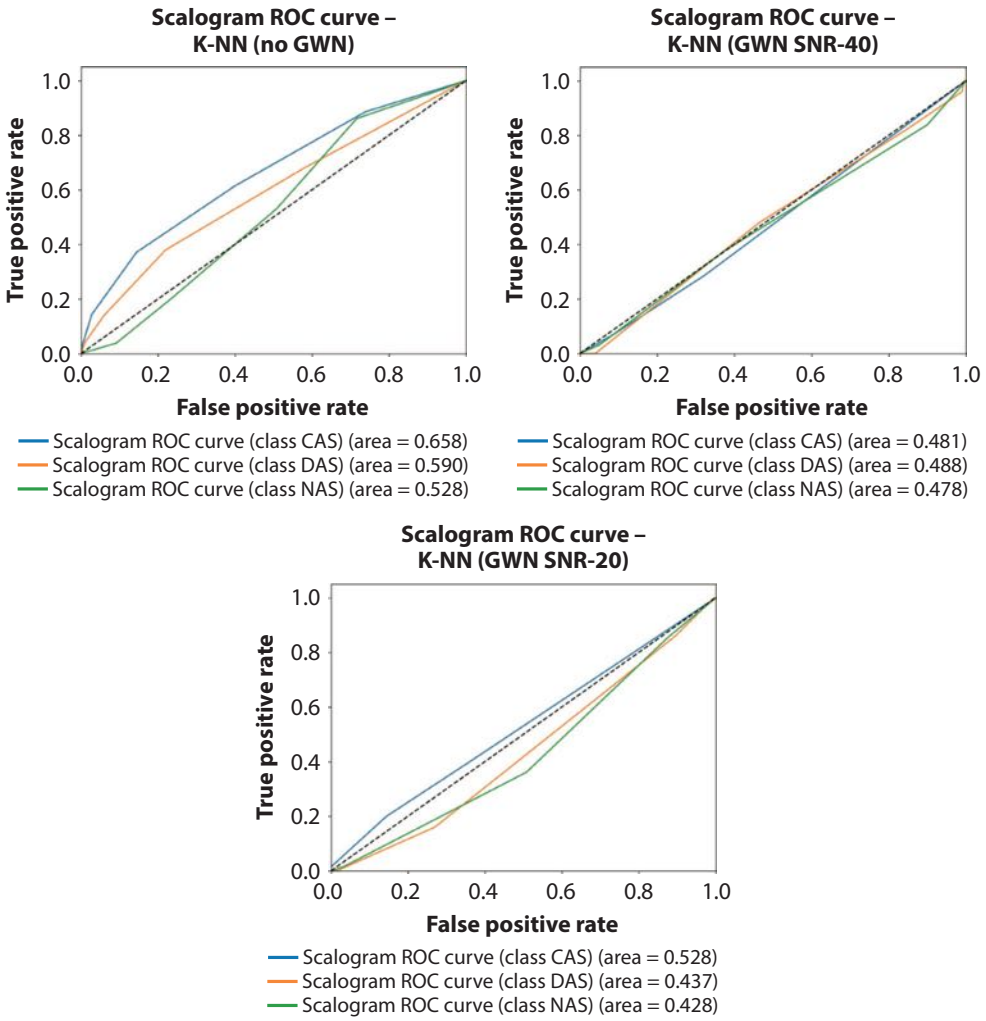


Fig. 5.2.54. *ROC curve showing significant impact (P = 0.000) of GWN on K-NN model's ability to identify lung sounds correctly (from top to bottom)*

Note this model's performance was extremely poor. GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.54, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve can be observed for the scalogram-based K-Nearest Neighbors (K-NN) model. At no GWN added level, the model shows only reasonable performance for the CAS class identification, whilst DAS and NAS classes recognition show poor performance of this ML model, as

seen from the area under the curve of the ROC graph (ROC-AUC). Additionally, the model performance drops significantly at medium levels of GWN (SNR-40) for all three sound groups identification. Finally, once GWN at SNR-20 is added to all three sound classes, the scalogram-based K-NN model's performance worsen for all three classes. Therefore, this model performs poorly overall without GWN added and lacks robustness even at medium ambient noise levels, as seen with TPR and ROC-AUC value drop.

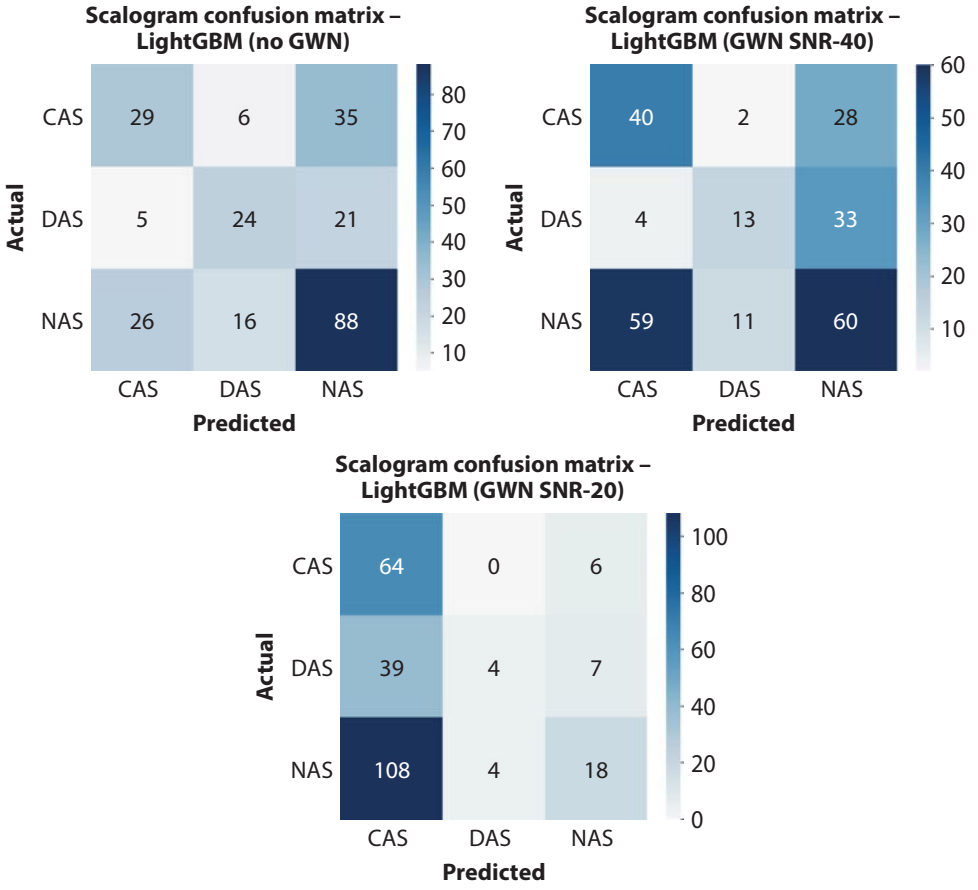


Fig. 5.2.55. *Confusion metrics showing significant impact (P = 0.000) of GWN on LightGBM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, LightGBM – Light Gradient Boosting Machine.

From Fig. 5.2.55, the impact of Gaussian white noise is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the scalogram-based Light Gradient Boosting Machine (LightGBM) model's confusion matrix. At no GWN level, 29/70 CAS was identified correctly, with 24/50 of DAS class correctly identified and 88/130 of NAS correctly identified. The confusion matrix at SNR-40 (medium GWN levels) shows a decrease in the model's performance with a drop in TP for DAS and NAS classes, with only 13/50 and 60/130, respectively, correctly identified for both. The CAS identification showed robustness with a value of 40/70. The increase in the GWN level to SNR-20 exaggerated contrasts between class identification with DAS and NAS standing at only 4/50 and 18/130, respectively. This showed that the scalogram-based LightGBM model had a reasonable performance at no GWN, but once GWN was added first at SNR-40, and later at SNR-20, the model lost its power to distinguish between TP and FP classes and with a tendency to classify all classes as CAS.

From Fig. 5.2.56, it can be observed that the scalogram-based Light Gradient Boosting Machine (LightGBM) model achieved an overall reasonable performance of precision to recall (PR) for NAS, DAS with slightly worse performance is observed for CAS as exemplified via the PR graph's area under the curve (PR-AUC). At medium levels of GWN (SNR-40) on all three sound classes, a drop of precision to recall is observed mainly for CAS and DAS classes, with the highest score at the medium level being scored by the NAS class. Finally, once GWN is increased to SNR-20 level, we see an extremely negative impact on all three sound classes' recognition, especially CAS, performing the worst. Therefore, even though the scalogram-based LightGBM model achieves reasonable PR performance for no GWN level, the model lacks robustness even at medium levels of GWN.

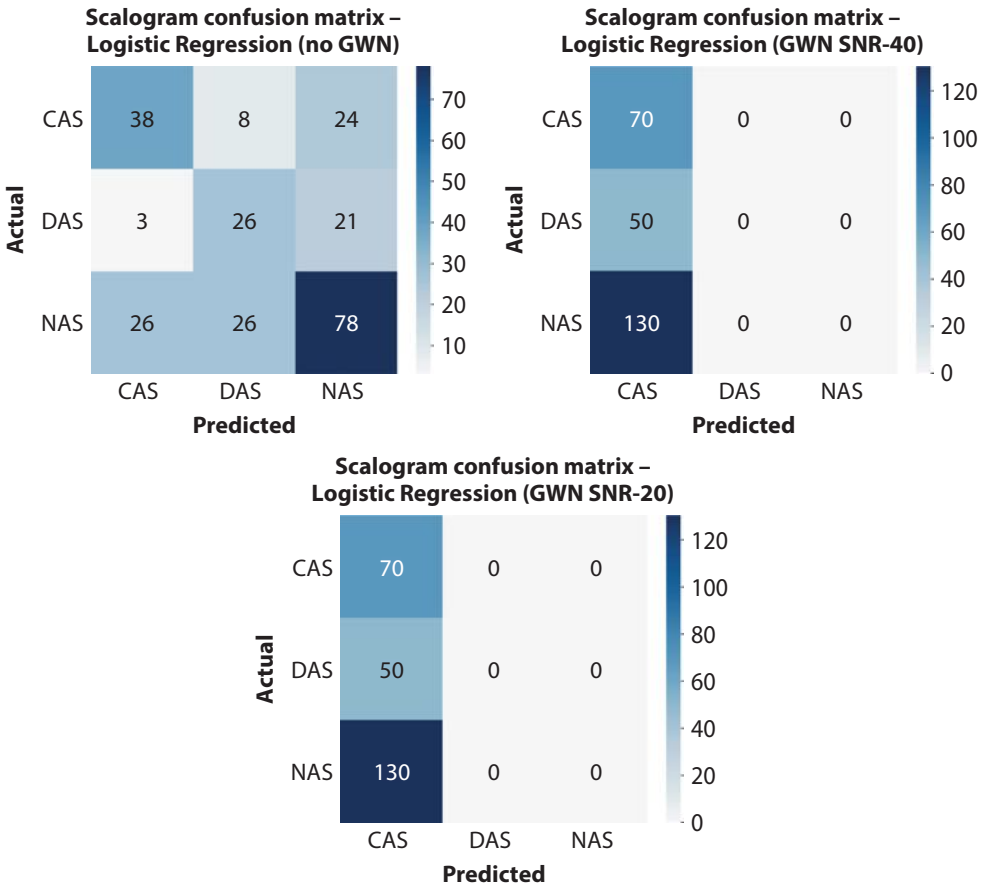


Fig. 5.2.56. *PR curve showing significant impact (P = 0.000) of GWN on LightGBM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, LightGBM – Light Gradient Boosting Machine.

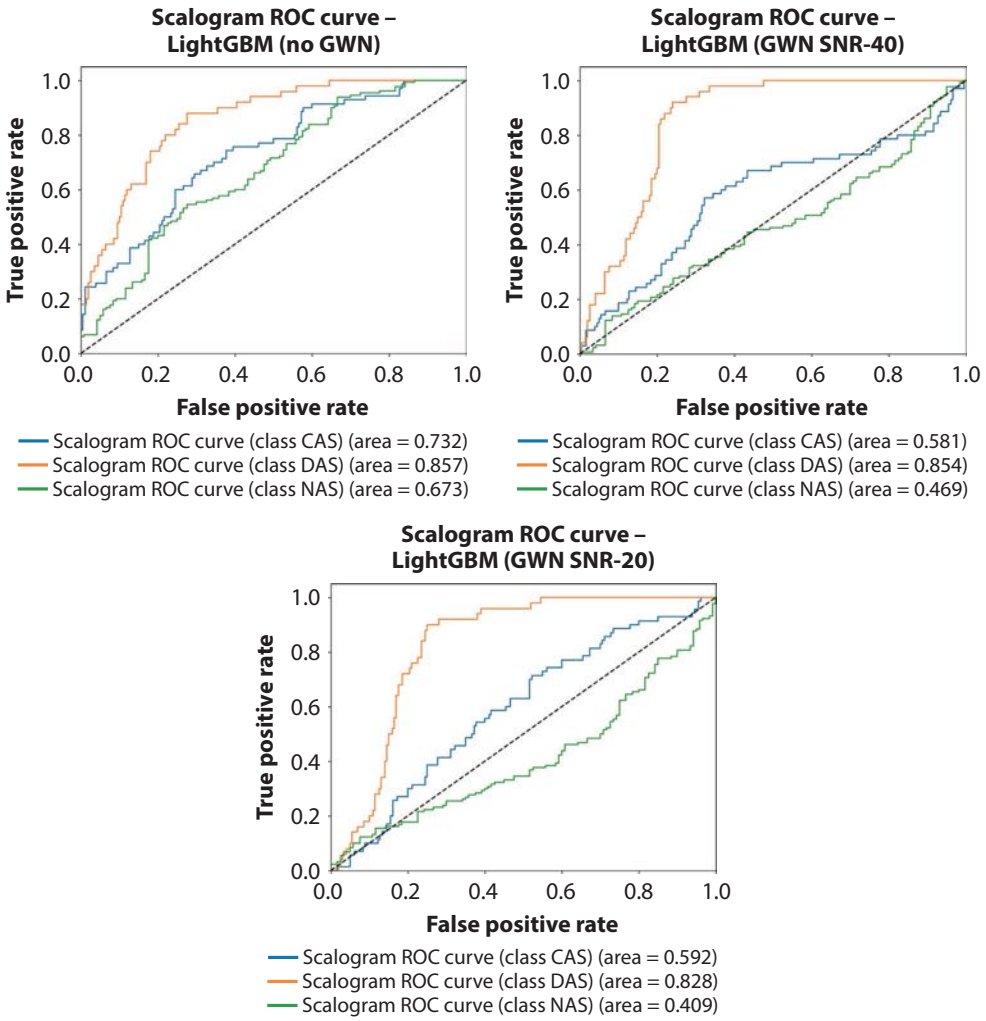


Fig. 5.2.57. *ROC curve showing significant impact (P = 0.000) of GWN on LightGBM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, LightGBM – Light Gradient Boosting Machine.

From Fig. 5.2.57, the impact of GWN levels on the true positive rate (TRP) as compared to the false positive rate (FPR) in the ROC curve can be observed for the scalogram-based Light Gradient Boosting Machine (LightGBM) model. At no GWN level, the model shows a very good performance from DAS, good performance for CAS and poor performance at NAS recognition, as seen from ROC-AUC (area) values. The model performance drops significantly at medium levels of GWN (SNR-40) for two sound groups: CAS and NAS. The drop is very slight for the DAS class. Finally, once GWN at SNR-20 is added, the LightGBM TRP rate drops for NAS, whilst DAS performance continues strong and CAS, though much weaker performance, is not impacted by increased levels of GWN as compared to medium GWN levels. Through DAS sound class recognition showed resilience and maintained very good TPR. However, the CAS shows poor performance, and it is even worse for NAS once GWN is increased to SNR-40 levels. Therefore, the scalogram-based LightGBM model has reasonable performance at no GWN levels, with caveat variability depending on the group. Additionally, the performance becomes very poor even at medium levels of GWN for CAS and NAS classes. Therefore, the model lacks robustness.

From Fig. 5.2.58, the impact of Gaussian white noise (GWN) is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the scalogram-based Logistic Regression ML model's confusion matrix. At no GWN level, 38/70 CAS was identified correctly, with 26/50 of DAS class correctly identified and 78/130 of NAS correctly identified. The confusion matrix at SNR-40 (medium GWN level) shows an extremely sharp decrease in model performance with a drop in TP for DAS and NAS class, with zero correctly identified for both, but a perfect score for TP for NAS and 130/130. With the increase in the GWN levels to SNR-20, the exact same score was maintained for all three classes. This showed that the scalogram-based Logistic Regression ML model had a good performance at no GWN, but even at medium levels of GWN (SNR-40), had an extremely sharp drop in the diagnostic ability for CAS and NAS sounds with a strong bias towards falsely diagnosing these sounds as CAS. This shows the extremely poor robustness of this model to GWN ambient noise.

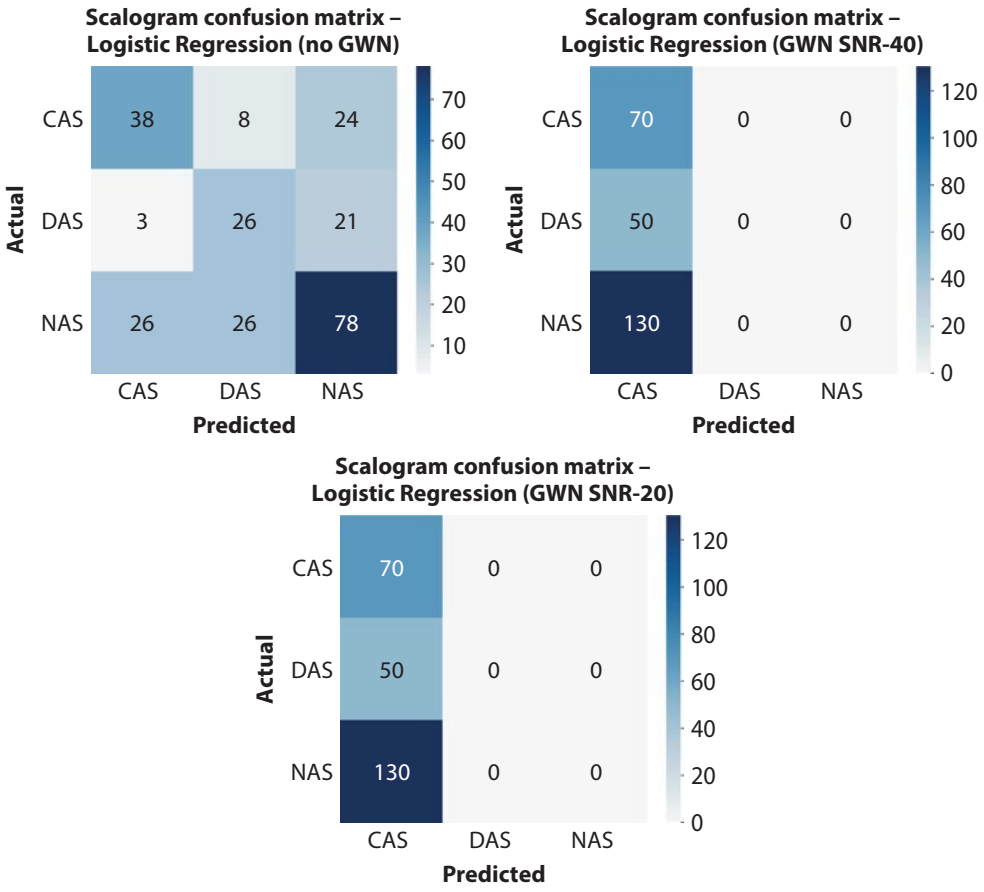


Fig. 5.2.58. *Confusion metrics showing significant impact (P = 0.000) of GWN on Logistic Regression model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.59, it can be observed that scalogram-based Logistic Regression ML model achieved an overall reasonable performance of precision to recall (PR) for NAS, CAS, with the worst performance observed for DAS as exemplified via area under the curve for PR (PR-AUC) graph. At medium levels of GWN (SNR-40) on all three sound classes, a drop of precision to recall is observed mainly for DAS, whilst NAS shows the greatest robustness to medium levels of GWN. Finally, once GWN was increased to SNR-20 level, an extremely negative impact on all three sound classes' recognition is observed. This is especially true for DAS class identification by the model. The scalogram-based LR ML model does not

seem to function properly, as three classes' lines are spread apart with overall very low precision.

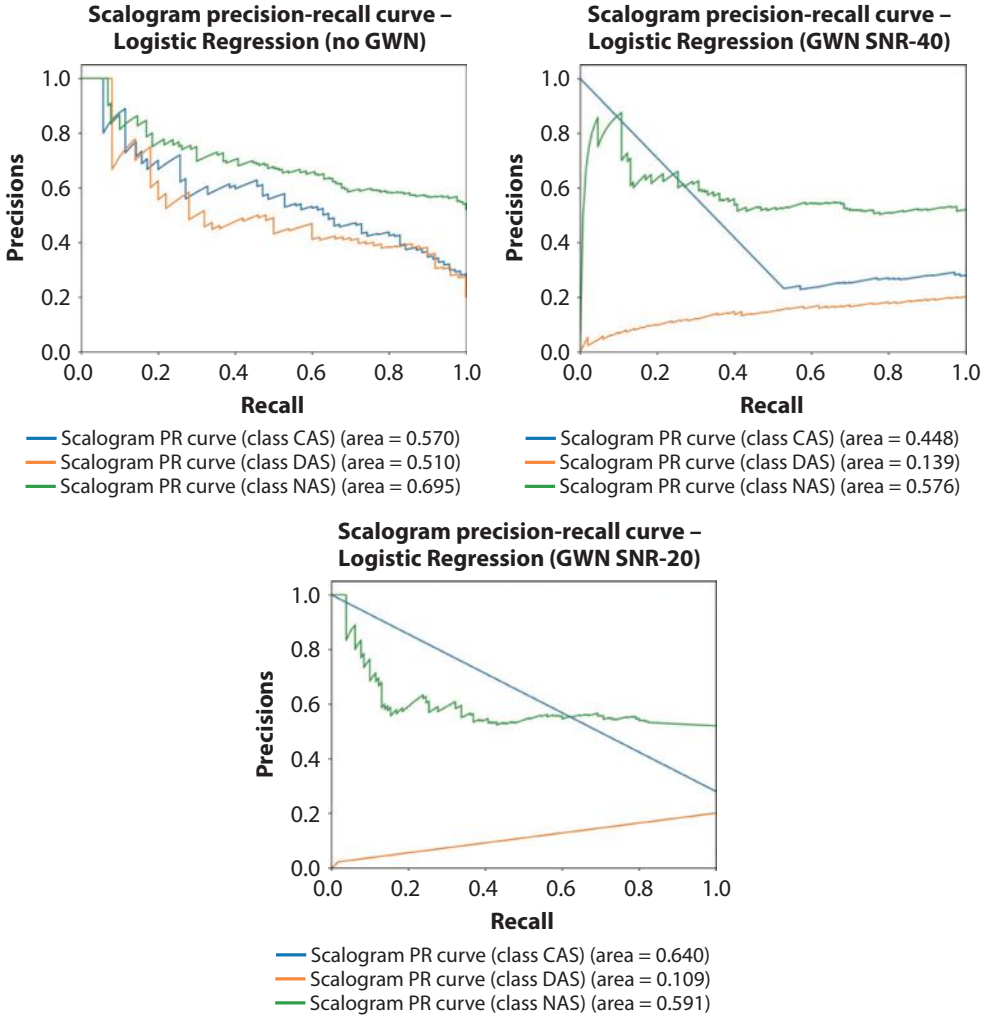


Fig. 5.2.59. *PR curve showing significant impact (P = 0.000) of GWN on Logistic Regression model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

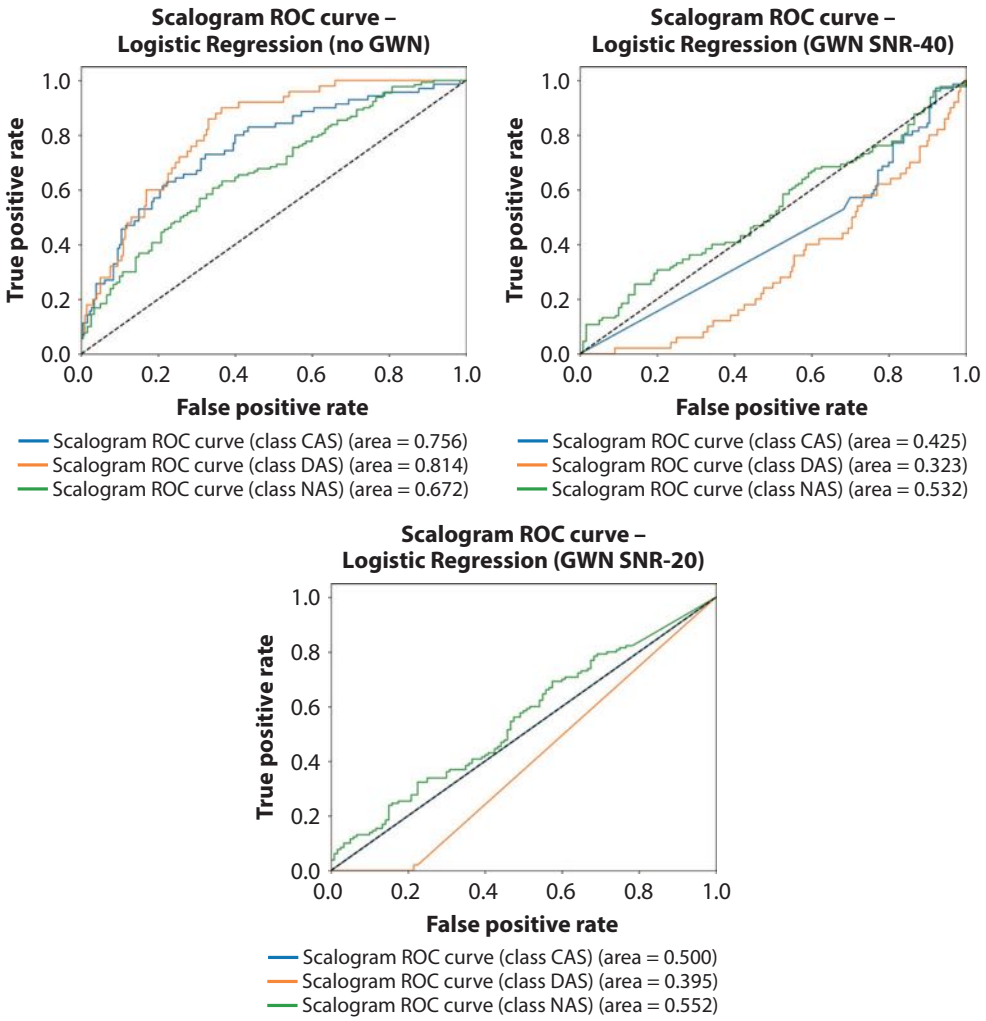


Fig. 5.2.60. *ROC curve showing significant impact (P = 0.000) of GWN on Logistic Regression model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.60, the scalogram-based Logistic Regression model can observe the impact of Gaussian white noise (GWN) levels on the true positive rate (TPR) compared to the false positive rate (FPR) in the receiver operating characteristic (ROC) curve. At no GWN added level, the model shows a very good performance for DAS, good performance for CAS and poorer performance for NAS sound class identification, as seen from the area under the curve of the ROC graph's (ROC-AUC) values. The model performance

drops significantly at medium levels of GWN (SNR-40) for all three classes. Finally, once GWN at SNR-20 level is added, the model's performance worsens for all three sound classes. Therefore, the scalogram-based Logistic Regression model shows diagnostic power only at no GWN level, but with the caveat of having variability between all three sound classes, the TPR drops off even at medium levels of GWN and gets worse at higher levels. Hence, the model lacks any robustness to GWN.

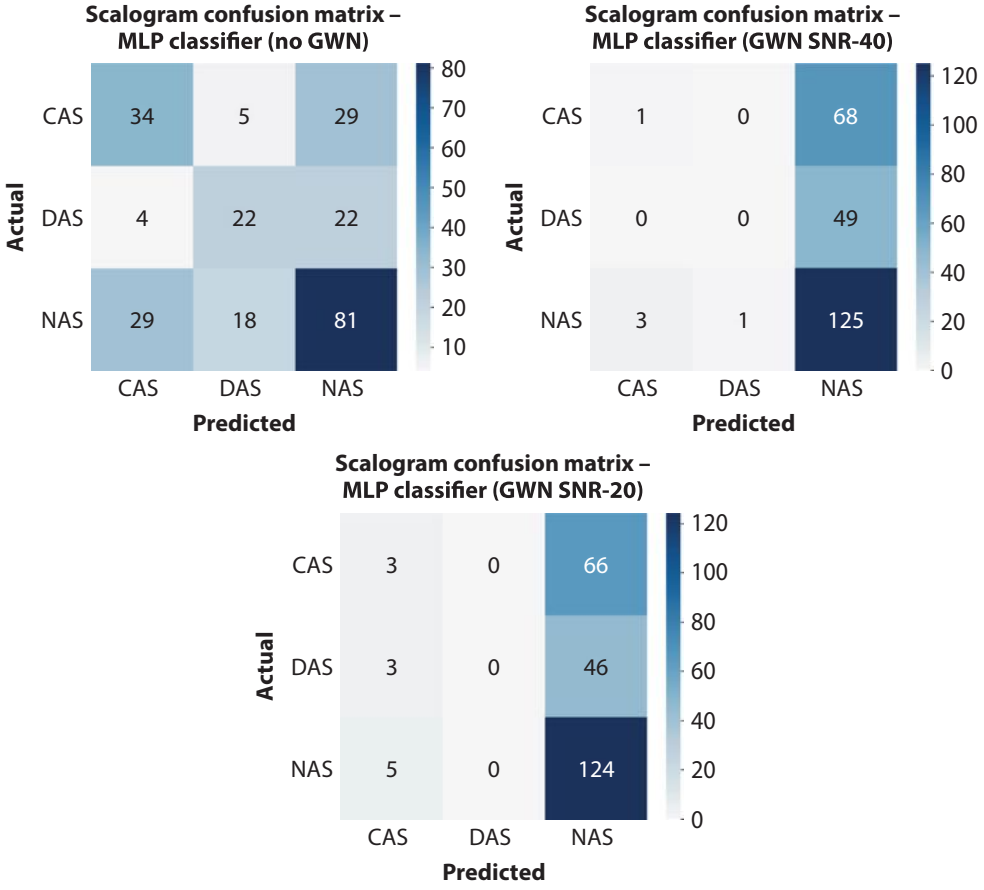


Fig. 5.2.61. *Confusion metrics showing significant impact (P = 0.000) of GWN on MLP model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, MLP classifier – Multilayer Perceptron.

From Fig. 5.2.61, the impact of Gaussian white noise (GWN) is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP), as seen in the scalogram-based Multilayer Perceptron (MLP) model's confusion matrix. At no GWN level, 34/70 CAS was identified correctly, 22/50 of the DAS class was correctly identified, and 81/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium GWN levels) shows a sharp decrease in the model performance with a drop in TP with only 1/70 correctly for the CAS class and zero correctly identified for DAS class, but a good score for TP for NAS and 125/130. The increase in the GWN levels to SNR-20 level continued DAS and CAS poor classification performance with 3/70 and 0/50, with only NAS scoring highly with 124/130. This showed that the scalogram-based MLP model had a good performance at no GWN level, but even at medium levels of GWN (SNR-40), it had a lack of diagnostic ability for CAS and DAS sounds with a strong bias towards falsely diagnosing these sounds as NAS.

From Fig. 5.2.62, it can be observed, that the scalogram-based Multilayer Perceptron (MLP) model achieved an overall reasonable to poor performance of precision to recall (PR) for NAS, CAS classes, with the worst performance observed for DAS sound class, as exemplified via area under the curve for PR graph (PR-AUC). At medium levels of GWN (SNR-40) on all three, a drop of precision to recall. The worst performance by the MLP model is observed for DAS class, whilst NAS class identification shows reasonable robustness to medium level of GWN. Finally, once GWN is increased to SNR-20 level, we see a significantly negative impact on all three sound classes' recognition, especially DAS and CAS identification. Therefore, at the higher GWN levels, the model stops being an effective diagnostic tool.

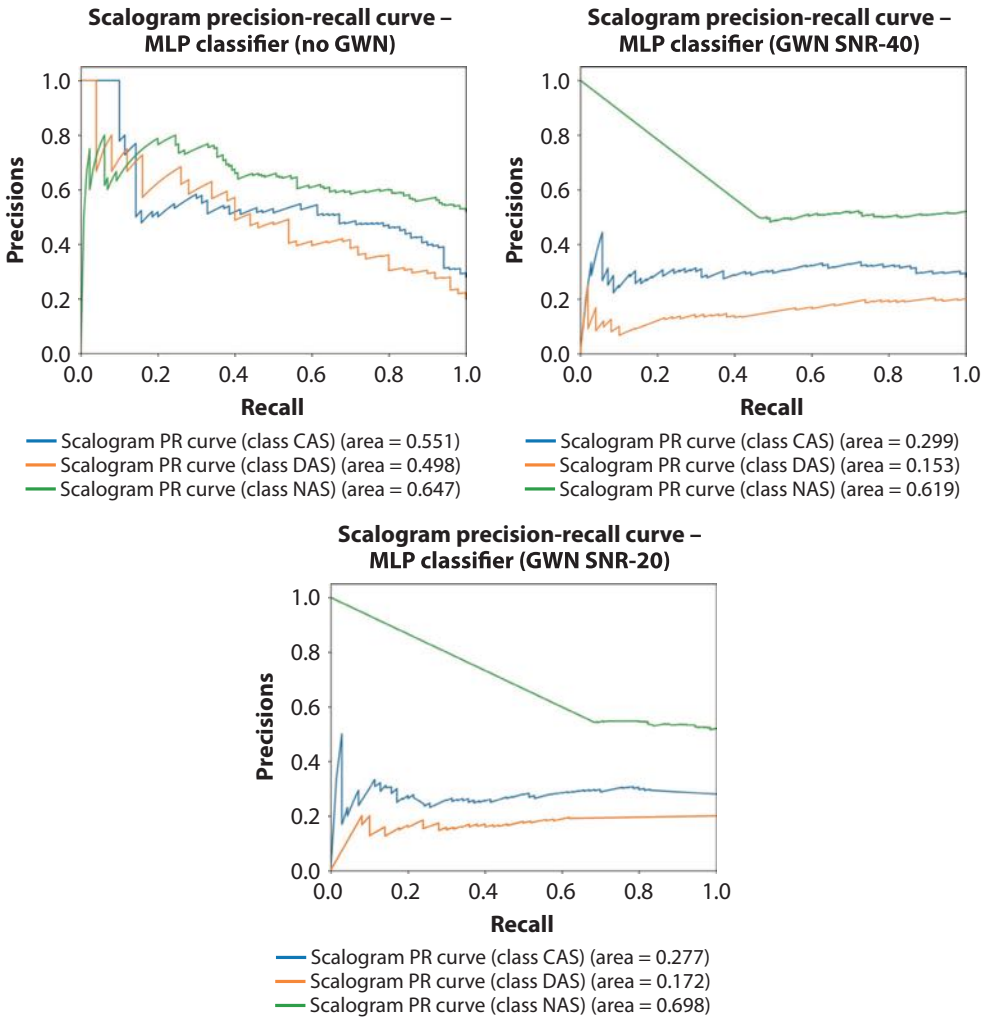


Fig. 5.2.62. *PR curve showing significant impact (P = 0.000) of GWN on MLP model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, MLP classifier – Multilayer Perceptron.

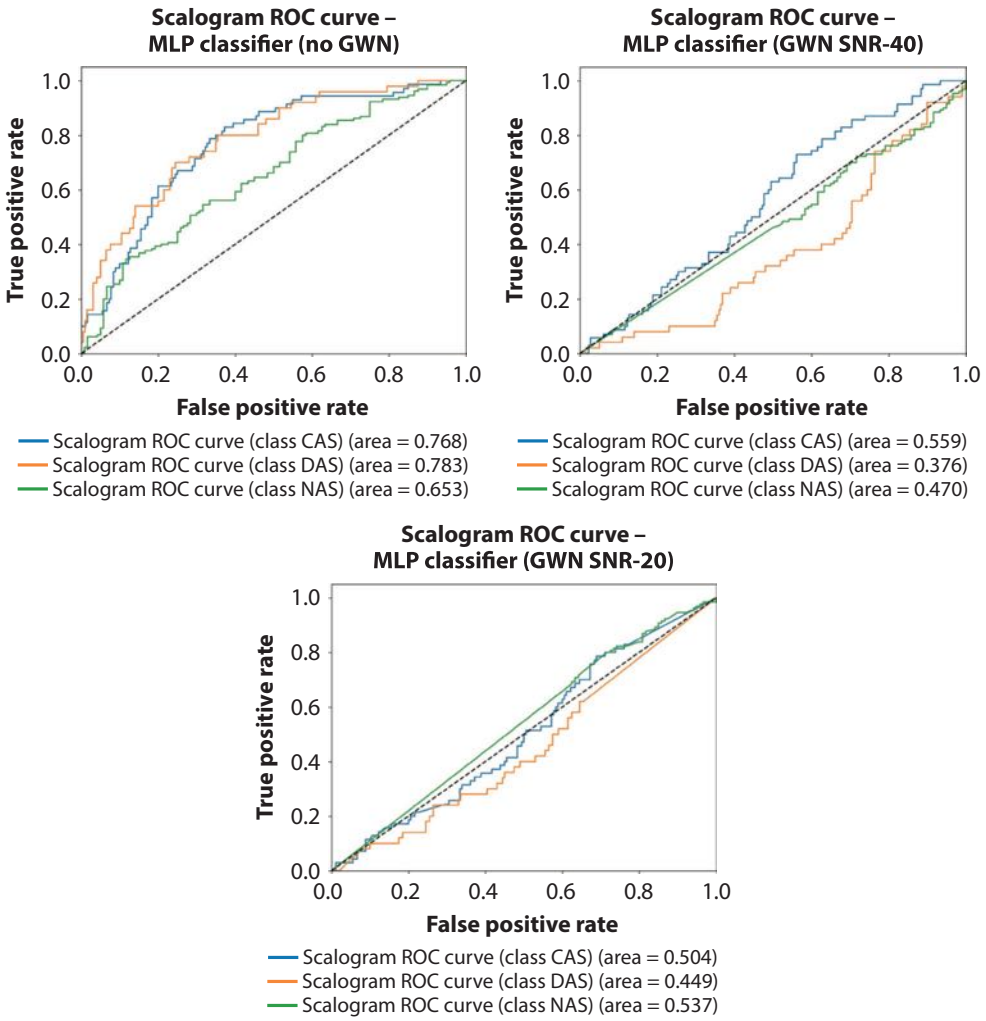


Fig. 5.2.63. *ROC curve showing significant impact (P = 0.000) of GWN on MLP model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS– continuous auscultated sound, DAS – discontinuous auscultated sound, MLP classifier – Multilayer Perceptron.

From Fig. 5.2.63, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate in the receiver operating characteristic (ROC) curve can be observed for the scalogram-based Multilayer Perceptron (MLP) model. At no GWN level, the model shows a good performance for DAS and CAS classes but poorer performance for NAS class recognition, as seen from the ROC graph's area under the curve (ROC-AUC)

values. The model performance drops significantly at medium level of GWN (SNR-40) for all three sound classes. Finally, once GWN at SNR-20 is added, the performance becomes even worse for all three sound classes. Therefore, the scalogram-based MLP model shows reasonable diagnostic power only at no GWN level. However, the TRP drops off even at medium GWN levels and worsens at higher levels. Hence, the model lacks any robustness to ambient noise.

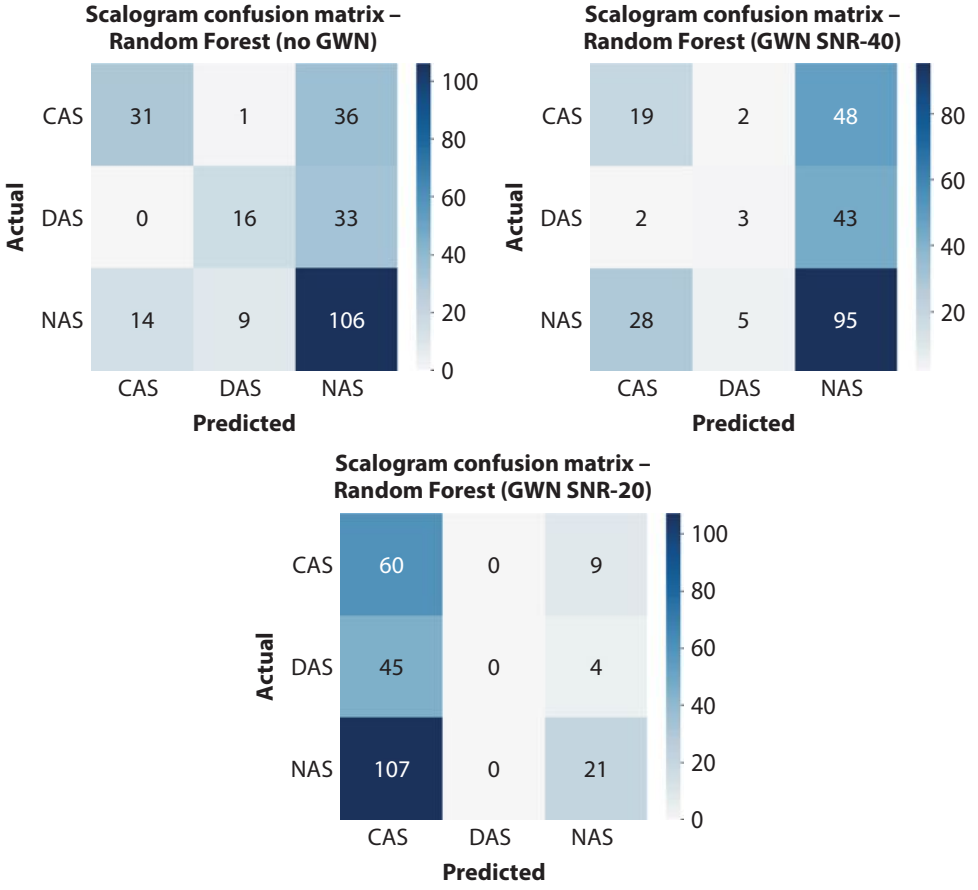


Fig. 5.2.64. *Confusion metrics showing significant impact (P = 0.000) of GWN on Random Forest model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.64, the impact of Gaussian white noise is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the scalogram-based Random Forest ML model's confusion matrix. At no GWN level, 31/70 CAS was identified correctly, 16/50 of the DAS class was correctly identified, and 106/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium GWN levels) shows an extremely sharp decrease of the model performance with a slight drop in TP for the NAS class with 95/130 correctly identified, whilst showing an extremely sharp drop in performance for DAS identification, with 3/50 score and CAS 19/70 score. The increase of the GWN levels to SNR-20 level shows a continued significant drop in the model's performance for DAS and NAS class sound identification with 0/50 and 21/130, respectively. Only the CAS class exhibiting a good TP score of 60/70. Overall, the scalogram-based Random Forest ML model showed very poor performance for DAS and NAS sound classes recognition and poor robustness to GWN. The model misdiagnosed DAS and CAS as NAS at GWN SNR-40. Whilst at GWN SNR-20, the bias turned towards misdiagnosing NAS and DAS as CAS classes. This showed the model's limited diagnostic ability, poor performance under GWN conditions, and unreliability as diagnostic tool.

From Fig. 5.2.65, it can be observed that the scalogram-based Random Forest model achieved an overall reasonable performance of precision to recall (PR) for NAS, CAS with slightly worse performance observed for DAS sound class identification, as exemplified via area under the curve for PR (PR-AUC) graph. At medium levels of GWN (SNR-40) on all three sound classes, a drop of precision to recall is observed mainly for DAS class, whilst NAS shows the greatest robustness to medium levels of GWN. Finally, once GWN is increased to SNR-20 (highest GWN level), we see a significantly negative impact on all three sound classes identification, especially DAS, whilst NAS continues to show the greatest robustness through all levels of GWN.

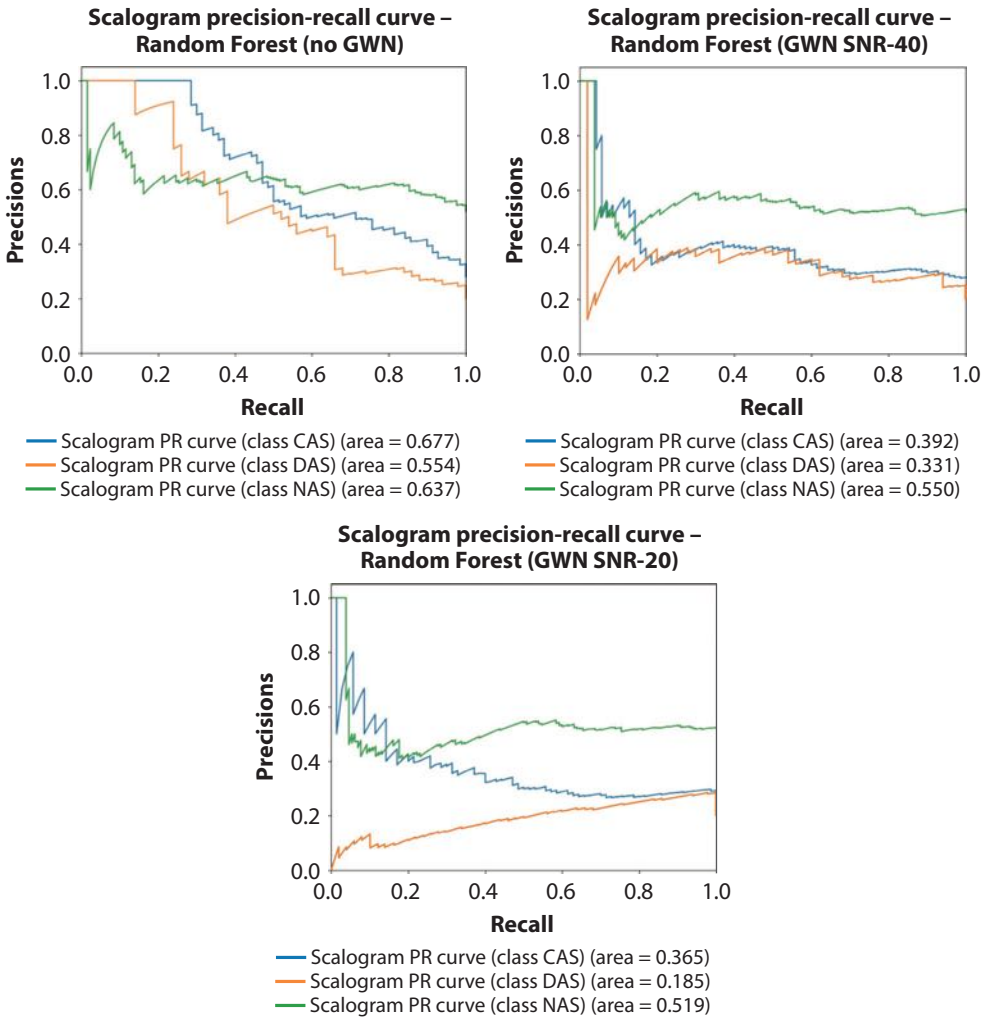


Fig. 5.2.65. *PR curve showing significant impact (P = 0.000) of GWN on Random Forest model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

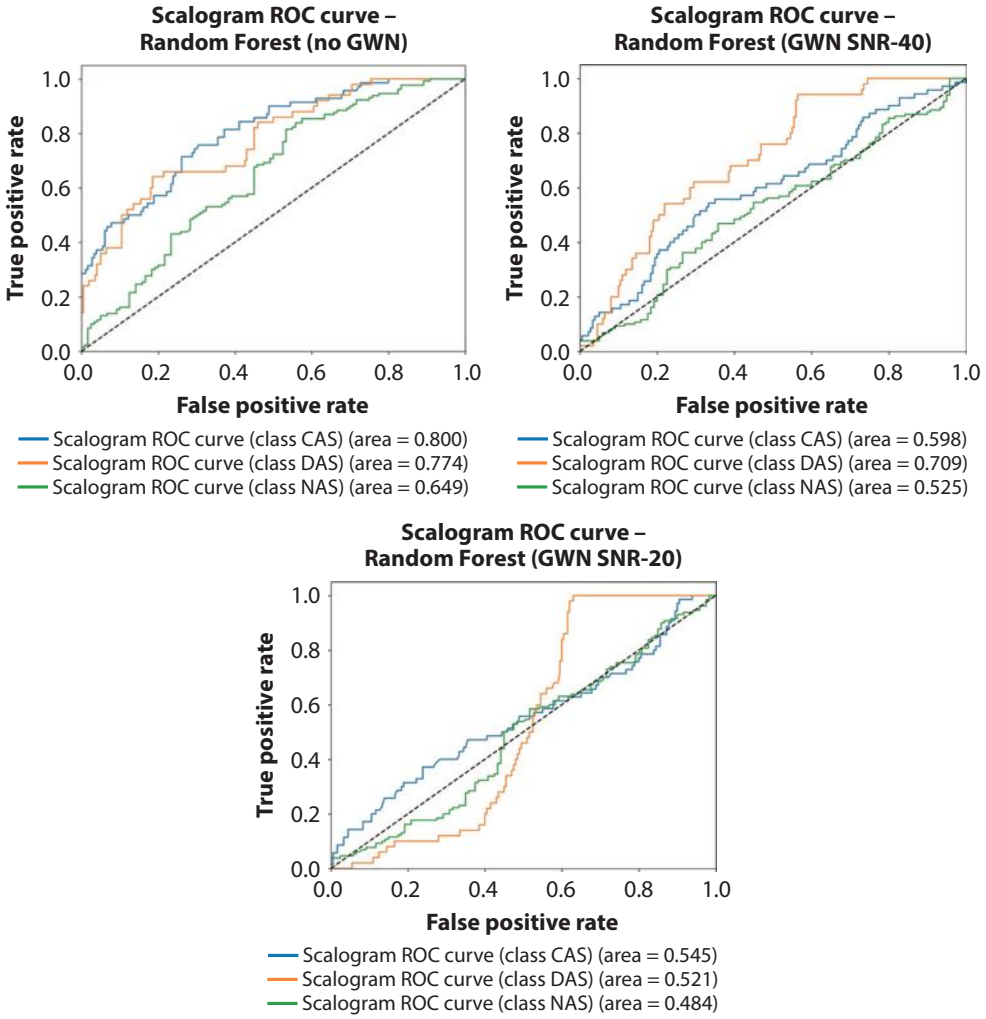


Fig. 5.2.66. *ROC curve showing significant impact (P = 0.000) of GWN on Random Forest model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound.

From Fig. 5.2.66, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive rate in the ROC curve can be observed for the scalogram-based Random Forest ML model. At no GWN level, the model shows a good performance for CAS class, above average for DAS class and poor performance for CAS class identification, as seen from ROC-AUC (area) values. The model performance drops significantly at medium levels of GWN (SNR-40) for all three sound classes, with some robustness

shown for the DAS class. Finally, once GWN at SNR-20 level is added, the Random Forest model's performance worsens for all three sound classes recognition. Therefore, the scalogram-based Random Forest model shows diagnostic reasonable power only at no GWN level, with some robustness shown at medium GWN levels, but only for the DAS sound class. Once the GWN is increased to the highest level, the model loses its ability to classify any of the three sound classes as it lacks robustness to the highest GWN levels.

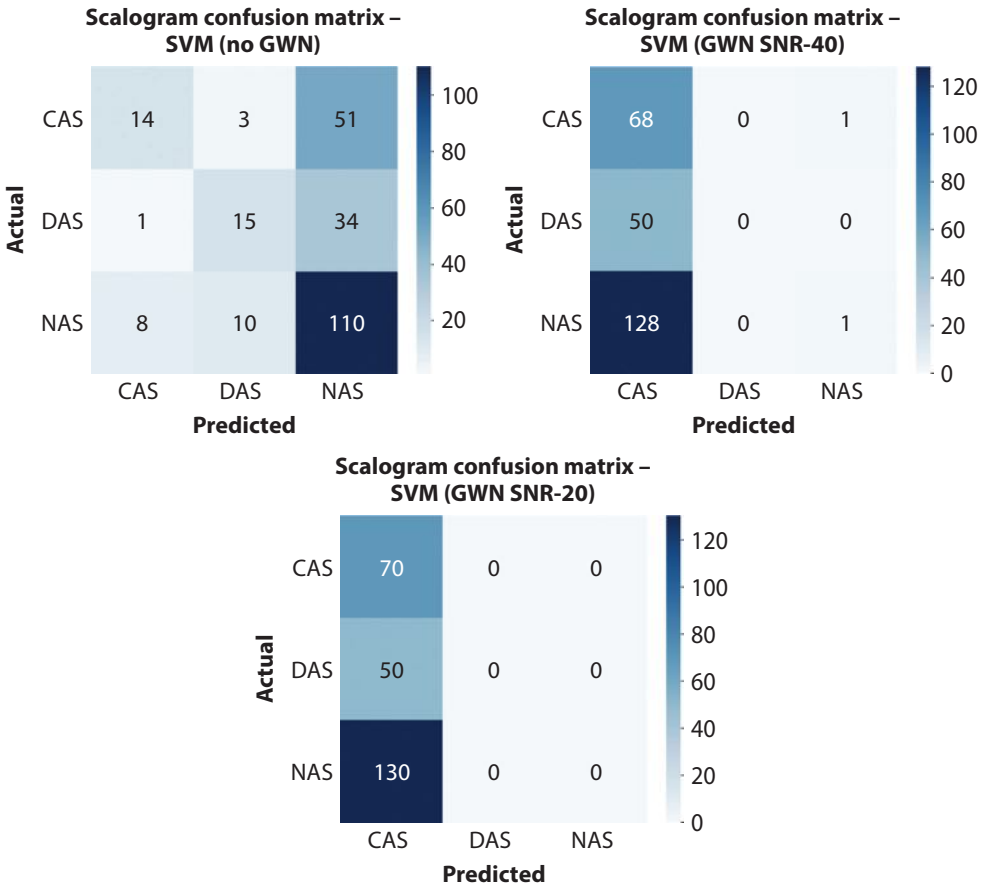


Fig. 5.2.67. *Confusion metrics showing significant impact (P = 0.000) of GWN on SVM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, SVM – Support Vector Machines.

From Fig. 5.2.67, the impact of Gaussian white noise is observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) scores in the scalogram-based SVM model's confusion matrix. At no GWN levels, 14/70 CAS was identified correctly, 15/50 of the DAS class was correctly identified, and 110/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium GWN level) shows a highly sharp decrease in the model's performance, the drop in TP for the NAS and DAS classes with extremely low with scores of 1/130 and 0/50, respectively. The increase in the GWN to SNR-20 level, shown a continued significant drop in the model's ability to identify DAS and NAS sound classes with 0/50 and 0/130, respectively, while only the CAS class maintained a TP score of 70/70. The SVM model overall showed abysmal performance at all three sound classes' identification, poor robustness to GWN levels and bias towards mislabelling sound classes as CAS at highest GWN levels.

From Fig. 5.2.68, it can be observed that scalogram-based Support Vector Machines (SVM) achieved an overall reasonable to poor performance of precision to recall (PR) for NAS, CAS identification with slightly worse performance is observed for DAS as exemplified by the area under the curve (AUC) for PR graph (PR-AUC). At medium levels of GWN (SNR-40) on all three classes precision to recall is reduced, especially for DAS class identification, whilst NAS shows the greatest robustness to medium levels of GWN. Finally, once GWN is increased to SNR-20 levels, we see a significantly negative impact on SVM models' ability to identify correctly all three classes, especially DAS. The scalogram based SVM model shows poor robustness at medium levels of GWN and complete loss to discriminated three classes of lung sounds at the highest levels of GWN.

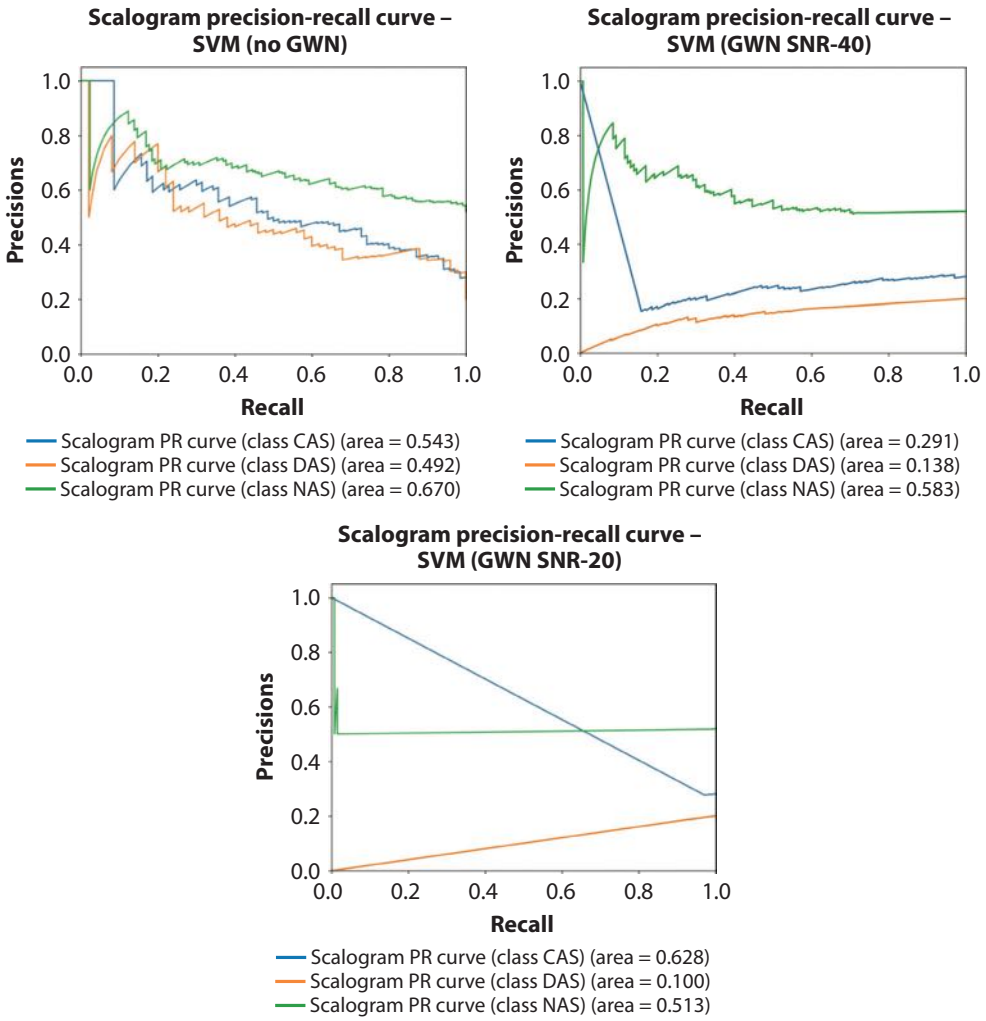


Fig. 5.2.68. *PR curve showing significant impact (P = 0.000) of GWN on SVM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, SVM – Support Vector Machines.

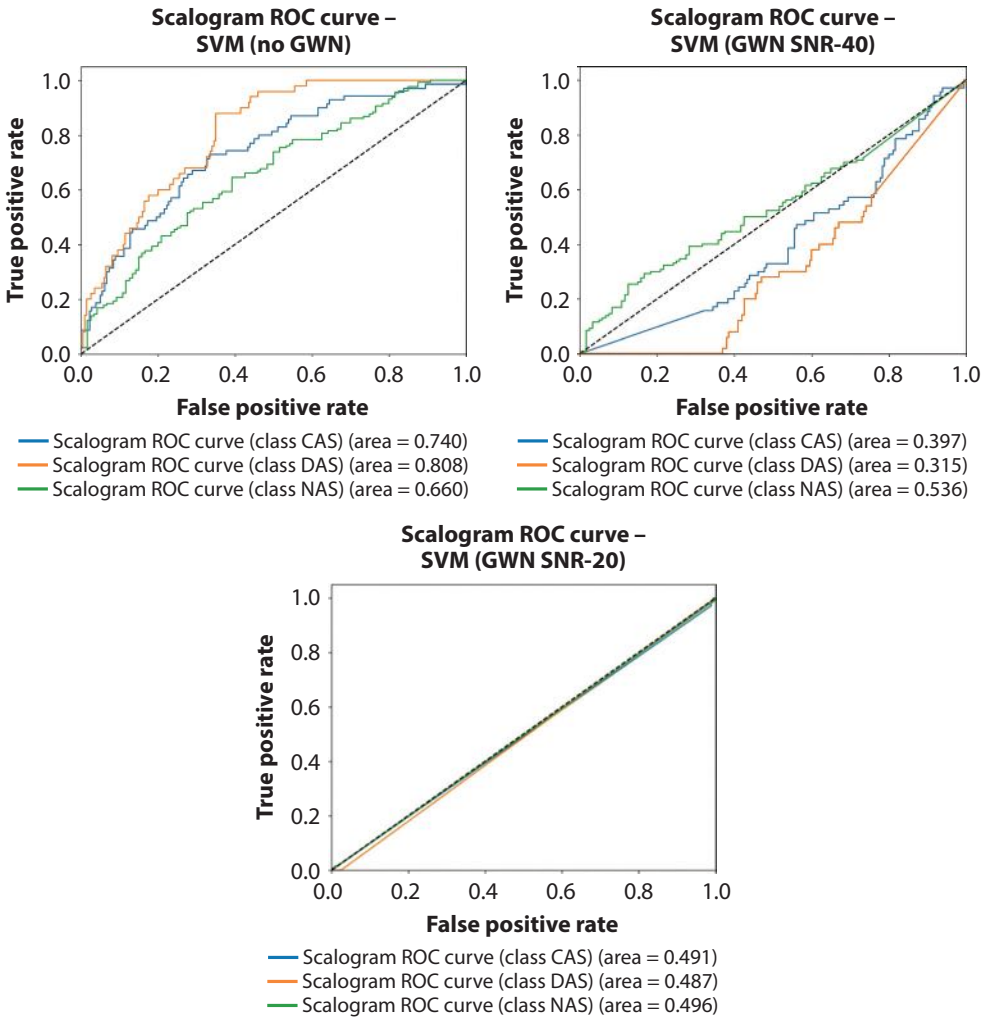


Fig. 5.2.69. *ROC curve showing significant impact (P = 0.000) of GWN on SVM model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, SVM – Support Vector Machines.

From Fig. 5.2.69, the impact of GWN levels on the true positive rate (TPR) as compared to the false positive (FPR) rate in the receiver operating characteristic (ROC) graph can be observed for the scalogram-based Support Vector Machines (SVM) ML model. At no GWN level, the model shows a good performance for DAS, above average for CAS and poorer performance for NAS sound classes identification, as seen from the area under the curve of ROC (ROC-AUC) values. The model performance drops significantly at medium levels of GWN (SNR-40) for all three classes. Finally, once GWN

with SNR-20 level is added, the performance becomes random (all the lines for all three classes adhere closely to the dashed random line). Therefore, the scalogram-based SVM model shows reasonable diagnostic power to identify lung sounds only at no GWN level, but lacks any robustness, even at medium level of GWN, as the model loses its capability to predict any of the three sound classes correctly.

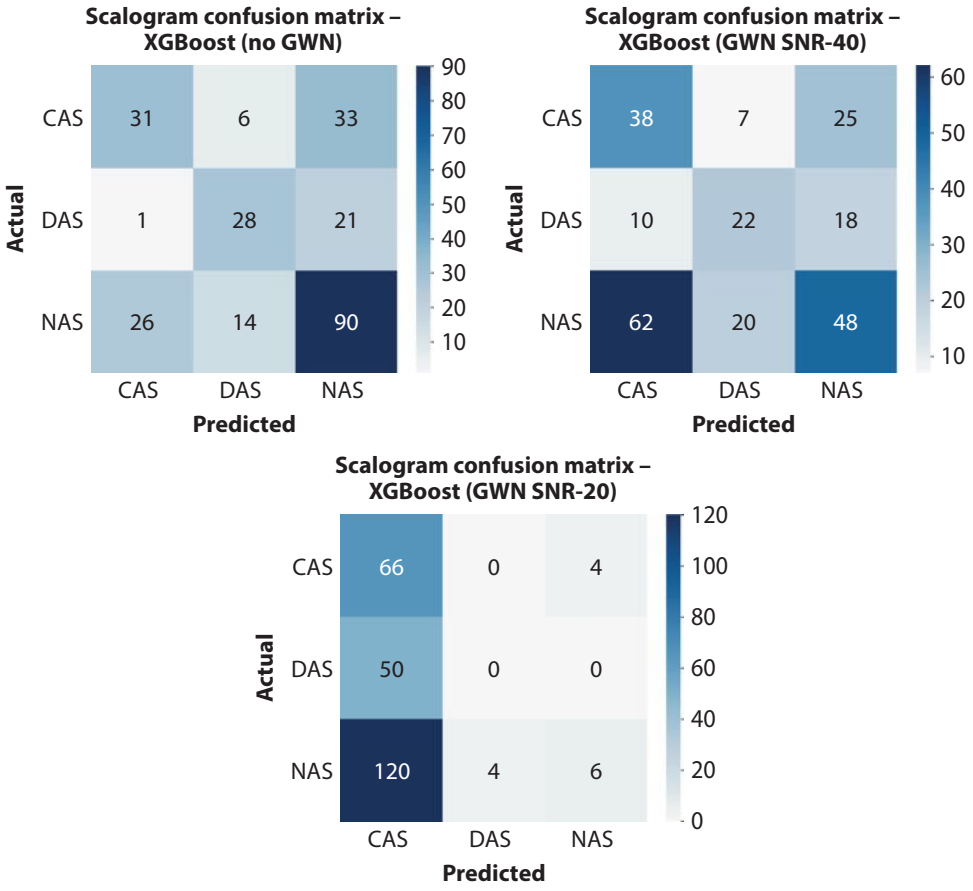


Fig. 5.2.70. *Confusion metrics showing significant impact (P = 0.000) of GWN on XGBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, XGBoost – Extreme Gradient Boosting classifier.

From Fig. 5.2.70, the impact of Gaussian white noise is clearly observed on true positive (TP), true negative (TN), false positive (FP) and false negative (FP) as seen in the scalogram-based XGBoost model's confusion matrix. At no GWN added levels, 31/70 CAS was identified correctly, 28/50 of the DAS class was correctly identified, and 90/130 of the NAS were correctly identified. The confusion matrix at SNR-40 (medium levels) shows a sharp decrease in performance of the model performance with a sharp drop in TP for the NAS class with only 48/130 correctly identified, whilst showing improved performance for DAS with 38/50 correctly and a slight decrease in CAS with 22/70 score. The increase in the GWN levels of SNR-20 saw a continued significant drop in the model's performance for DAS and NAS class sound identification with 0/50 and 6/130, respectively. Only the CAS class maintaining an extremely good TP score of 66/70. Overall, the scalogram-based XGBoost ML model showed reasonable performance for identifying DAS and CAS sounds at no GWN added levels, with some robustness to medium GWN levels but a bias towards mislabelling sound as NAS at highest ambient noise level.

From Fig. 5.2.71, it can be observed that scalogram-based XGBoost model achieved an overall reasonable performance except for one sound class. The precision to recall (PR) graph shows reasonable performance of the model in identifying NAS and DAS classes with worse performance observed for DAS, as exemplified via PR area under the curve (PR-AUC) scores. At medium level of GWN (SNR-40) recall scores for all three sound classes are negatively impacted. Only NAS shows the greatest robustness to medium level of GWN. Finally, once GWN is increased to SNR-20 level, we see an extremely negative impact on all three classes, especially CAS, followed by DAS, whilst NAS continues to show robustness to extreme noise. Nonetheless, once GWN levels are increased to SNR-20, the precision levels drop significantly for CAS and DAS lung sound classes, meaning that the model becomes useless at the highest GWN.

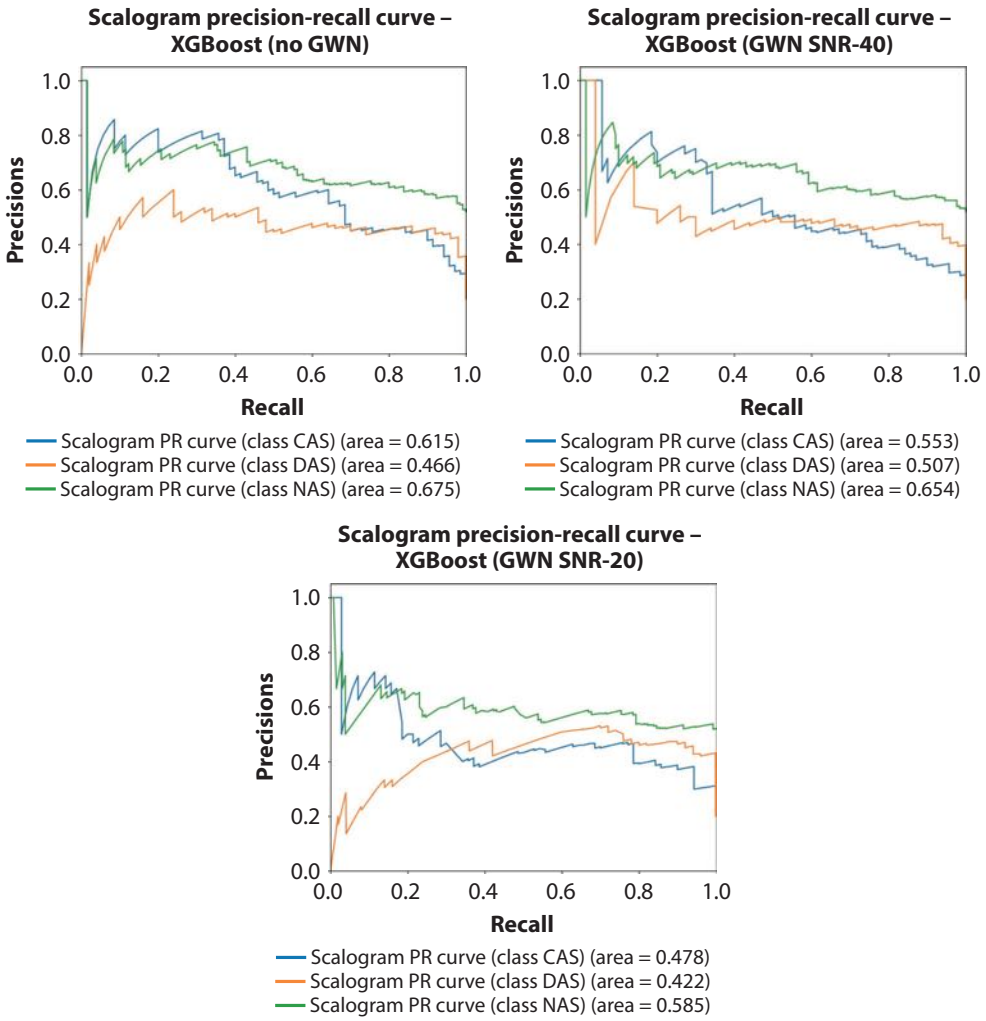


Fig. 5.2.71. *PR curve showing significant impact (P = 0.000) of GWN on XGBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, XGBoost – Extreme Gradient Boosting classifier.

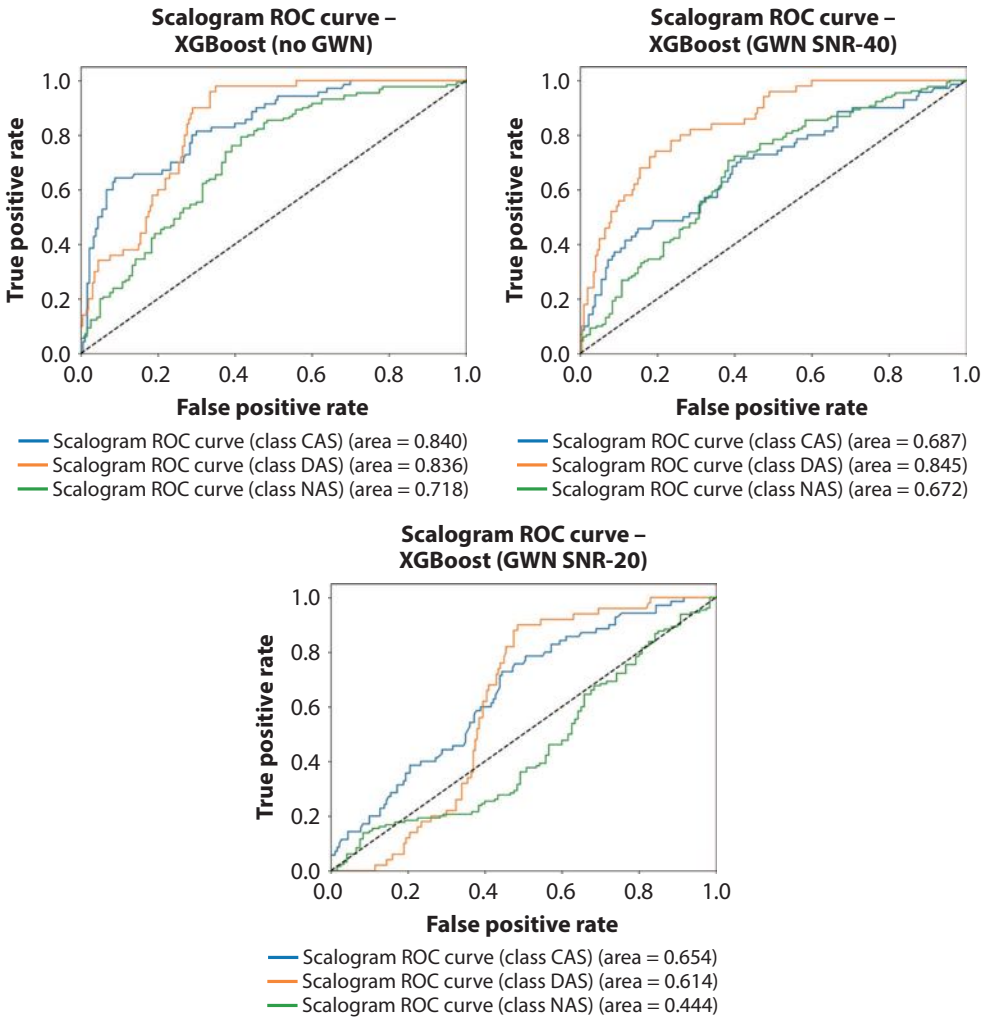


Fig. 5.2.72. *ROC curve showing significant impact (P = 0.000) of GWN on XGBoost model's ability to identify lung sounds correctly (from top to bottom)*

GWN – Gaussian white noise, NAS – normal auscultated sound, CAS – continuous auscultated sound, DAS – discontinuous auscultated sound, XGBoost – Extreme Gradient Boosting classifier.

From Fig. 5.2.72, the impact of GWN levels on the true positive rate as compared to the false positive rate in the ROC curve can be observed for the scalogram-based Extreme Gradient Boosting classifier (XGBoost) model. At no GWN level, the model shows a good performance for all three sound classes, with NAS class identification being the worst out of the three. The model's performance drops significantly at medium level of GWN (SNR-40)

for two classes: CAS and NAS. XGBoost shows strong robustness for the DAS class at medium level of GWN. Finally, once GWN is increased to SNR-20 level, the model loses its discrimination power to identify sound classes correctly, with greatest negative impact on ML model's identification of NAS sound class. Therefore, the scalogram-based XGBoost model is only viable at no GWN level, and higher levels of GWN (SNR-20) makes the model unviable to distinguish all three classes.

Table 5.2.1. *Twelve-spectrogram based models' performance according to ROC-AUC scores*

Spectrogram-based model	**ROC-AUC, median (IQR)**	**Test statistic**	**Degrees of freedom**	***P*-value**
AdaBoost	0.800 (0.689–0.853)	803	11	< 0.001
CatBoost	0.857 (0.764–0.880)			
Extra Trees	0.820 (0.691–0.859)			
Gradient Boosting	0.874 (0.772–0.897)			
Histgradient	0.865 (0.802–0.894)			
K-NN	0.751 (0.638–0.753)			
LightGBM	0.856 (0.782–0.879)			
Logistic Regression	0.863 (0.781–0.876)			
MLP	0.863 (0.786–0.902)			
Random Forest	0.833 (0.694–0.873)			
SVM	0.836 (0.746–0.853)			
XGBoost	0.871 (0.782–0.895)			

The best-performing spectrogram-based algorithms, according to the median performance, were all boosting models: Gradient Boosting, XGBoost, Histgradient. Histgradient had second highest medium, but narrower interquartile range with highest Q1 out of the top three models. IQR – interquartile range.

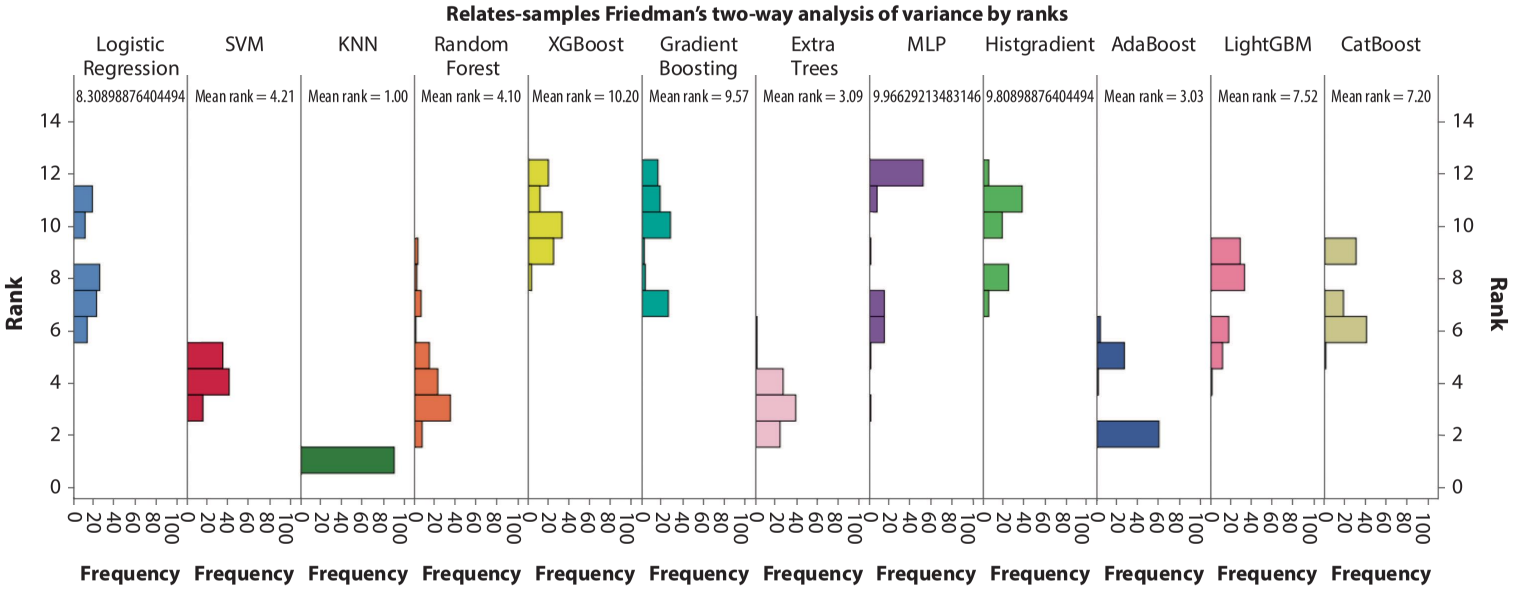


Fig. 5.2.73. *Post hoc analysis of 12 spectrogram-based ML models revealed significant differences in the algorithms' performance. XGBoost, Gradient Boosting, and Histgradient were the top-performing models, showing significant superiority over weaker models*

XGBoost demonstrated strong performance, surpassing K-NN (Z = inf, $P = 0.000$) and Extra Trees (Z = 8.210, $P = 0.000$). Gradient Boosting also performed well, outperforming K-NN (Z = 8.210, $P = 0.000$), AdaBoost (Z = 8.210, $P = 0.000$), and Extra Trees (Z = 8.210, $P = 0.000$). Histgradient was significantly better than K-NN (Z = inf, p = 0.000), AdaBoost (Z = inf, p = 0.000), and Random Forest (Z = 8.077, $P = 0.000$). K-NN was the worst-performing model, showing significant inferiority to all top models, including XGBoost (Z = inf, $P = 0.000$), Gradient Boosting (Z = 8.210, $P = 0.000$), and Histgradient (Z = inf, $P = 0.000$).

Table 5.2.2. *Twelve-scalogram based models' performance according to ROC-AUC scores*

Scalogram-based model	ROC-AUC, median (IQR)	Test statistic	Degrees of freedom	*P*-value
AdaBoost	0.735 (0.658–0.847)	574	11	< 0.001
CatBoost	0.794 (0.679–0.881)			
Extra Trees	0.746 (0.590–0.788)			
Gradient Boosting	0.752 (0.685–0.867)			
Histgradient	0.733 (0.671–0.850)			
K-NN	0.590 (0.528–0.658)			
LightGBM	0.732 (0.673–0.847)			
Logistic Regression	0.756 (0.671–0.814)			
MLP	0.741 (0.590–0.788)			
Random Forest	0.768 (0.635–0.808)			
SVM	0.740 (0.658–0.810)			
XGBoost	0.727 (0.659–0.859)			

The best-performing scalogram-based algorithms according to the median performance were one boosting, one tree based and one classical model, they were in the upper quartile of performance: CatBoost, Random Forest, and Logistic Regression. IQR – interquartile range.

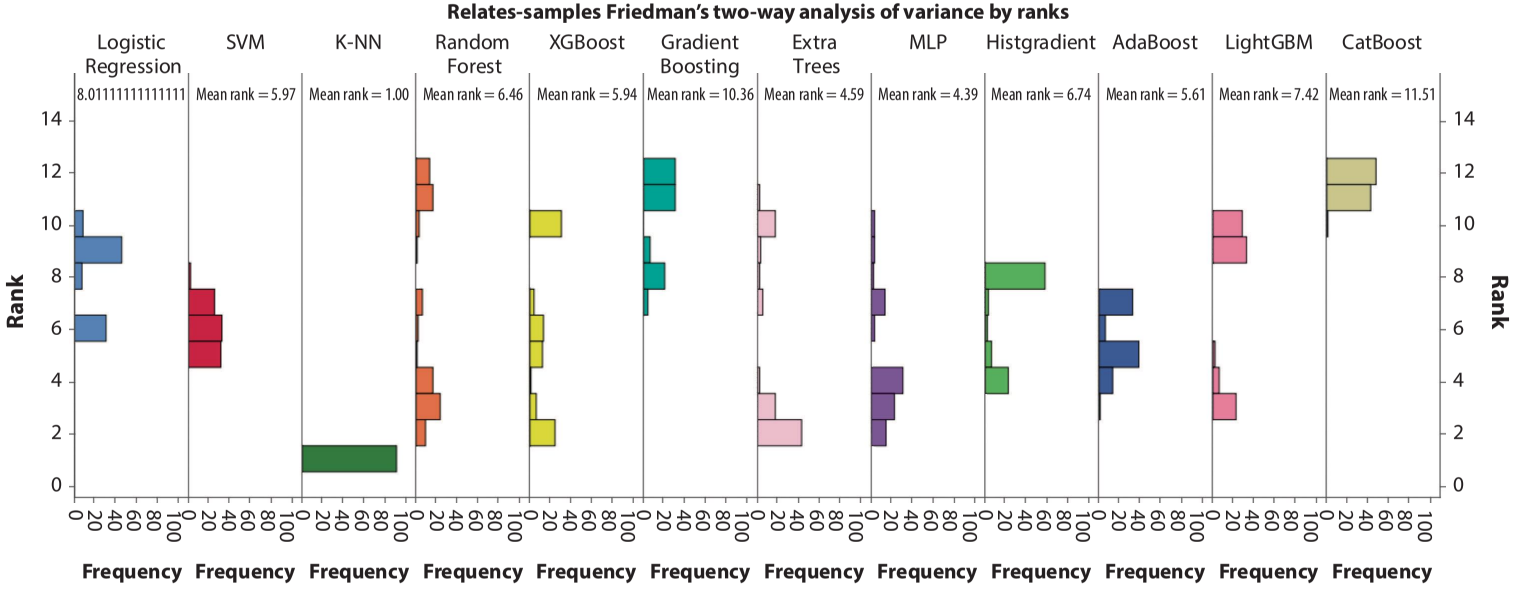


Fig. 5.2.74. *Post hoc analysis of 12 scalogram-based ML models revealed significant differences in the algorithms' performance. CatBoost, Gradient Boosting, and Logistic Regression showed statistically significant superiority over other models*

CatBoost outperformed K-NN (Z = –19.556, P = 0.000), Extra Trees (Z = –12.879, P = 0.000), and AdaBoost (Z = –10.977, P = 0.000). Gradient Boosting was significantly better than K-NN (Z = –17.406, P = 0.000), Extra Trees (Z = 10.729, P = 0.000), and Logistic Regression (Z = –4.362, P = 0.001). Logistic Regression also performed better than K-NN (Z = 13.044, P = 0.000) and Extra Trees (Z = 6.367, P = 0.000). K-NN was statistically the worst-performing model, showing significant inferiority to all top models (e.g., CatBoost: Z = –19.556, P = 0.000).

Table 5.2.3. *Twelve spectrogram-based models' compared to twelve scalogram-based models' performance according to ROC-AUC scores*

Spectrogram of 12 ML model, median (IQR)	Scalogram of 12 ML model, median (IQR)	Test statistic	*P*-value
0.837 (0.638–0.902)	0.735 (0.528–0.881)	583275	< 0.001

Wilcoxon Test shows significant difference between 12 ML models based on spectrogram and scalogram, with spectrogram-based models having a much higher median values as compared to scalograms-based models. IQR – Interquartile range.

5.3. Medical Faculty students' performance

In total 45 medical students attempted to learn three classes of lung sounds, over a period of 4 days and then performed a test under three levels of GWN noise (no added noise, GWN SNR-40 and GWN SNR-20).

The models all tested for overall impact of GWN on their performance via Friedman test and post hoc analysis.

From Fig. 5.3.1 the impact of different levels of Gaussian white noise (GWN) can be observed on three classes of lung sound's identification.

The noise levels are expressed in signal-to-noise ratio (SNR) from lowest levels (no GWN), medium (SNR-40) and to highest levels (SNR-20). Friedman test showed ability to identify NAS and DAS significantly varied ($P = 0.042$, 0.021, respectively) at the three levels of GWN, whilst no significant impact of GWN levels on CAS sounds were observed ($P = 0.311$).

Post hoc comparison was performed to evaluate the influence of the three levels of GWN on the ability to recognise NAS and DAS sound classes.

Statistically significant differences were found in lung sound recognition between no GWN and SNR-40 for NAS, between no GWN and SNR-40 and between SNR-40 and SNR-20 for DAS ($P = 0.016$, 0.013, 0.023, respectively).

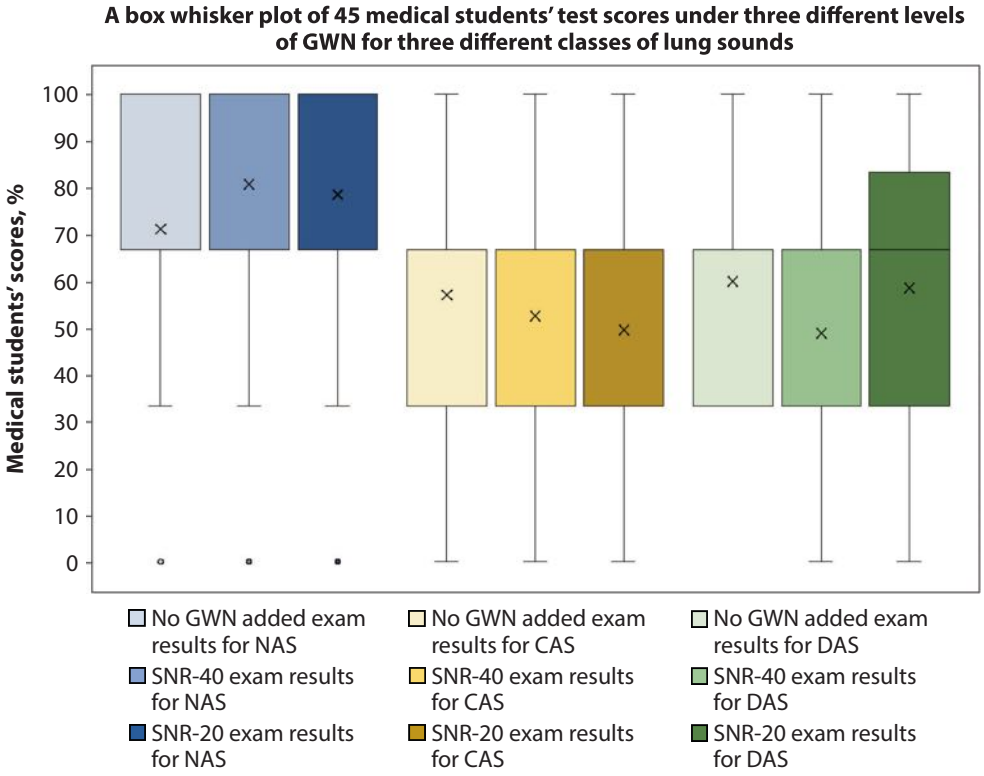


Fig. 5.3.1. *Medical students' exam scores for three classes of lung sounds under different levels of GWN. Impact of three levels of GWN on the ability for students to recognise continuous (CAS), discontinuous (DAS) and normal (NAS) lung sound classes.*

GWN – Gaussian white noise, SNR – signal to noise ratio, NAS – normal auscultated sound, DAS – discontinuous auscultated sound, CAS – continuous auscultated sound.

5.4. Comparison of best ML model's performance against Medical Faculty students' performance under different levels of GWN

Finally, the TN, TP, FN, FP values of students' scores were used to calculate MCC, specificity and sensitivity for each class of the sound under each level of GWN (no GWN, GWN SNR-40, GWN SNR-20).

The results were used to plot a box and whisker graph and Friedman test with post hoc analysis was performed (Fig. 5.4.1 to Fig. 5.4.3).

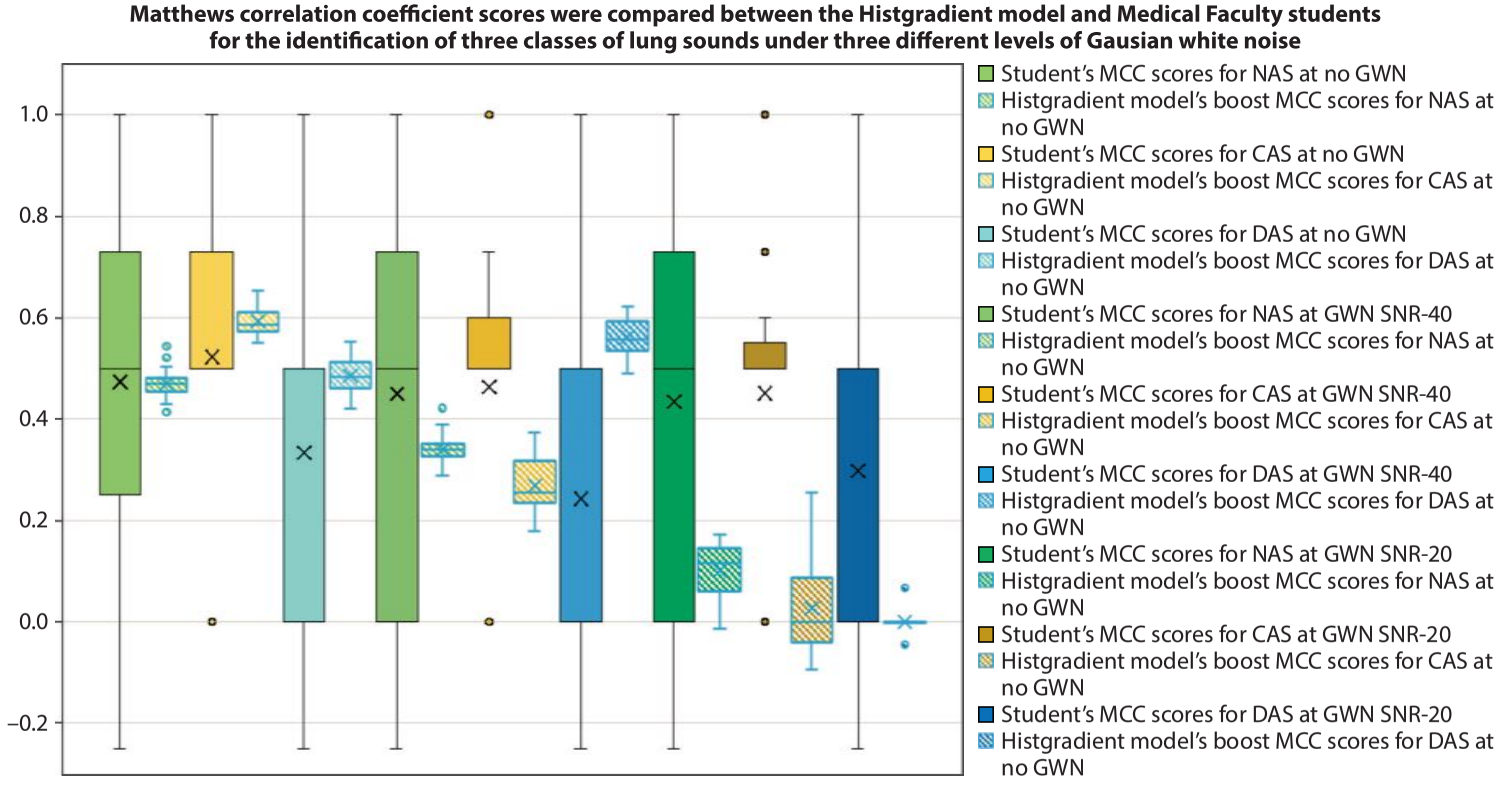


Fig. 5.4.1. *The box and whisker plot shows comparison of Histgradient model's and MF students' MCC scores for three sound classes (NAS, CAS, DAS) identification under 3 levels of GWN.*

GWN – Gaussian white noise, MCC – Matthews correlation coefficient, SNR – signal to noise ratio, NAS – normal auscultated sound, DAS – discontinuous auscultated sound, CAS – continuous auscultated sound, Histgradient boost – Histogram-based Gradient Boosting Classification Tree spectrogram-based model (Histgradient).

From Fig. 5.4.1 the impact of GWN on machine learning model and Medical Faculty students' diagnostic accuracy can be observed, by the comparison of Matthews correlation coefficient (MCC) score.

Students performed similarly to the spectrogram-based Histgradient model in the no GWN added condition, as no significant differences between the two study groups (MFS *vs.* ML) were observed for the two sound classes identification rates: NAS and CAS ($P > 0.05$). However, there was a statistically significant difference between the two study groups for the DAS class of lung sound identification ($P = 0.002$). The ML models' MCC scores of 0.471 (0.415 to 0.543), 0.587 (0.522 to 0.654), 0.485 (0.422 to 0.552) vs. MFS 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.00), 0.500 (–0.250 to 1.000) for NAS, CAS, DAS, respectively. The ML model showed superior MCC scores for DAS class identification under no GWN conditions.

With GWN at SNR-40 level, there was statistical significance between all three sound groups: NAS, CAS, DAS ($P = 0.035$, $P = 0.002$, $P = 0.000$) with ML scores of 0.341 (0.288 to 0.422), 0.256 (0.180 to 0.374), 0.557 (0.491 to 0.621) *vs.* MFS 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.000), 0.000 (–0.250 to 1.000 for NAS, CAS, DAS respectively. The MF students showed superior performance in identifying NAS and CAS classes whilst ML model outperformed human subjects under SNR-40 for DAS sound class recognition.

Whist at GWN SNR-20 level, students showed statistically significantly better results for all classes of sounds recognition, than Histgradient spectrogram-based ML model ($P = 0.000$ for NAS and CAS classes and $P = 0.009$ for DAS class) with ML scores of 0.116 (–0.013 to 0.173), 0.001 (–0.095 to 0.255), 0.000 (–0.045 to 0.067) vs. 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.000), 0.500 (–0.250 to 1.000) for NAS, CAS, DAS respectively. Therefore, at the highest levels, the MMC scores of the MFS group were significantly higher than those of the ML Histgradient model, strongly indicating human subjects' robustness to the highest GWN levels compared to the ML model.

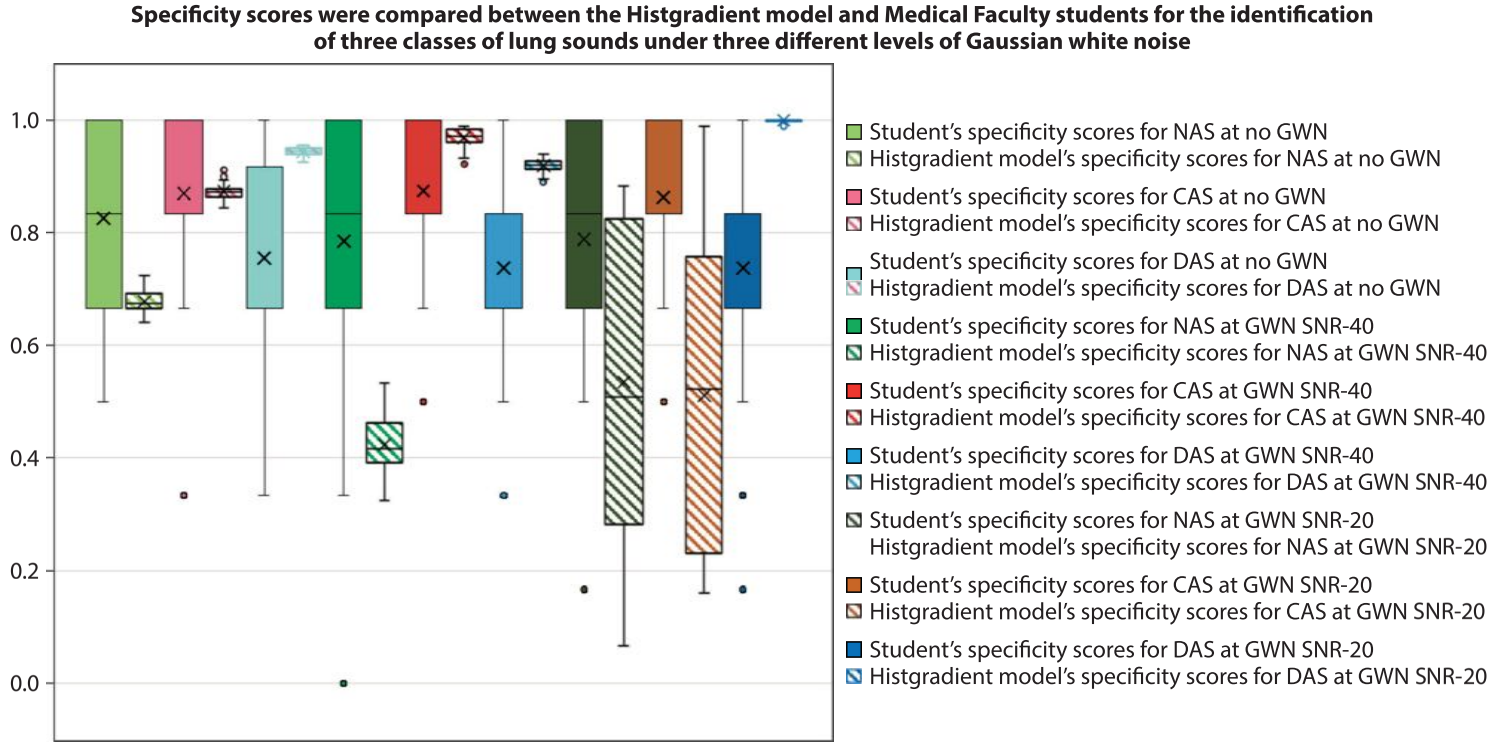


Fig. 5.4.2. *The box and whisker plot shows comparison of Histgradient model's and MF students' specificity scores for three sound classes (NAS, CAS, DAS) identification under 3 levels of GWN.*

GWN – Gaussian white noise, SNR – signal to noise ratio, NAS – normal auscultated sound, DAS – discontinous ausculatated sound, CAS – continuous auscultated sound, ML – machine learning.

From Fig. 5.4.2 the impact on machine learning (ML) model and medical faculty students (MFS) of different levels of Gaussian white noise (GWN) can specificity be observed on three classes of lung sounds.

In the no GWN added condition, the specificity distributions of the spectrogram Histogram-based Gradient Boosting Classification Tree machine learning (ML) model and the students scores were not significantly different for NAS and CAS ($P > 0.05$ for both classes), with ML model specificities of 0.471 (0.415 to 0.543) and 0.587 (0.522 to 0.654) for NAS and CAS, respectively, compared to medical faculty students' (MFS) specificities of 0.500 (–0.250 to 1.000) and 0.500 (0.000 to 1.000) for the same classes. However, for DAS, the ML model showed significantly higher specificity than the students' scores ($P = 0.000$), with ML specificity of 0.485 (0.422 to 0.552) compared to specificity of 0.500 (–0.250 to 1.000).

At GWN SNR-40, no significant differences were observed between the two study groups for CAS class identification ($P > 0.05$). However, significant differences were found between the NAS and DAS classes (P = 0.000, P = 0.024), with Histgradient ML model scores for DAS class identification being higher than those of MFS. However, the MFS group identified better NAS class sounds. The ML model showed specificities of 0.341 (0.288 to 0.422), 0.256 (0.180 to 0.374), and 0.557 (0.491 to 0.621) for NAS, CAS, and DAS, respectively, while the MFS had specificities of 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.000), and 0.000 (–0.250 to 1.000) for the same classes.

In contrast, at GWN SNR-20, the MFS demonstrated statistically significantly better specificity results than the spectrogram-based Histgradient ML model for all sound classes ($P = 0.000$ for all classes). This ML model specificities were 0.116 (–0.013 to 0.173), 0.001 (–0.095 to 0.255), and 0.000 (–0.045 to 0.067) for NAS, CAS, and DAS, respectively, compared to the MFS specificities of 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.000), and 0.500 (–0.250 to 1.000) for the same classes. Therefore, at the highest levels of GWN, the MFS study group outperformed in specificity scores for NAS and CAS classes. At the same time, the ML Histgradient model maintained its statistically significant advantage for DAS class identification over students, even at the highest levels of GWN.

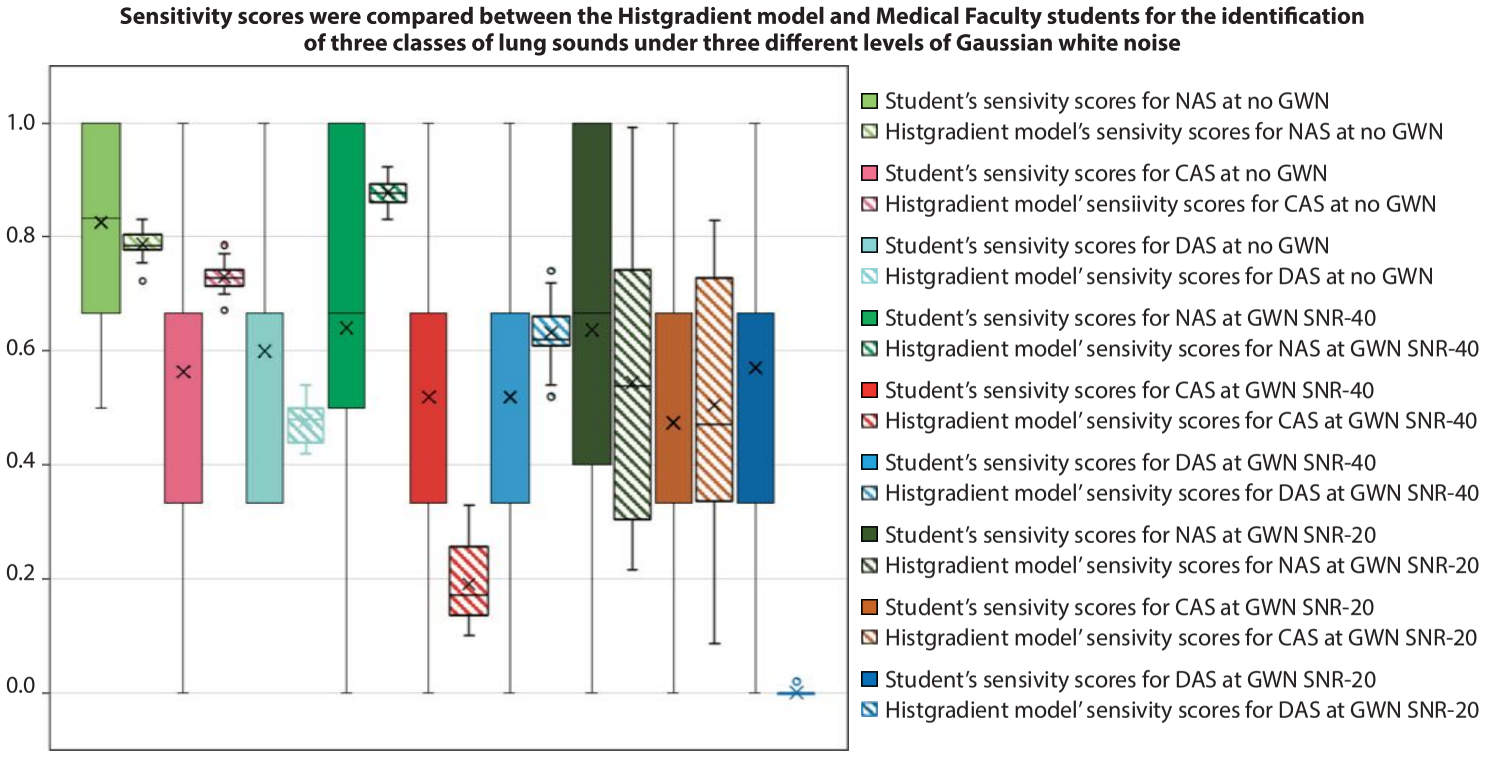


Fig. 5.4.3. *The box and whisker plot shows comparison of Histgradient model's and MF students' sensitivity scores for three sound classes (NAS, CAS, DAS) identification under 3 levels of GWN.*

GWN – Gaussian white noise, SNR – signal-to-noise ratio, NAS – normal auscultated sound, DAS – discontinuous auscultated sound, CAS – continuous auscultated sound.

From Fig. 5.4.3 The impact of different levels of Gaussian white noise (GWN) on machine learning (ML) models and medical faculty students' (MFS) sensitivity scores.

At no GWN added condition, the sensitivity scores of the spectrogram-based Histgradient ML model's and the medical faculty students' (MFS) scores do not significantly different for DAS class identification ($P > 0.05$). However, that was statistically significant difference between study groups in identifying NAS and CAS class of lung sounds ($P = 0.030$ and $P = 0.000$, respectively). The ML sensitivities scores were as follows: 0.471 (0.415 to 0.543), 0.587 (0.522 to 0.654), and 0.485 (0.422 to 0.552) for NAS, CAS, and DAS, respectively, compared to MFS sensitivities of 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.000), and 0.500 (–0.250 to 1.000) for the same classes. Therefore, data analysis shows that ML model hold statistically significant advantage at no GWN levels for NAS and CAS classes and evenly matches MFS for DAS class.

At GWN SNR-40, no significant differences were observed for DAS class sensitivity median scores ($P > 0.05$), but significant differences were found for NAS and CAS classes ($P = 0.000$ for both groups), with the Hist-gradient ML model having higher sensitivity than MFS for the NAS class and lower for the CAS class of lung sounds. The ML model showed sensitivities of 0.341 (0.288 to 0.422), 0.256 (0.180 to 0.374), and 0.557 (0.491 to 0.621) for NAS, CAS, and DAS, respectively, while the MFS scores had sensitivities of 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.000), and 0.000 (–0.250 to 1.000) for the same classes.

At GWN SNR-20, both study groups (ML Histgradient model and MFS) were not statistically different in their sensitivity while identifying NAS and CAS classes of lung sounds ($P > 0.05$ for both). However, the sensitivity of spectrogram-based Histgradient ML model was statistically lower for DAS class than that of MFS under the highest GWN level ($P = 0.000$). The ML model sensitivities were 0.116 (–0.013 to 0.173), 0.001 (–0.095 to 0.255), and 0.000 (–0.045 to 0.067) for NAS, CAS, and DAS, respectively, compared to the MFS sensitivities of 0.500 (–0.250 to 1.000), 0.500 (0.000 to 1.000), and 0.500 (–0.250 to 1.000) for the same classes.

Therefore, the ML model performs better in sensitivity at no GWN levels and has some robustness at medium GWN levels. However, at the highest levels of GWN (SNR-20), human subjects catch up with the ML model and outperform it for DAS class lung sounds.

6. DISCUSSION

The research was set up to explore and better understand the integration of machine learning tools as an AI-based decision assistant for healthcare workers under different ambient noise levels.

The research had two main parts: training machine models and medical students, and comparing their sensitivity, specificity and MCC scores under three levels of GWN for both groups. The project was set up to achieve these goals with four objectives in mind. These objectives were used to create a methodology where medical students and machine learning models were trained and assessed to identify three lung sound classes under three levels of GWN. Additionally, spectrogram and scalogram-based models were compared for their performance under different GWN levels for three lung sound classes identification. The evaluation of the models was performed using the following assessment metrics: ROC-AUC and MCC and supported by PR curve and confusion matrix, to evaluated GWN impact at SNR-40 and SNR-20 levels on three classes of lungs sounds identification (NAS, CAS, DAS), and finally to compare the ability of machine learning models and medical students to identify three classes of lung sounds under three different levels of GWN.

During the project, 124 patients were auscultated, and 108 patients were selected for the research project. 52 medical students rolled into auscultation training and assessment under GWN conditions, of which 45 completed the study fully.

A proprietary website with a training and assessment section was created for students.

First, ML models were trained using extracted features from scalograms and spectrograms. The research training ran 30 times, with average data used to create a confusion matrix, PR-AUC, ROC-AUC and calculate MCC. Statistical significance between models was evaluated using Friedman's test with a post hoc analysis.

The models were evaluated based on the overall performance of the three classes of lung sound detection under no GWN conditions (the best conditions for the model). ROC-AUC was the primary criteria. Secondary criteria were that the IQR range would be narrow, with the lower quartile as high as possible and the highest possible median score. The evaluation of spectrogram showed GB, XGBoost, MLP, Histgradient and LR models being the top 5 models, with Histgradient being in top 3 in median score and having the best IQR range.

The middle performing spectrogram-based ML models were CatBoost, LightGBM, followed by similarly performing SVM, Random Forest, Extra Trees. Finaly the worst two models were AdaBoost and K-NN, from which two K-NN had the worst performance out of the two.

The Friedman test results indicate statistically significant differences among the evaluated algorithms. Post-hoc analysis was applied with an adjusted *P*-value set at less than 0.05.

In summary, the statistical analysis confirms that GB, Histgradient and XGBoost are the top performers in this evaluation, while K-NN and AdaBoost results were underwhelming.

The study results concur with past research, showing very strong performance by Hisgradient and XBBoost ML models and emphasize the importance of selecting the right model. Previous research has shown that XGBoost can outperform other models in respiratory sound detection. [127]. The other research shows that XGBoost and Histgradient outperforming MLP, RF, AdaBoost [128]. Additionally, this research adds to the body of scientific knowledge by comparing statistically spectrogram-based 12 models' performance for lung class recognition.

Whilst the story was slightly different for scalogram-based ML models, especially concerning the question of best performing ones. The training and assessment of scalogram-based models showed that there was a significant difference between the 12 ML models, and CatBoost model came out on top as best performing ML model, significantly better than several other models, for instance, better than LightGBM, SVM, Logistic Regression ($P = 0.001$).

Random Forest model was also a strong, but not always significantly better. XGBoost took third place as a good, but slightly behind RF and CatBoost. LightGBM was similar to XGBoost with no significant difference, but slightly weaker performance. To worst performing models were SVM and Logistic Regression and were significantly weaker than top ones. SVM and Logistic Regression performed significantly worse than the tree-based models (CatBoost, Random Forest and XGBoost).

Though there are no direct studies comparing CatBoost and other ML models for lung sounds under ambient noise conditions, yet, interesting glimpses can be acquired by looking at other types of studies, which used this ML model.

For instance, Qin Yifan study shows that the best performing model was CatBoost in predicting diabetes through lifestyle, followed by XGBoost, Random Forest (RF), Logistic Regression (LR), and Support Vector Machines (SVM) [129]. These findings align somewhat with our study on ML scalogram-based model performance, though the type of predictive model being built was very different.

The study by Zaman R. Syed had more similar research topic, where voice recordings used to predict biometric features. This research showed that CatBoost ML model performs best among all in predicting human biometric information from voice timbre with 96.4% test accuracy, compared to Random Forest and XGBoost. On the other hand, Random Forest performs best for predicting age, among all ML models used, with 70.4% test accuracy. For emotion prediction, XGBoost performs best with 66.1% test accuracy [130]. This research emphasis, again, on how models' accuracy varies depending on type of data being classified.

Hence, we can see that each ML model has its advantages and disadvantages, furthermore, only by running experiments empirically and comparing them, we can deduce an optimal model for pulmonary lung sound recognition.

The 12 spectrogram-based ML models were compared to their 12 scalogram-based counterparts. The results showed a significant difference between the two groups, with spectrogram-based models outperforming their equivalent counterparts.

This indeed was a bit of surprise as scalograms, technically, should preserve more information then spectrograms and, in theory, perform better even in noisy conditions [131].

The phenomena could be explained by two main points. First, the dataset might not have been large enough. The article by Pratham N. Soni stated that limited datasets could lead to overfitting in small datasets due to their high dimensional feature space [132].

The other issue with scalograms is that they require more finetuning to get the most out of them. This topic has been explored by Addison S. Paul in 2002 book [133]. Therefore, the limited conclusion can only be drawn that due to relative spectrogram simplicity, models adaptability and limited datasets spectrogram-based models were significancy better in recognising lungs sounds.

The second stage of the study was to assess the human subjects' ability to recognise different class of lung sound at GWN environment, trained on the same data.

Medical students were chosen for the study, as auscultation training typically begins with them, whilst them being motivated to learn lung auscultation skills, as part of development, towards becoming physicians. Younger subjects were also less likely to experience hearing impairments [134]. Furthermore, confounding variables such as age, subjects' environment and training hours could be more easily controlled.

The study enrolled 52 LMSU second- and third-year students and after training for 4 days students took an exam under three levels of GWN.

The results showed that GWN had a statistically significant impact on the ability of subjects to recognise specific classes of lung sounds. This ability to identify NAS and DAS significantly varied ($P = 0.042$, 0.021, respectively) at all three levels of GWN, whilst no significant impact of GWN levels on CAS sound recognition was observed ($P = 0.311$).

Post hoc analysis of the NAS and DAS classes revealed a statistically significant difference in students' scores for the NAS class between no GWN and SNR-40 ($P = 0.016$). For the DAS class, significant differences were found between no GWN and SNR-40 ($P = 0.013$) and between SNR-40 and SNR-20 ($P = 0.023$).

The hypothesis that ambient noise uniformly impacts all lung sound classes' identification was rejected. The findings indicated that background noise especially affected DAS class, which was the most difficult to identify at SNR-40 level of noise pollution.

Existing research shows that crackles are more difficult to identify correctly than wheezes, which belong to the DAS and CAS classes of lung sounds respectively [135]. Particularly, research by Ye *Peitao* examined the ability of 56 subjects to distinguish fake crackles from real ones and concluded that the former has a statistically significant impact on misdiagnosis [136]. This research indicates another contributing factor, noise, as demonstrated by different levels of GWN. This factor is concerning because DAS lung sounds are associated with heart failure and pneumonia; therefore, a lack of early diagnosis could adversely impact the care of these patients and negatively affect preliminary treatment plan.

Assessing acoustic properties is a key in understanding why DAS is affected more than CAS. Amongst the two classes, adventitious lung sounds and wheezes are continuous, high-pitched sounds with a frequency of 400 Hz, lasting more than 80 ms. In contrast, crackles are discontinuous, exhibiting a wider frequency range of 100–2000 Hz but with a notably shorter duration of less than 20 ms [137].

Fine crackles are hard to hear due to their short duration. In a previous study by Moriki Dafni, which included 296 physicians with different specialities and levels of expertise, only 55.2% correctly identified fine crackles, compared to 72.2% who correctly recognised wheezes [138]. They can also be more easily confused with the rubbing of the stethoscope membrane sound [136]. The study used only five audio-recorded respiratory sounds that physicians had to listen to and document their responses.

Whilst CAS appears not to be impacted by GWN, this finding may not hold true if different types of background noise, such as babbling or car sounds, are used.

Another major reason why CAS is least affected by GWN is that wheezes have the most distinct audio qualities amongst the three classes. Whilst NAS could potentially be confused with DAS, especially when GWN is introduced, students misidentify these lung sounds even at no GWN and SNR-20 levels.

Regarding the DAS class of sounds, a fascinating observation is obtained: identifying lung sounds at SNR-40 is more difficult compared to SNR-20. Previous research has already identified crackles as problematic to identify and easy to confuse, particularly due to fake crackles, a wide frequency range and their short duration [136, 137]. This research indicates that not only is the DAS class harder to identify, but it is also the most affected by noise pollution. Interestingly, this class is impacted most at the medium noise level (SNR-40) rather than at higher intensity (SNR-20).

Finally, the final stage of the study looked at comparing MCC, sensitivity and specificity values. MCC value was chosen in additional to standard evaluation coefficients as there was an imbalanced dataset in machine learning with lower DAS class and therefore a more balanced matric was needed than a standard accuracy.

Though, as previously mentioned students' ability to recognise sounds was impacted now it was time to compare it to machine learning models.

The comparison of the Histgradient model and MF students' MCC performance under different levels of GWN revealed interesting trends. Under the no GWN condition, MF students performed similarly to the Histgradient model, as no statistically significant differences were observed across NAS, CAS, and DAS classes ($P > 0.05$ for all). The ML scores for Histgradient were 0.471 (IQR: 0.415–0.543), 0.587 (IQR: 0.522–0.654), and 0.485 (IQR: 0.422–0.552) for NAS, CAS, and DAS, respectively, whereas MF students consistently scored 0.500 (IQR: –0.250–1.000), 0.500 (IQR: 0.000–1.000), and 0.500 (IQR: –0.250–1.000) across all classes, indicating comparable performance in noise-free conditions.

Whilst GWN at SNR-40 level, the results showed a mixed pattern. While there was no overall statistical significance between the two subject groups, significant differences were observed for each sound class ($P = 0.035$ for NAS, $P = 0.002$ for CAS, and $P = 0.000$ for DAS). The ML model's performance declined for NAS and CAS, with MCC scores of 0.341 (IQR: 0.288–0.422) and 0.256 (IQR: 0.180–0.374), respectively. However, for DAS, the Histgradient model achieved a notably higher MCC score of 0.557 (IQR: 0.491–0.621), outperforming MF students, who scored just 0.000 (IQR: –0.250–1.000). This suggests that while the ML model struggled with NAS and CAS under moderate noise, it demonstrated superior performance

in identifying DAS, a class of lung sounds that proved particularly difficult for human listeners.

Finally, at the highest GWN level (SNR-20), MF students significantly outperformed the Histgradient model across all sound classes ($P = 0.000$). The ML model's performance deteriorated sharply, with MCC scores of 0.116 (IQR: –0.013–0.173), 0.001 (IQR: –0.095–0.255), and 0.000 (IQR: –0.045–0.067) for NAS, CAS, and DAS, respectively, whereas MF students maintained a consistent score of 0.500 across all classes' recognition.

Concerning sensitivity, varied GWN conditions reveals important differences in performance between the spectrogram-based Histgradient ML model and MF students. In the no GWN added condition, while both groups showed comparable sensitivity for DAS classification (ML: 0.485, IQR: 0.422-0.552 vs MFS: 0.500, IQR –0.250–1.000; $P > 0.05$), the ML model demonstrated significantly better performance for NAS (0.471, IQR: 0.415–0.543 vs 0.500, IQR: 0.250–1.000; $P = 0.030$) and CAS (0.587, IQR: 0.522–0.654 vs 0.500, IQR: 0.000–1.000; $P = 0.000$).

At medium levels of GWN (SNR-40), the ML model maintained higher sensitivity than MFS for NAS (0.341, IQR: 0.288–0.422 vs 0.500, IQR: 0.250–1.000; $P = 0.000$) but showed lower sensitivity for CAS (0.256, IQR: 0.180–0.374 vs 0.500, IQR: 0.000–1.000; $P = 0.000$), with no significant difference in DAS classification ($P > 0.05$). Notably, the ML model's DAS sensitivity (0.557, IQR: 0.491-0.621) contrasted sharply with MFS performance (0.000, IQR: 0.250–1.000).

Under the highest tested GWN conditions (SNR-20), the ML model's sensitivity dropped substantially across all classes: NAS (0.116, IQR: 0.013–0.173), CAS (0.001, IQR: 0.095–0.255), and DAS (0.000, IQR: 0.045–0.067). While no significant differences existed between two study groups for NAS and CAS class identification ($P > 0.05$), as students were also impacted by ambient noise. There was a significant performance difference between the study groups in DAS lung class identification sensitivity scores, with MFS outperforming ML model (0.500, IQR: 0.250–1.000 vs ML; $P = 0.000$).

These results demonstrate that while the ML model shows superior sensitivity in noise-free conditions, particularly for NAS and CAS classifications, its performance degrades with increasing noise levels. Therefore, ambient noise levels significantly affect ML model performance, with degradation as noise increases. This analysis highlights the ML model's good sensitivity scores in low-noise with some robustness to moderate (SNR-40) GWN conditions but vulnerability to higher GWN levels. At the same time, MFS show more consistent (though variable across lung sound classes)

performance across noise levels although with a trade-off that students have higher variability, as shown by interquartile range as compared to ML model.

Specificity comparison findings reveal distinct performance patterns between the Histgradient ML model and MF students under varying GWN conditions.

Under no GWN conditions, the ML model demonstrated comparable specificity to MFS for NAS (ML: 0.471, IQR: 0.415–0.543 vs MFS: 0.500, IQR: –0.250–1.000; $P > 0.05$) and CAS (ML: 0.587, IQR 0.522–0.654 vs MFS: 0.500, IQR: 0.000–1.000; $P > 0.05$), while showing significantly better performance for DAS (ML: 0.485, IQR: 0.422–0.552 vs MFS: 0.500, IQR: –0.250–1.000; $P = 0.000$). Additionally, the ML model holds an advantage over human subjects by exhibiting lower interquartile values (IQR), indicating lower variability than human subjects, but this is only true under low ambient noise conditions.

When moderate noise was introduced (SNR-40), the ML model maintained its advantage in DAS classification (ML: 0.557, IQR: 0.491–0.621 vs MFS: 0.000, IQR: –0.250–1.000; $P = 0.024$) but was outperformed by MFS in NAS (ML: 0.341, IQR: 0.288–0.422 vs MFS: 0.500, IQR: –0.250–1.000; $P = 0.000$), with comparable performance in CAS ($P > 0.05$).

Under high noise conditions (SNR-20), MFS showed consistently superior specificity across all classes: NAS (0.500, IQR: –0.250–1.000 vs ML: 0.116, IQR: –0.013–0.173; $P = 0.000$), CAS (0.500, IQR: 0.000–1.000 vs ML: 0.001, IQR: –0.095–0.255; $P = 0.000$), and DAS (0.500, IQR: –0.250–1.000 vs ML: 0.000, IQR: –0.045–0.067; $P = 0.000$).

These specificity results demonstrate that while the ML model performs well in no GWN added and medium levels of GWN (SNR-40), particularly for DAS classification, MFS exhibit greater robustness in noisy environments, maintaining stable performance where the ML model's accuracy deteriorates significantly.

The substantial performance gap at GWN SNR-20 for specificity, MCC scores, and sensitivity suggests that ML models require additional noise resilience improvements to match healthcare workers performance in real-world clinical settings where acoustic interference is common such as emergency room settings.

These findings indicate that ML models could serve as valuable clinical assistants, particularly for detecting discontinuous auscultated sounds in quiet and medium-noisy environments where human perception is significantly impaired. This could lead to better sensitivity for diagnosing DAS types of adventitious lung sounds, such as crackles, which are associated with leading morbidities such as HF and pneumonia.

Moreover, study results highlight both the potential strengths and limitations of current ML models in lung sound classification under ambient noise conditions. Human listeners generally outperform ML models in highly noisy environments, suggesting a need for further research that explores improving robustness for lung sound classification models.

The model's diagnostic sensitivity to detect DAS sounds was not just due to model or spectrogram visualisation features, but also due to rudimentary fine tuning where sensitivity threshold for DAS was set at 1.30, for CAS at 1.1 and for NAS at 1.0. This was done due to the fact that DAS class of dataset was smaller as compared to other two classes and additionally it is known that discontinuous lung sounds are more difficult to detect. Therefore, showing potential and importance of fine-tuning ML models.

Future work should focus on enhancing ML models' robustness through noise-adaptive training strategies, such as data augmentation and advanced denoising methods, to further improve performance across all lung sound classes.

The literature overviews on machine learning model's robustness have shown limited data available of lung ambient noise impact on auscultation, but our research empathises paramount importance in such research kind.

This clearly shows the need not only training and assessing performance of ML models under different levels of GWN and other types of ambient noise, but also comparing trained models to human subjects' abilities. Only then ML models can fully be integrated as diagnostic tool to assist healthcare workers.

Advantages and disadvantages of the study

There are several advantages to this study design. First, this is the first research to compare human subjects' ability to learn and identify three classes of lung sounds under three levels of GWN. Secondly, the research project has achieved statistically significant results in showing the impact of GWN on ML models. The research used a substantial number of ML model variations; 24 in total (12 spectrogram-based and 12 scalogram-based), which allowed for comparing a large number of ML models under the same conditions as a tool for classification of lung sounds. The study compared human subjects and ML models across different sound classes and ambient noise levels. This allowed for the revelation of advantages and disadvantages of organic intelligence versus AI, i.e., top ML model (Histgradient) performing better under low noise conditions for DAS class, whilst human subjects performed better under noisiest test conditions.

The drawback of the work is that even though 250 recordings were used for training and assessing ML models, this might still not be sufficient to fully exploit the models' potential. Though threshold and fine tuning were attempted with some success for boosting ML models, the scalogram results were very disappointing in their diagnostic accuracy. This might be due to the real noise environment conditions in which the recordings were collected.

The student number was also relatively small at 52 subjects, of which 45 completed the study fully. Nonetheless, meaningful and statistically significant results were achieved. Additionally, this was not a multi-clinical study and physicians and nurses were not involved. To expand and make the results even more applicable, it would be important to include all types of healthcare workers to better understand how assisted ML diagnostics could support the specialist in diagnosing lung pathologies under different types of ambient noise.

CONCLUSION

1. Machine learning models and medical students can be trained to identify three classes of lung sounds but with various levels of accuracy.
2. Spectrograms based models showed significant better accuracy as compared to scalograms across 12 machine learning models.
3. Increased Gaussian white noise levels affected the ability of both medical students and machine learning models to recognise lung sounds. Machine learning models were more often affected by the highest level of sound contamination (SNR-20), where the ability to recognise normal lung sounds, discontinuous, and continuous lung sound classes significantly decreased. Meanwhile, out of three lung sound classes, the medical students' accuracy in recognising discontinuous was most significantly affected by the medium level of Gaussian white noise (SNR-40).
4. The machine learning Histgradient model at the SNR-40 GWN level outperformed medical students in recognising discontinuous lung sounds with higher Matthews correlation coefficient and sensitivity while maintaining reasonable specificity. Therefore, this ML shows potential for use as a diagnostic assistant in low ambient noise conditions.

PRACTICAL RECOMMENDATIONS

1. All physicians, nurses, residents and medical students should know their diagnostic accuracy for a given lung sound class and noise level in their work environment. This assessment could be performed via web-based examination format.
2. All future machine learning models should be evaluated for environmental noise conditions. The models performance should be freely available for access and scrutiny.
3. The Histgradient Boost machine learning model should be further explored and developed to help identify lung sounds belonging to the discontinues lung sounds class under sound pollution conditions.
4. All models used as an aid to the diagnosis of lung sounds in the clinical work of healthcare professionals should be evaluated at different ambient noise levels.
5. The machine learning model diagnostic assistant integration with healthcare worker performing auscultation should always take into account the specialist performance for that particular sound under certain ambient noise conditions, this should be weight against machine learning models performance under the same conditions and once diagnosis assistance is provided by the model it could do so in a context of this information.

SANTRAUKA

1. SUTRUMPINIMAI

AdaBoost	–	adaptacinis stiprinimas (mašininio mokymosi modelis) (angl. *Adaptive Boosting*)
ANN	–	dirbtinis neuroninis tinklas (angl. *artificial neural network*)
AIF	–	asmens informavimo forma
boxplot	–	dėžutės ir ūsų pobūdžio diagrama (ang. *boxplot diagram*)
BT	–	baltasis triukšmas
CatBoost	–	kategorijų gradientinis stiprinimo modelis (mašininio mokymosi modelis) (angl. C*ategorical data Gradient Boosting*)
DAK	–	drėgni auskultaciniai karkalai
dB	–	decibelai
DI	–	dirbtinis intelektas
ET	–	„Papildomų medžių“ mašininio mokymosi modelis (angl. *Extra Trees*)
el. stetoskopas	–	elektroninis stetoskopas
GB	–	Gradientinio stiprinimo mašininio mokymosi modelis (angl. *Gradient Boosting*)
GBT	–	Gausso baltas triukšmas
Histgradient	–	Histogramomis pagrįsto gradientinio stiprinimo klasifikacinio medžio mašininio mokymosi modelis (angl. *Histogram-based Gradient Boosting Classification Tree*)
Hz	–	hercai
IFN	–	inkstų funkcijos nepakankamumas
ISD	–	informuoto sutikimo dokumentas
K-NN	–	K-artimiausi kaimynai (mašininio mokymosi modelis) (angl. *K-Nearest Neighbors*)
KNT	–	Konvoliucinis neuroninis tinklas
LightGBM	–	lengvas gradientinis stiprinimo mašinio mokymosi modelis (angl. *Light Gradient Boosting Machine*)
LIL	–	lėtinė inkstų liga
LOPL	–	Lėtinė obstrukcinė plaučių liga
LR	–	logistinės regresijos mašininio mokymosi modelis (angl. *Logistic Regression*)
MF	–	Medicinos fakultetas
MFS	–	Medicinos fakulteto studentas
MKK	–	Motiejaus koreliacijos koeficientas
MLP	–	daugiasluoksnis perceptronas (mašininio mokymosi modelis) (angl. *Multilayer Perceptron*)
MM	–	mašininis mokymasis
ms	–	milisekundės
NAG	–	normalūs auskultaciniai garsai
OI	–	organinis intelektas
PSO	–	Pasaulio sveikatos organizacija

RF	–	atsitiktinių medžių tipo mašininio mokymosi modelis (angl. *Random Forest*)
ROC	–	sprendimus priimančiojo ypatybių kreivė (angl. *Receiver Operating Characteristic*)
ROC-AUC	–	plotas po sprendimus priimančiojo ypatybių kreive (angl. *Receiver Operating Characteristic Area Under the Curve*)
s	–	sekundės
SAK	–	sausi auskultaciniai karkalai
SITS	–	signalų ir triukšmo santykis
SN	–	standartinis nuokrypis
SPS	–	Skubios pagalbos skyrius
SVM	–	palaikymo vektorių mašininio mokymosi modelis (angl. *Support Vector Machines*)
ŠN	–	širdies nepakankamumas
XGBoost	–	Ekstremalus gradientinis stiprinimo klasifikatorius (mašininio mokymosi modelis) (angl. *Extreme Gradient Boosting classifier*)

2. ĮVADAS

Stetoskopas sveikatos priežiūros specialistų klinikiniame darbe kasdien naudojamas daugiau nei 200 metų, tačiau šio prietaiso naudojimas vis dar apribotas tyrėjo subjektyviu gebėjimu efektyviai atlikti auskultaciją ir ją vertini bei aplinkos triukšmo lygiu [1–3].

Pastebėta, jog pastarąjį dešimtmetį krito kardiopulmoninės auskultacijos atlikimo dažnis ir šių auskultacijų interpretavimo lygis [4, 5]. Plaučių auskultacija išlieka vis dar svarbiausias iš keturių plaučių sistemos klinikinio ištyrimo metodų. Plaučių ligos – trečia pagal dažnumą mirties priežastis visame pasaulyje [6, 7], todėl geri plaučių auskultaciniai įgūdžiai, kaip pradinis, greitas ir efektyvus, neinvazinis ištyrimo metodas klinikiniame sveikatos priežiūros specialistų darbe aktualus ir būtinas.

Mokslas ir inžinerija nestovi vietoje, dešimtmečius kuriami bei tobulinami elektroniniai stetoskopai (el. stetoskopai), būtent tai leido plėtoti kompiuterinę auskultaciją [8, 9].

Naujausi pokyčiai mikroschemų industrijoje, skaičiavimo galios spartos progresas, paremtas Moore‘o dėsniu, kartu su patobulintais matematiniais modeliais lėmė vis didesnius proveržius ir mašininio mokymosi (MM) priemonių taikymą diagnostikos srityje [11–13].

Sinerginis elektroninių stetoskopų ir dirbtinio intelekto (DI), o konkrečiau MM modelių, derinys išryškėjo kaip galimas sprendimas, siekiant pagerinti plaučių auskultacijos diagnostinį tikslumą [6].

Tačiau yra labai nedaug straipsnių, kuriuose būtų lyginamas žmonių auskultacijų interpretavimo tikslumas su dideliu skaičiumi MM modelių gebėjimu interpretuoti auskultacinius duomenis.

Išlieka esminis klausimas, kaip ir koks MM modelis galėtų tapti pagalbine priemone gydytojui, esant apsunkintoms klinikinės auskultacijos sąlygoms, t. y. esant įvairiems aplinkos triukšmo lygiams. Neatsakius į šį klausimą, šių priemonių integracija į kasdienį klinikinį darbą problematiška, o netinkamų ar nepritaikytų interpretuoti esant garso užterštumui MM modelių naudojimas gali sukelti daugiau problemų nei sprendimų, kuriuos jie turėtų išspręsti.

Plaučių garsai skiriasi ir yra klasifikuojami du pagrindiniai auskultacijos garsų tipai: normalūs (NAG) ir patologiniai. Patologiniai auskultaciniai garsai gali būti sausi auskultaciniai karkalai (SAK) ir drėgni auskultaciniai karkalai (DAK). Drėgnų karkalų savybės išreiškiamos kaip smulkūs ir grubūs traškėjimai, o sausų karkalų garsai gydytojui girdimi kaip švilpimai ir bronchų garsai. Tipiniai sausi patologiniai garsai paprastai yra nuo 80 iki 1600 Hz, trunka ilgiau nei 250 ms ir yra susiję su astma, lėtine obstrukcine plaučių liga. Drėgni patologiniai plaučių garsai yra trumpesni, paprastai trunka mažiau nei 20 ms, turi plačią dažnių ribą nuo 100 iki 2000 Hz ir yra susiję su širdies nepakankamumu (ŠN), pneumonija [14].

Ši disertacija gilinasi į tai, kaip anotuoti ir kontroliuojami MM modeliai geba atpažinti tris skirtingas plaučių garsų klases esant trims skirtingiems aplinkos triukšmo lygiams ir lygina šių modelių klaidų matricos parametrus su žmonių gebėjimais, naudojant tą patį auskultacinių duomenų rinkinį.

Šio darbo sukauptos žinios turėtų prisidėti prie žinių pažangos, siekiant ateityje integruoti ir plėtoti ekonomiškai efektyvius, neinvazinius, kokybiškus ir objektyvius pirminio ištyrimo sprendimus, kurie galėtų būti taikomi ambulatorinėje kvėpavimo sistemos ištyrimo ir stebėjimo srityje [15–17].

3. TIKSLAS IR UŽDAVINIAI

3.1. Tikslas

Įvertinti ir palyginti mašininio mokymosi modelių ir medicinos studentų diagnostinį tikslumą, teisingai identifikuojant tris auskultacinių plaučių garsų klases, esant trims skirtingiems Gausso baltojo triukšmo (GBT) lygiams.

3.2. Uždaviniai

1. Išmokyti mašininio mokymosi modelius bei medicinos studentus identifikuoti tris auskultacinių plaučių garsų klases ir įvertinti jų gebėjimą identifikuoti jas skirtinguose GBT lygiuose.
2. Įvertinti spektrogramos ir skalogramos įtaką 12 skirtingų anotuotų mašininio mokymosi modelių gebėjimui tiksliai identifikuoti skirtingas plaučių garsų klases.
3. Palyginti mašininio mokymosi modelių ir medicinos studentų gebėjimą identifikuoti tris plaučių garsų klases, esant trims skirtingiems GBT lygiams, naudojant pagrindinę diagnostinę metriką.
4. Nustatyti mašininio mokymosi modelio galimybes veikti kaip diagnostiniam pagalbininkui GBT sąlygomis, identifikuojant tris pagrindines plaučių garsų klases.

3.3. Darbo naujumas

Tyrimo projektas yra unikalus pasaulyje. Pirmas tokio pobūdžio, palyginantis žmogiškųjų tyrimo subjektų ir mašininio mokymosi modelių gebėjimus identifikuoti tris plaučių garsų klases trijose GBT lygio sąlygose.

Šiuo metu nėra atliktų tyrimų, kurie lygintų 12 MM modelių rezultatus su žmonių gebėjimais, naudojant tuos pačius plaučių garsų duomenų rinkinius. Tyrimų, nagrinėjančių žmogaus gebėjimą atpažinti plaučių garsus skirtingomis triukšmo lygio sąlygomis yra nedaug.

Moksliniai straipsniai, kurie tyrinėtų aplinkos triukšmo poveikį MM modelių tikslumui yra taipogi reti, o kai kurie daugiau nei dešimties metų senumo, tuo tarpu per šį laiką MM modeliai tobulėjo, atsirado naujų įrankių, kurie dar nebuvo išbandyti minėtose sąlygose. Todėl, pasitelkiant kelis skirtingus MM modelius ir dvi garso atvaizdavimo formas, galima pateikti naujų įžvalgų apie tai, kurie modeliai galėtų būti atspariausi triukšmo poveikiui ir kaip jie galėtų būti naudojami sprendimų priėmimui. Tyrimai apie žmonių gebėjimą atpažinti skirtingų klasių plaučių garsus taip pat yra menki, senesni nei 5 metų, atlikti įvairiose aplinkose, jų pernelyg neklasifikuojant arba auskultaciniai duomenys rinkti vaikų populiacijoje [30, 31].

Tyrimų straipsniai netgi pateikia prieštaringas išvadas, pavyzdžiui, kad daugumos tyrėjų gebėjimas girdėti širdies ir plaučių garsus nėra reikšmingai paveiktas ekstremalaus aplinkos triukšmo garsumo, kuris pasitaiko skubios pagalbos skyriuose [31].

2019 m. Rory Wallis apžvalgos straipsnis padarė išvadą, kad aplinkos triukšmo lygio matavimai ligoninėse yra netikslūs ir nestandartizuoti [32]. Aukščiau minėti faktoriai apsunkina hipotezių tikrinimą, bandant pakartoti metodiką. GBT savybės svarbios standartizuojant triukšmo lygį, kadangi jis

tolygiai apima visas dažnių juostas. Todėl GBT panaudojimas, kaip standartizuoto aplinkos triukšmo, vertinant žmonių ir MM mokymosi auskultacinių garsų atpažinimo rezultatus yra naujo pobūdžio tyrimas.

Tuo tarpu tyrimų, kurie nagrinėja MM modelius skirtingomis aplinkos triukšmo sąlygomis, taip pat yra labai mažai ir literatūros apžvalgoje galima rasti tik tris straipsnius [10, 30, 33]. Be to, kai kurie tyrimai neturi statistiškai reikšmingo duomenų kiekio, kad būtų galima atlikti statistinę analizę [10].

Šiuolaikinių MM modelių pritaikymas auskultacijų interpretavimui, esant standartizuotam GBT, trijuose skirtinguose triukšmo lygiuose bei minėtų MM modelių rezultatų palyginimas su žmogiškųjų tyrimo subjektų gebėjimu interpretuoti tą patį auskultacinių duomenų rinkinį, tomis pačiomis standartizuotomis garso užterštumo sąlygomis paverčia šį tyrimą visiškai unikaliu.

4. METODIKA

4.1. Tyrimo dizainas, tyrimo vieta

Prospektyvusis tyrimas buvo atliktas Lietuvoje 2020–2024 m.

Tyrimo dalyviai:

Auskultacinių duomenų rinkimo etape tyrimo dalyviai: pacientai, hospitalizuoti dėl diagnozuotos pneumonijos, širdies nepakankamumo (ŠN), lėtinės obstrukcinės plaučių ligos (LOPL), astmos, inkstų nepakankamumo, lėtinės inkstų ligos (LIL) arba hidrotorakso, remiantis Lietuvos sveikatos mokslų universiteto Kauno ligoninės protokolais [95–104], kuriems pirminio klinikinio tyrimo duomenimis buvo nustatyti patologiniai ir nepatologiniai plaučių garsai.

Žmogiškųjų tyrimo subjektų auskultacijų interpretacijos apmokymo etape tyrimo dalyviai: savanoriai II ir III kurso LSMU medicinos studijų studentai, iki tyrimo neturėję auskultacinių įgūdžių.

Tyrimo vieta plaučių garsų (auskultacinių duomenų) rinkimo etape: tyrimas buvo atliktas Lietuvos sveikatos mokslų universiteto Kauno ligoninės Kardiologijos ir Vidaus ligų diagnostikos skyriuose (Josvainių g. 2 ir Hipodromo g. 13, Kaunas).

Tyrimo vieta medicinos studentų auskultacijų mokymo etape: Lietuvos sveikatos mokslų universiteto, Vidaus ligų katedra (Josvainių g. 2, Kaunas).

Tyrimas dalinai atliktas bendradarbiaujant su Kauno technologijos universiteto profesoriumi Evaldu Vaičiukynu ir jo kolegomis, remiant Kauno technologijos universiteto (dotacijos Nr. PP2023/39/4) ir Lietuvos sveikatos mokslų universiteto švietimo ir mokslo fondams.

4.2. Imties dydžio apskaičiavimas

Patologinių ir nepatologinių plaučių garsų turinčių pacientų (auskultacinių duomenų rinkimo etape) ir auskultacijų apmokymo etapo studentų imties dydžiai buvo apskaičiuoti naudojant G*Power programinę įrangą (versija 3.1.9.4; Heinrich-Heine-Universität Düsseldorf, Düsseldorfas, Vokietija) [106, 107].

Medicinos studentų imties dydžio skaičiavimai buvo grindžiami prieš tai atliktu pilotiniu tyrimu. Programinė įranga naudojo šiuos nustatymus vidurkiams apskaičiuoti: Wilcoxon suporuotų rangų testas (sulygintos poros) funkcija. Buvo taikomos šios prielaidos: galia (1 − β klaidos tikimybė) – 0,95 ir α klaidos tikimybė – 0,05. Efekto dydis (Cohen dz) iš pilotinio tyrimo buvo 0,61, pagal reikšmes prieš ir po mokymo bei standartinius nuokrypius (SD), kurie buvo atitinkamai 4,80 ± 0,49 ir 5,07 ± 0,36. Šios reikšmės buvo įvestos į funkciją, o rezultatas parodė, kad reikalinga 33 dalyvių imtis. Pilotiniame tyrime buvo pridėta papildomai 30 proc. daugiau turimųjų dėl atkritimo galimybės. Todėl, atsižvelgiant į numanomą tiriamųjų atkritimo tikimybę, šiam tyrimui reikalingas bendras dalyvių skaičius apskaičiuotas buvo 48.

Plaučių garsų pacientų imties dydis buvo apskaičiuotas remiantis prielaida, kad efekto dydis bus 0,50, galia (1 − β klaidos tikimybė) – 0,95 ir α klaidos tikimybė – 0,05, o grupių skaičius buvo nustatytas 3. G*Power programinės įrangos (versija 3.1.9.4; Heinrich-Heine-Universität Düsseldorf, Düsseldorfas, Vokietija) funkcija buvo nustatyta kaip ANOVA: fiksuotas efektas. Įvestis davė rezultatą – 85 dalyviai (įskaitant kontrolinę grupę). Plaučių garsų įrašai turėjo būti peržiūrėti taikant „double-blind“ metodą, darant prielaidą, kad apie 30 proc. atrinktųjų nebus tinkami, o tai reiškia, kad į tyrimą turėjo būti įtraukta apie 122 dalyviai.

4.3. Įtraukimo ir neįtraukimo kriterijai

Įtraukimo kriterijai – pacientams, plaučių garsų įrašymui:

1. Pacientas, kuriam diagnozuota pneumonija;
2. Pacientas, kuriam diagnozuota astma;
3. Pacientas, kuriam diagnozuotas širdies nepakankamumas (ŠN);
4. Pacientas, kuriam diagnozuotas inkstų nepakankamumas (IFN);
5. Pacientas, kuriam diagnozuota lėtinė obstrukcinė plaučių liga (LOPL) paūmėjimas;
6. Pacientas, turintis papildomų plaučių garsų;
7. Pacientas, turintis normalius plaučių auskultacinius garsus;
8. Pacientas, vyresnis nei 18 metų;
9. Pacientas, neturintis psichikos sutrikimų;

10. Pacientas buvo sąmoningas ir galėjo teisingai atsakyti į klausimus;
11. Pacientas, sutinkantis savanoriškai dalyvauti ir pasirašęs informuoto sutikimo formą.

Įtraukimo kriterijai – medicinos studentams, auskultacijų apmokymams:
1. LSMU medicinos studentas, esantis antrame ar trečiame kurse;
2. Dalyvis, vyresnis nei 18 metų;
3. Dalyvis, neturintis ankstesnės auskultacijos patirties;
4. Sutinkantis savanoriškai dalyvauti ir pasirašęs informuoto sutikimo formą.

Neįtraukimo kriterijai – pacientams, plaučių garsų įrašymui:
1. Pacientai, atsisakę dalyvauti tyrime;
2. Pacientai, kurie negalėjo kalbėti lietuviškai ir suteikti sutikimo;
3. Pacientai, kurie negalėjo stovėti ar sėdėti ramiai, kad būtų atlikta auskultacija.

Neįtraukimo kriterijai – medicinos studentams, auskultacijų apmokymams:
1. Studentai, turintys klausos sutrikimų;
2. Studentai, vyresni nei 40 metų;
3. Studentai, kurie atsisakė dalyvauti arba nepasirašė sutikimo formų.

4.4. Tyrimo metodika

Siekiant palygti medicinos studentų ir mašininio mokymosi (MM) modelių rezultatus, metodika buvo suskirstyta į keletą mažesnių užduočių.

Pirma užduotis buvo įrašyti plaučių garsus ir juos apdoroti mokymo ir mokymosi tikslais.

Pacientų auskultaciniai įrašai buvo atlikti per maždaug tris mėnesius (neįskaitant pertraukų). Elektrinio stetoskopo nustatymai: režimas nustatytas į diafragmą, o garso stiprinimas – į 3 lygį (maksimalus lygis – 9). Tyrėjas įrašus atlikdavo palatose, kuriose paprastai būdavo nuo 2 iki 4 pacientų. Buvo naudojamas 3M™ Littmann® CORE skaitmeninis stetoskopas (3M Company, St Paul, Minesota, JAV), HP ProBook 450 G4 nešiojamas kompiuteris (HP Inc., Palo Alto, Kalifornija, JAV) su „Microsoft® Windows® 10“ operacine sistema (Microsoft Corporation, Redmondas, Vašingtonas, JAV) ir Intel® Core™ i5 i5-7200U procesoriumi (Intel Corporation, Santa Clara, Kalifornija, JAV), skirtas garso failams saugoti naudojant 3M™ Littmann® StethAssist – 1.3.230 programinę įrangą (3M Company, St Paul, Minesota, JAV).

Norint įvertinti medicinos studentų ir MM modelių atsparumą skirtingiems signalo ir triukšmo santykio (SITS) lygiams, kiekvienam įrašui buvo pridėtas Gauso baltasis triukšmas (GBT) pagal Samit Ari metodiką (108). Trijų lygių GBT: be GBT, GBT su SITS-40 ir GBT su SITS-20. GBT buvo pridėtas naudojant „Anaconda®“ (Austinas, Teksasas, JAV) su „Jupyter Notebook 6.4.7“ ir Python paketais mašininio mokymosi treniravimui ir vertinimui. Garso ypatybės buvo išgautos naudojant Python biblioteką ir išsaugotos CSV formatu.

Antra užduotis buvo apmokyti MM modelius plaučių garsų interpretacijos.

Iš viso buvo pasirinkta 12 anotuotų mašininio mokymosi modelių: AdaBoost, CatBoost, Extra Trees (ET), gradient boosting (GB), Histgradient, K-NN, LightGBM, logistinė regresija, MLP, Random Forest, SVM, XGBoost. Šie modeliai buvo pasirinkti dėl jų potencialo ankstesniuose plaučių garsų ar kitų klausos biosignalų diagnostikos tyrimuose bei galimybės taikyti mažesniems duomenų rinkiniams. Modeliai buvo mokomi naudojant metodiką, kuri išskiria ypatybes iš skalogramų ir spektrogramų [111]. Mokymas buvo atliktas specialiai tyrimui sukomplektuotame kompiuteryje su „Windows® 10“ operacine sistema (Microsoft Corporation, Redmondas, Vašingtonas, JAV), kuris buvo aprūpintas Intel® Core™ i7-12700K procesoriumi, 64 GB RAM ir NVIDIA GeForce RTX 3060 vaizdo plokšte su 12 GB VRAM (NVIDIA Corporation, Santa Clara, Kalifornija, JAV).

Duomenų rinkinys buvo padalintas santykiu 80/20 mokymui ir testavimui [112]. Padalintuose duomenyse buvo proporcingai paskirstyti NAG, SAK ir DAK plaučių garsai trijuose skirtinguose GBT lygiuose (be GBT, GBT SITS-40 lygyje, GBT SITS-20 lygyje). MM metu mokymo duomenys buvo suskirstyti į devynias dalis, kad būtų užtikrintas panašus tikslo klasių pasiskirstymas kiekvienoje dalyje ir pagerintas MM modelių veikimas.

Trečia užduotis buvo įvertinti MM modelių gebėjimą atpažinti tris plaučių garsų klases esant trims skirtingiems GBT lygiams. Kiekvienos dalies veiklos metrika buvo surinkta ir suskaičiuota vidutinė vertė, kad būtų pateiktas geriausias modelio veiklos įvertinimas. Iš viso buvo atlikta 30 iteracijų (paleidimų) kiekvienam modeliui, įskaitant klasių disbalanso valdymą, kryžminės patikros atlikimą, modelių mokymą ir veiklos metrikų skaičiavimą [114]. Pasirinkus geriausią rezultatą gavusį MM modelį iš 24 kurtų variantų (12 MM modelių, pagrįstų spektrogramomis, ir 12 – skalogramomis), modelis buvo dar kartą patikslintas, o vidutiniam MMC skaičiavimui buvo atlikti 45 paleidimai.

Ketvirta užduotis buvo įtraukti medicinos studentus ir juos apmokyti plaučių auskultacinių garsų interpretavimo, naudojant specialiai sukurtą mokymo (-si) svetainę.

Svetainė, sukurta su mokymo ir testavimo skyriais bei buvo sėkmingai panaudota ankstesniame tyrime [109]. Mokymo skyriuje buvo pateikti normalūs ir patologiniai auskultaciniai plaučių garsai bei jų savybių apibūdinimas žodžiais. Mokymo skyriuje buvo 101 plaučių garsų įrašas, iš kurių 54 proc. buvo DAK ir SAK. Testavimo skyrius buvo sudarytas atsitiktiniu būdu ir apėmė 54 garso įrašus, susidedančius proporcingai iš NAG, SAK ir DAK klasių plaučių garsų. Prieš inicijuojant tyrimą svetainė buvo išbandyta su 15 studentų pilotinėje studijoje, siekiant įvertinti tinklalapio funkcionalumą, efektyvumą ir pašalinti galimus tinklalapio veikimo nesklandumus, o surinkti duomenys panaudoti imties dydžio skaičiavimui. Papildomai svetainę peržiūrėjo gyd. pulmonologas dėl kokybės užtikrinimo. Į galutinį tyrimą buvo įtraukti 52 antrojo ir trečiojo kurso medicinos fakulteto studentai (MFS), atitinkantys įtraukimo kriterijus ir pateikę informuotą sutikimą.

Penkta užduotis buvo įvertinti studentų gebėjimą atpažinti tris plaučių garsų klases esant trims skirtingiems GBT lygiams.

Po 4 dienų mokymo studentai buvo išbandyti, ar geba teisingai atpažinti NAG, SAK ir DAK, atlikdami 3 testus, kurių kiekvienas turėjo skirtingus GBT lygius (be GBT, GBT SITS-40 lygyje, GBT SITS-20 lygyje). Vertinimas buvo atliktas toje pačioje svetainėje, tinklalapio testavimo skyriuje.

Šešta užduotis buvo įvertinti galimą skirtingą GBT poveikį medicinos studentų ir geriausių MM modelių gebėjimui atpažinti skirtingas plaučių garsų klases.

Galiausiai, septinta užduotis buvo taikyti Friedmano testą su poriniu palyginimu, kad būtų palyginti geriausio MM modelio ir medicinos studentų MKK reikšmės visuose skirtinguose GBT lygiuose ir visose trijose plaučių garsų klasėse. Rezultatų statistinis reikšmingumo lygmuo vertintas, esant $p < 0,05$.

4.5. Statistinė analizė

Duomenų analizė MM modelių ir LSMU studentų auskultacijų interpretavimo rezultatams įvertinti buvo atlikta naudojant „Microsoft® Excel®“ (Microsoft Corporation) skaičiuoklę ir JASP (ver. 0.18.3; Jeffreys’ Amazing Statistics Programme, The Jamovi project, Sidnėjus, Australija) statistikos paketą [126]. Bei IBM® SPSS® ver. 29 (IBM Inc., Armonkas, Niujorkas, Jungtinės Amerikos Valstijos). p reikšmė, mažesnė arba lygi 0,05, buvo laikoma statistiškai reikšminga. Rezultatai buvo pateikti lentelėse ir apibendrinti „dėžutės ir ūsų“ pobūdžio (boxplot) diagramose.

Atliekant duomenų valymą, septyni subjektai buvo pašalinti iš tolimesnės statistinės analizės, nes jie nesugebėjo užbaigti visų trijų testavimų. Todėl statistinė analizė buvo atlikta 45 iš 52 subjektų.

Rezultatai neatitiko normalaus pasiskirstymo, todėl tolesniam vidurkių analizavimui buvo naudojami neparametriniai testai. Wilcoxono rangų sumažinimo testas įvertino mokymų poveikį studentų gebėjimui tiksliai atpažinti plaučių garsus, o Friedmano testas buvo naudojamas analizuoti trijų GBT lygių poveikį skirtingų plaučių garsų klasių identifikavimui su dviem laisvės laipsniais. Galiausiai buvo atlikta *post hoc* palyginimo analizė, kad būtų įvertintas medicinos studentų gebėjimas atpažinti plaučių garsų klases (NAG, SAK ir DAK) atskirai pagal tris skirtingus GBT lygius.

Studentų garsų interpretavimo testų rezultatai iš mokomosios/testavimo svetainės surinkti naudojant MongoDB® (MongoDB, Inc., Niujorkas, JAV) programinę įrangą ir įrašyti į „Microsoft® Excel®“ (Microsoft Corporation) skaičiuoklę tolimesnei statistinei analizei.

MM modelių veikimas buvo įrašytas naudojant Anaconda® (Austin, TX, JAV) su Jupyter Notebook 6.4.7, naudojant „Python“ paketų mašininio mokymosi mokymui ir vertinimui, ir išsaugotas CSV formatu.

Norint palyginti mašininio mokymosi įrankius, buvo atliktas Friedmano testas su *post hoc* porinių palyginimų analize, siekiant palyginti 24 skirtingų MM modelių variacijų diagnostinį tikslumą. Norint palyginti 12 spektrogramų ir 12 skalogramų pagrįstų MM modelių našumą, buvo naudojamas Wilcoxono pasirašytų rangų testas. $p < 0,05$ buvo laikoma statistiškai reikšminga.

5. REZULTATAI

5.1. Tyrimo populiacijos charakteristikos

5.1.1 lentelė. *Aprašomoji lentelė, kurioje pateikiama populiacija, iš kurios buvo gauti plaučių garsai tyrimui*

Plaučių garsai	Moterys	Moterų amžius (SN)	Vyrai	Vyrų amžius (SN)	Bendras skaičius	Bendras amžius (SN)
NAG	26	69.5 (16.9)	26	56.5 (18.6)	52	63.0 (18.0)
SAK	10	75.5 (8.4)	13	66.0 (12.2)	23	70.1 (11.5)
DAK	12	78.7 (12.3)	21	69.0 (11.7)	33	72.5 (12.7)
Bendras	48	73.1 (14.7)	60	62.9 (16.0)	108	67.4 (16.2)

SN – standartinis nuokrypis.

5.1.2 lentelė. *Aprašomoji lentelė, tyrime dalyvavusių medicinos studentų lyties ir amžiaus analizė*

Moterys	Moterų amžius (SN)	Vyrai	Vyrų amžius (SN)	Bendras skaičius	Bendras amžius (SN)
32	21.9 (2.4)	13	21.6 (3.1)	45	21.8 (2.6)

SN – standartinis nuokrypis.

5.2. Mašininio mokymosi modelių efektyvumo analizė

Iš viso 24 mašininių modelių variantai buvo išbandyti naudojant spektrogramas ir skalogramas, esant trims GBT triukšmo lygiams (be pridėtinio triukšmo, GBT SITS-40 ir GBT SITS-20). Poveikis buvo stebimas trims pagrindinėms plaučių garsų klasėms: NAG, SAK, DAK.

Modelio efektyvumo palyginimui pagrinde buvo naudojamas „receiver operating characteristic area under the curve“ – plotas po sprendimus priimančiojo ypatybių kreive (ROC-AUC). Visi modeliai buvo tikrinami dėl bendro GBT poveikio jų veikimui taikant Friedmano testą.

5.2.1 lentelė. *Dvylikos spektrogramų pagrįstų modelių našumas pagal ROC-AUC balus*

Spektrogramomis pagrįstas modelis	ROC-AUC, mediana (IQR)	Testo statistika	Laisvės laipsniai	p
AdaBoost	0.800 (0.689–0.853)	803	11	< 0.001
CatBoost	0.857 (0.764–0.880)			
Extra Trees	0.820 (0.691–0.859)			
Gradient Boosting	0.874 (0.772–0.897)			
Histgradient	0.865 (0.802–0.894)			
K-NN	0.751 (0.638–0.753)			
LightGBM	0.856 (0.782–0.879)			
Logistic Regression	0.863 (0.781–0.876)			
MLP	0.863 (0.786–0.902)			

5.2.1 lentelės tęsinys

Spektrogramomis pagrįstas modelis	ROC-AUC, mediana (IQR)	Testo statistika	Laisvės laipsniai	p
Random Forest	0.833 (0.694–0.873)	803	11	< 0.001
SVM	0.836 (0.746–0.853)			
XGBoost	0.871 (0.782–0.895)			

Geriausiai spektrogramomis pagrįsti algoritmai pagal rezultatų medianą buvo visi stiprinimo modeliai: Gradient Boosting ir XGBoost, Histgradient. Histgradient buvo antras pagal vidurkį MM modelis, tačiau jo siauresnis tarpkvartilinis intervalas su didžiausiu Q1 kvartiliu iš trijų geriausiai pasirodžiusių MM modelių. IQR – tarpkvartilis (interkvartilis). ROC-AUC (angl. *Receiver Operating Characteristic Area Under the Curve*) – plotas po sprendimus priimančiojo ypatybių kreive.

5.2.2 lentelė. *Dvylikos skalogramų pagrįstų modelių našumas pagal ROC-AUC balus*

Skalograma pagrįstas modelis	ROC-AUC, mediana (IQR)	Testo statistika	Laisvės laipsniai	p
AdaBoost	0.735 (0.658–0.847)	574	11	< 0.001
CatBoost	0.794 (0.679–0.881)			
Extra Trees	0.746 (0.590–0.788)			
Gradient Boosting	0.752 (0.685–0.867)			
Histgradient	0.733 (0.671–0.850)			
K-NN	0.590 (0.528–0.658)			
LightGBM	0.732 (0.673–0.847)			
Logistic Regression	0.756 (0.671–0.814)			
MLP	0.741 (0.590–0.788)			
Random Forest	0.768 (0.635–0.808)			
SVM	0.740 (0.658–0.810)			
XGBoost	0.727 (0.659–0.859)			

Geriausiai pasirodę skalogramų pagrindu veikiantys MM algoritmai, remiantis medianiniu rezultatu, iš jų vienas buvo stiprinamojo (angl. *boosting)* tipo, vienas „papildomų medžių“ (angl. *extra trees*) ir vienas klasikinis modelis. MM modeliai kurie pateko į viršutinį veiklos kvartilį: CatBoost, Random Forest ir Logistic Regression. IQR – tarpkvartilis (interkvartilis). ROC-AUC (angl. *Receiver Operating Characteristic Area Under the Curve*) – plotas po sprendimus priimančiojo ypatybių kreive.

5.2.3 lentelė. *Dvylikos spektogramų ir dvylikos skalogramų modelių palyginimas pagal ROC-AUC balus*

12 MM spectrogramų mediana (IQR)	12 MM scalogramų mediana (IQR)	Testo statistika	p
0.837 (0.638–0.902)	0.735 (0.528–0.881)	583275	< 0.001

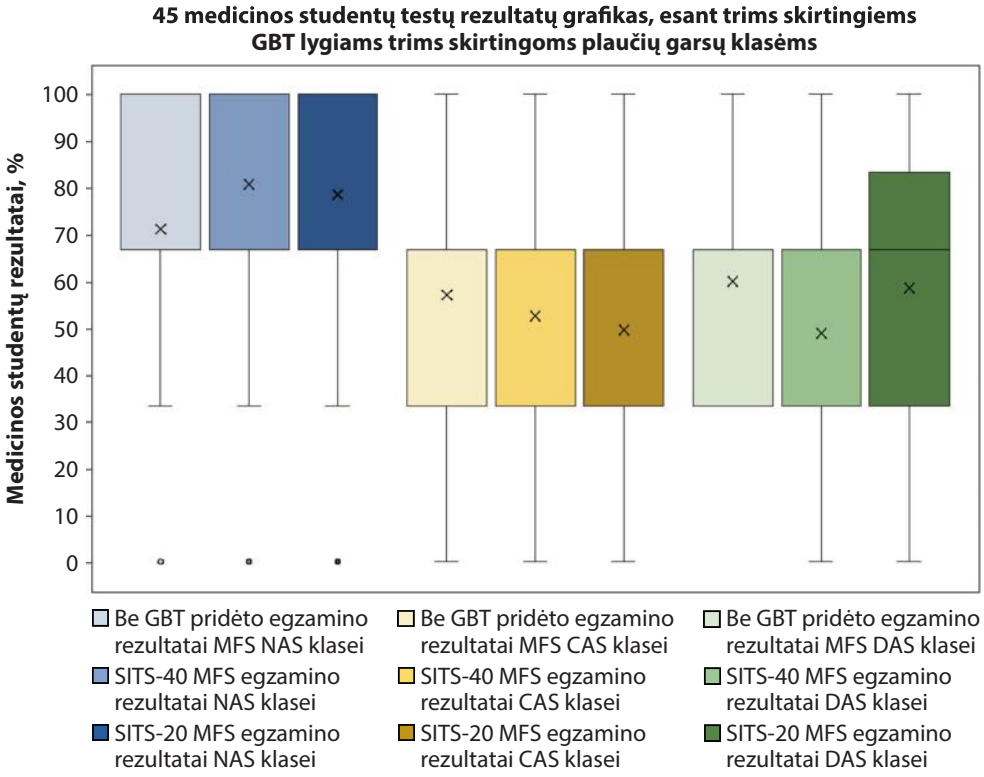
Wilcoxono testas rodo, kad 12 MM modelių, pagrįstų spektrogramomis ir skalogramomis, reikšmingai skiriasi, o spektrograma pagrįstų modelių medianos reikšmės yra daug didesnės, palyginti su skalograma pagrįstais modeliais. „IQR“ – tarpkvartilis (interkvartilis). „ROC-AUC“ (angl. *Receiver Operating Characteristic Area Under the Curve*) – plotas po sprendimus priimančiojo ypatybių kreive.

5.3. Medicinos fakulteto studentų veiklos rezultatai

Iš viso 45 medicinos studentai per 4 dienas bandė išmokti trijų klasių plaučių garsus ir atlikti testą esant trims GBT triukšmo lygiams (be papildomo triukšmo, GBT SNR-40 ir GBT SNR-20).

Boxplot diagramoje 5.3.1 pav. pavaizduota Medicinos studentų trijų klasių plaučių garsų atpažinimo testų rezultatai, kuriuose įvertinama trijų Gauso baltojo triukšmo lygių įtaka studentų gebėjimui atpažinti sausų auskultacinių karkalų (SAK), drėgnų auskultacinių karkalų (DAK) ir normalių auskultacinių garsų (NAG) plaučių garsų klases.

Friedmano testas parodė, kad gebėjimas atpažinti NAG ir DAK reikšmingai skyrėsi (atitinkamai p = 0,042, 0,021), esant trims GBT lygiams, o reikšmingo BT lygių poveikio SAK garsams nepastebėta (P = 0,311). siekiant įvertinti trijų GBT lygių įtaką gebėjimui atpažinti NAG ir DAK, atliktas *post hoc* palyginimas. Nustatyti statistiškai reikšmingi skirtumai atpažįstant plaučių garsus tarp be GBT ir SITS-40 NAG atveju, tarp be GBT ir SITS-40 bei tarp SITS-40 ir SITS-20 DAK atveju (atitinkamai p = 0,016, 0,013, 0,023).



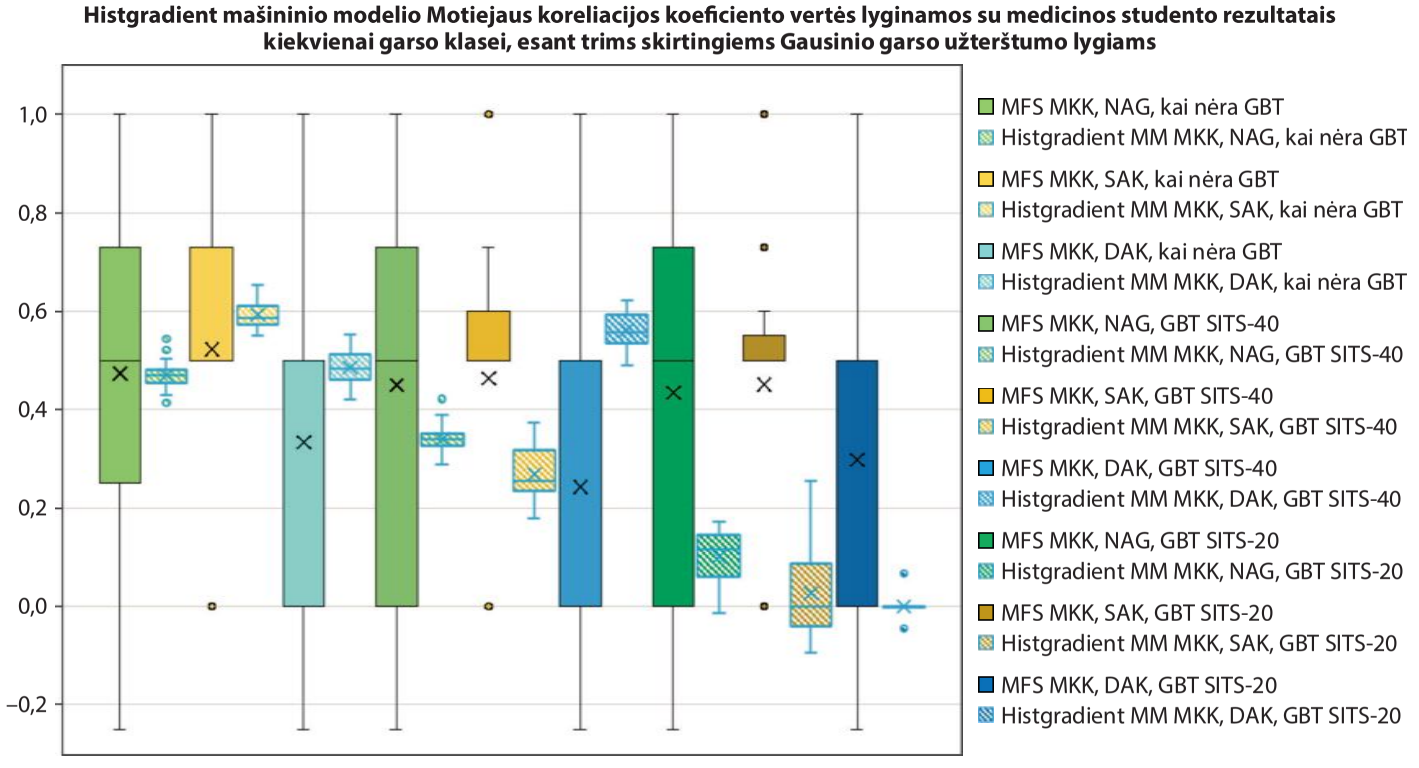
5.3.1 pav. *Medicinos studentų trijų klasių plaučių garsų atpažinimo testų rezultatai, esant skirtingiems GBT lygiams. Trijų Gauso baltojo triukšmo (GBT) lygių įtaka studentų gebėjimui atpažinti sausų karkalų (SAK), drėgnų karkalų (DAK) ir normalių plaučių garsų (NAG) klases*

GBT – Gauso baltasis triukšmas, MKK – Motiejaus koreliacijos koeficientas, SITS – signalo ir triukšmo santykis, NAG – normalūs auskultaciniai garsai, DAK – drėgni auskultaciniai karkalai, SAK – sausi auskultaciniai karkalai.

5.4. Geriausio MM modelio rezultatų palyginimas su medicinos studentų tikslumu

Studentų rezultatai buvo perskaičiuoti į tikras teigiamas, klaidingas teigiamas, tikras neigiamas, klaidingas neigiamas, šios reikšmės buvo naudojamos apskaičiuojant kiekvienos garso klasės MKK, jautrumą, specifiškumą, pagal kiekvieną GBT lygį (be GBT, GBT SITS-40 lygyje, GBT SITS-20 lygyje).

Gauti rezultatai buvo panaudoti nubraižant dėžutės ir ūsų diagramą (boxplot) ir atliekant Friedmano testą su *post hoc* analize (5.4.1 pav.).



5.4.1 pav. *Boxplot diagramoje pavaizduotas Histgradient mašininio mokymo ir medicinos fakulteto studentų (MFS) Motiejaus koreliacijos koeficientų palyginimas trims garsų klasėms (NAG, SAK, DAK) esant trims GBT lygiams*

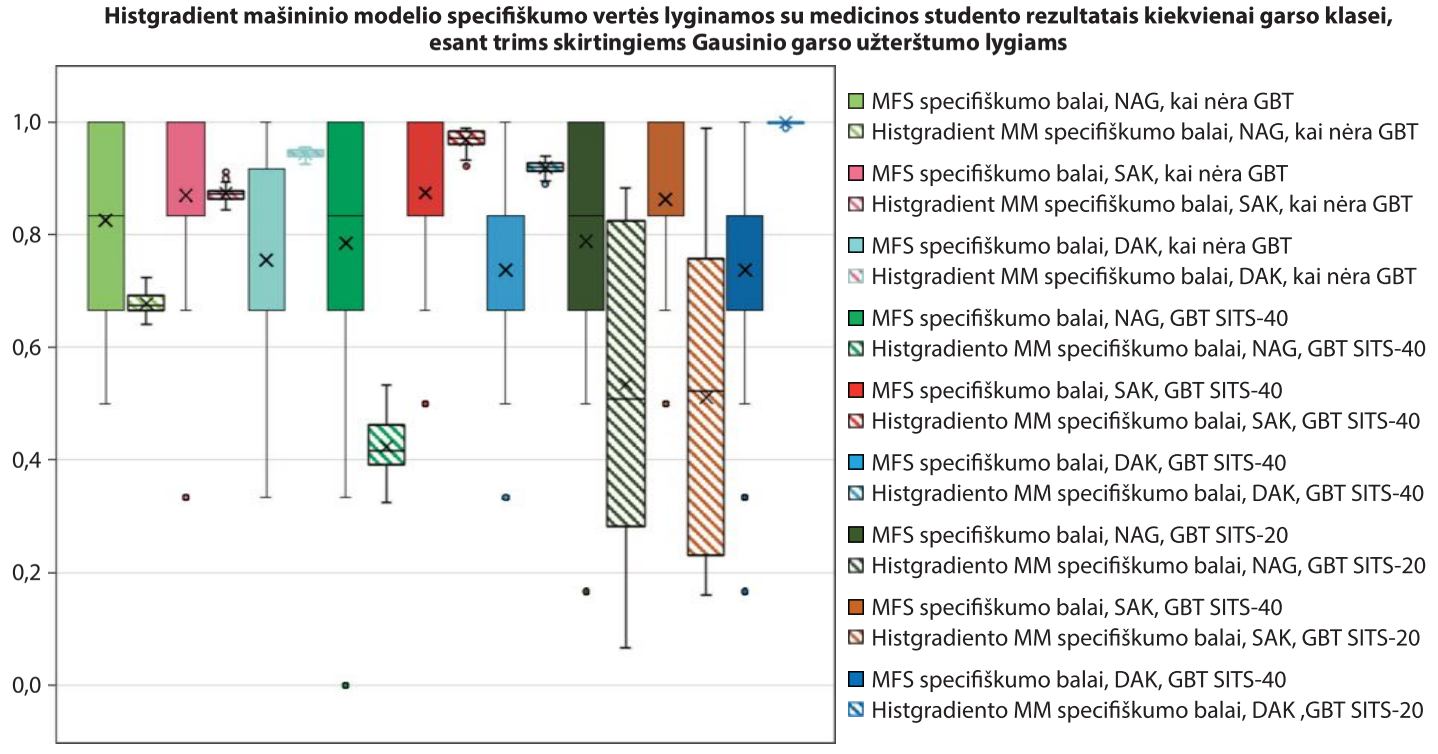
GBT – Gauso baltasis triukšmas, Histgradient – (angl. *Histogram-based Gradient Boosting Classification Tree*) Histogramomis pagrįsto gradientinio stiprinimo klasifikacinio medžio mašininio mokymosi modelis, SITS – signalo ir triukšmo santykis, NAG – normalūs auskultaciniai garsai, DAK – drėgni auskultaciniai karkalai, SAK – sausi auskultaciniai karkalai, MFS – Medicinos Fakulteto studentai, MKK – Motiejaus koreliacijos koeficientas.

Boxplot diagramoje (5.4.1 pav.) matomas skirtingų Gauso baltojo triukšmo (GBT) lygių poveikis Histgradient mašininio mokymosi (MM) modeliui ir Medicinos Fakulteto studentams (MFS) identifikuoti trijų klasių auskultacinius plaučių garsus, matuojamas Motiejaus koreliacijos koeficientas (MKK).

Studentų auskultacinių garsų atpažinimo rezultatai buvo panašūs į spektrogramomis pagrįsto Histgradient mašininio modelio rezultatus, kai į modelį nepridėta GBT, nes nepastebėta reikšmingų skirtumų tarp NAG, SAK, DAK ($p > 0,05$ visoms klasėms). Mašininio mokymosi modelio Motiejaus koreliacijos koeficiento rezultatai buvo 0,471 (0,415–0,543), 0,587 (0,522–0,654), 0,485 (0,422–0,552), lyginant su MFS Motiejaus koreliacijos koeficiento rezultatais 0,500 (–0,250–1,000), 0,500 (0,000–1,00), 0,500 (–0,250–1,000), 0,500 (–0,250–1,000) atitinkamai NAG, SAK, DAK auskultacinių garsų klasėse.

SITS-40 Gausinio užterštumo lygmenyje tarp visų trijų auskultacinių garsų klasių atpažinimo buvo statistinis reikšmingumas: NAG, SAK, DAK ($p = 0,035$, $p = 0,002$, $p = 0,000$), mašininio mokymosi Histgradient rezultatų balai buvo 0,341 (0,288–0,422), 0,256 (0,180–0,374), 0,557 (0,491–0,621), palyginus su medicinos fakulteto studentų gautais rezultatais 0,500 (–0,250–1,000), 0,500 (0,000–1,000), 0,000 (–0,250–1,000), 0,000 (–0,250–1,000) atitinkamai šiose garsų klasėse: NAG, SAK, DAK. Tuo tarpu medicinos fakulteto studentai parodė geresnius rezultatus atpažįstant normalius plaučių garsus (NAG) ir sausus karkalus (SAK), o MM Histgradient modelis, esant SITS-40 Gausinio užterštumo lygmeniui pranoko žmogiškųjų tiriamųjų subjektų (t. y. studentų) rezultatus geriau atpažįstant drėgnus auskultacinius karkalus (DAK).

Esant Gausinio užterštumo lygmeniui SITS-20, studentai parodė statiškai reikšmingai geresnius auskultacinių garsų atpažinimo rezultatus visose auskultacinių garsų klasėse, lyginant su Histgradient MM modeliu ($p = 0,000$ NAG ir SAK klasėms, $p = 0,009$ DAK klasei), rezultatai buvo 0.116 (–0,013–0,173), 0,001 (–0,095–0,255), 0,000 (–0,045–0,067), palyginti su studentų gautais rezultatais 0,500 (–0,250–1,000), 0,500 (0,000–1,000), 0,500 (–0,250–1,000), 0,500 (–0,250–1,000) atitinkamai NAG, SAK, DAK auskultacinių garsų klasėse.



5.4.2 pav. *Boxplot diagramoje pavaizduotas Histgradient modelio ir medicinos studentų specifiškumo palyginimas trims garsų klasėms (NAG, SAK, DAK) esant trims GBT lygiams*

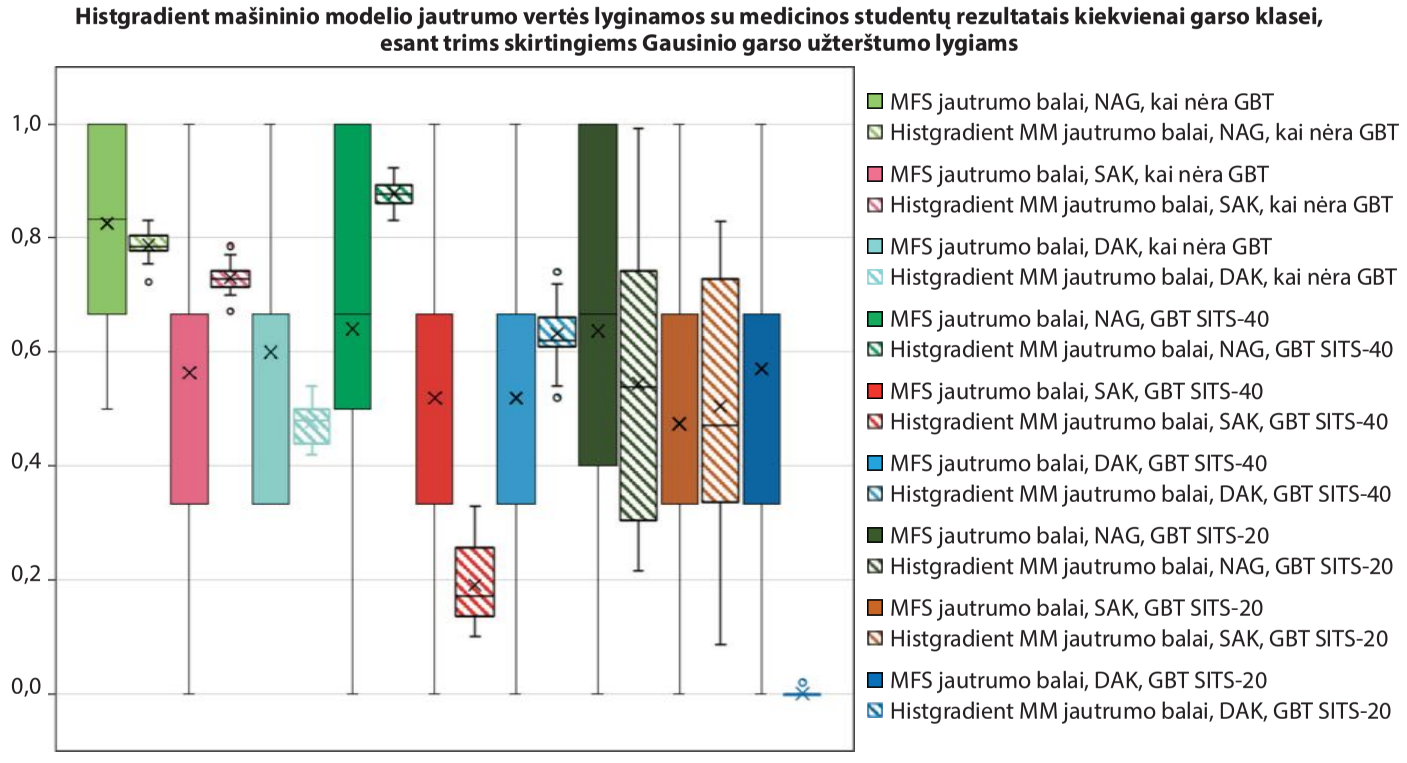
GBT – Gauso baltasis triukšmas, Histgradient – (angl. *Histogram-based Gradient Boosting Classification Tree*) Histogramomis pagrįsto gradientinio stiprinimo klasifikacinio medžio mašininio mokymosi modelis, SITS – signalo ir triukšmo santykis, NAG – normalūs auskultaciniai garsai, DAK – drėgni auskultaciniai garsai, SAK – sausi auskultaciniai garsai, MFS – medicinos fakulteto studentai.

Boxplot diagramoje (5.4.2 pav.) pavaizduotas Histgradient mašininio mokymo (MM) ir medicinos fakulteto studentų (MFS) specifiškumo palyginimas trims auskultacinių garsų klasėms (NAG, SAK, DAK), esant skirtingiems Gauso baltojo triukšmo (GBT) lygiams.

Kai į modelį nepridėta GBT, Histgradient modelio ir MFS specifiškumo pasiskirstymas NAG ir SAK auskultacinių garsų klasių atpažinimui, reikšmingai nesiskyrė ($p > 0,05$ abiem klasėms): Histgradient specifiškumas buvo 0,471 (nuo 0,415 iki 0,543) ir 0,587 (nuo 0,522 iki 0,654) NAG ir SAK klasėms, o MFS specifiškumas 0,500 (nuo –0,250 iki 1,000) ir 0,500 (nuo 0,000 iki 1,000) toms pačioms auskultacinių garsų klasėms. Tačiau drėgnų auskultacinių karkalų atveju Histgradient MM modelis parodė reikšmingai didesnį specifiškumą nei MF studentai ($p = 0,000$): Histgradient specifiškumas buvo 0,485 (nuo 0,422 iki 0,552), palyginus su MFS, kurių specifiškumas buvo 0,500 (nuo –0,250 iki 1,000).

Esant Gauso baltojo triukšmo SITS-40 lygmeniui, reikšmingų skirtumų tarp abiejų lyginamųjų grupių pagal SAK auskultacinių garsų klasę nenustatyta ($p > 0,05$), tačiau reikšmingi skirtumai nustatyti atpažįstant normalius plaučių garsus ir drėgnus auskultacinius karkalus ($p = 0,000$, $p = 0,024$). Histgradient MM modelio NAG, SAK ir DAK auskultacinių garsų klasių atpažinimo specifiškumas atitinkamai buvo 0,341 (nuo 0,288 iki 0,422), 0,256 (nuo 0,180 iki 0,374) ir 0,557 (nuo 0,491 iki 0,621), o MF studentų specifiškumas atpažįstant tas pačias auskultacinių garsų klases buvo 0,500 (nuo –0,250 iki 1,000), 0,500 (nuo 0,000 iki 1,000) ir 0,000 (nuo –0,250 iki 1,000). Histgradient modelis pasižymėjo statiškai reikšmingai geresniu specifiškumu atpažįstant drėgnus karkalus (DAK), bet statistiškai prastesniu rezultatu nepatologiniams plaučių garsams (NAG) esant Gauso baltojo triukšmo SITS-40 lygmeniui.

Priešingai, esant Gauso baltojo triukšmo aukščiausiajam, SITS-20, lygmeniui, medicinos fakulteto studentai parodė statistiškai reikšmingai geresnius garsų atpažinimo specifiškumo rezultatus nei Histgradient MM modelis visoms garsų klasėms ($p = 0,000$ visoms klasėms). Histgradient modelio specifiškumas buvo 0,116 (–0,013–0,173), 0,001 (–0,095–0,255) ir 0,000 (–0,045–0,067) NAG, SAK ir DAK auskultacinių garsų klasėms, lyginant su MF studentų minėtų auskultacinių klasių garsų atpažinimo specifiškumu 0,500 (–0,250–1,000), 0,500 (0,000–1,000) ir 0,500 (–0,250–1,000).



5.4.3 pav. *Boxplot diagramoje pavaizduotas Histgradient modelio ir medicinos studentų jautrumo palyginimas trims garsų klasėms (NAG, SAK, DAK) esant trims GBT lygiams*

GBT – Gauso baltasis triukšmas, Histgradient – (angl. *Histogram-based Gradient Boosting Classification Tree*) Histogramomis pagrįsto gradientinio stiprinimo klasifikacinio medžio mašininio mokymosi modelis, SITS – signalo ir triukšmo santykis, NAG – normalus auskultuotas garsas, DAK – drėgni auskultaciniai karkalai, SAK – sausi auskultaciniai karkalai, MFS – medicinos fakulteto studentai.

Boxplot diagramoje (5.4.3 pav.) pavaizduotas Histgradient mašininio mokymosi (MM) ir medicinos fakulteto (MF) studentų jautrumo pasiskirstymas atpažįstant tris auskultacinių plaučių garsų klases (NAG, SAK, DAK), esant skirtingiems Gauso baltojo triukšmo (GBT) lygiams.

Kai į modelį nepridėta Gauso baltojo triukšmo, Histgradient MM modelio ir medicinos fakulteto studentų jautrumo, atpažįstant auskultacinius garsus, pasiskirstymai reikšmingai nesiskyrė tik vienoje auskultacinių plaučių garsų klasėje, tai yra atpažįstant drėgnus karkalus (DAK) ($p > 0,05$). Tačiau statistiškai reikšmingas skirtumas yra tarp studijos grupių jautrumo, atpažįstant normalius auskultacinius garsus (NAG) ir sausus karkalus (SAK) ($p = 0,030$ ir $p = 0,000$). MM Histgradient modelio jautrumas buvo 0,471 (415–0,543) normalių auskultacinių garsų (NAG), 0,587 (0,522–0,654) sausų auskultacinių karkalų (SAK) ir 0,485 (0,422–0,552) drėgnų auskultacinių karkalų (DAK) garsų klasėse, lyginant su MF studentų jautrumu 0,500 (–0,250–1,000), 0,500 (0,000–1,000) ir 0,500 (–0,250–1,000) tose pačiose garsų atpažinimo klasėse.

Esant Gausinio užterštumo vidutiniam lygmeniui, SITS-40, reikšmingų skirtumų tarp abiejų tiriamųjų grupių jautrumo identifikuojant DAK auskultacinę plaučių garsų klasę nenustatyta ($p > 0,05$), tačiau reikšmingi skirtumai nustatyti atpažįstant normalius auskultacinius garsus ir sausus auskultacinius karkalus ($p = 0,000$ abiem klasėms). Histgradient MM modelio jautrumo pasiskirstymas identifikuojant NAG, SAK ir DAK auskultacinius plaučių garsus atitinkamai buvo 0,341 (0,288–0,422), 0,256 (0,180–0,374) ir 0,557 (0,491–0,621), o MF studentų jautrumas atpažįstant plaučių garsus tose pačiose auskultacinių garsų klasėse buvo 0,500 (–0,250–1,000), 0,500 (0,000–1,000) ir 0,000 (–0,250–1,000).

MM Histgradient modelio jautrumas, esant SITS-40 Gausinio užterštumo lygmeniui, pranoko žmogiškųjų tiriamųjų subjektų (t. y. studentų) jautrumą atpažįstant nepatologinius plaučių garsus (NAG). MM modelio ir studentų jautrumas tolygus identifikuojant drėgnus auskultacinius karkalus (DAK) ir statistiškai prastesnis jautrumas, atpažįstant sausus auskultacinius karkalus (SAK.)

Esant Gauso užterštumo aukščiausiam lygmeniui, SITS-20, MF studentai parodė statistiškai reikšmingai geresnius jautrumo rezultatus nei Histgradient MM modelis drėgnų auskultacinių karkalų (DAK) atpažinimo atžvilgiu ($p = 0,000$). Tačiau statistiškai reikšmingo skirtumo tarp abiejų lyginamųjų grupių nebuvo, žvelgiant į NAG ir SAK plaučių klasių identifikavimą ($p > 0,05$). MM Histgradient modelio jautrumas buvo 0,116 (–0,013–0,173), 0,001 (–0,095–0,255) ir 0,000 (–0,045–0,067) NAG, SAK ir DAK auskultacinių garsų klasėms, tuo tarpu MF studentų jautrumas 0,500 (–

0,250–1,000), 0,500 (0,000–1,000) ir 0,500 (–0,250–1,000) toms pačioms auskultacinių garsų klasėms identifikuoti.

IŠVADOS

1. Mašininio mokymosi modeliai bei medicinos studentai gali būti išmokyti identifikuoti tris auskultacinių plaučių garsų klases, per trumpą laiką, tačiau su skirtingais tikslumo lygiais.
2. Spektrogramų pagrindu sukurti mašininio mokymo modeliai parodė žymiai geresnį tikslumą lyginant su skalograminiais mašininio mokymosi modeliais.
3. Tiek medicinos studentus, tiek MM modelių gebėjimus atpažinti plaučių garsus paveikė padidėjęs GBT lygis, o mašininio mokymosi modeliai dažniau buvo paveikiami didžiausio garso užterštumo (SITS-20) lygmens, kuriame gebėjimas atpažinti normalius plaučių garsus, drėgnus ir sausus karkalus reikšmingai sumažėjo. Tuo tarpu medicinos studentų drėgnų karkalų atpažinimo tikslumą reikšmingai paveikė vidutinio lygio GBT užterštumas (SITS-40).
4. Mašininio mokymosi Histgradient modelis, esant SITS-40 Gausinio užterštumo lygmeniui, pranoko žmogiškųjų tiriamųjų subjektų (t. y. medicinos fakulteto studentų) rezultatus atpažįstant drėgnus auskultacinius karkalus didesniu Motiejaus koreliacijos koeficientu, todėl šis MM modelis gali būti pritaikomas kaip diagnostinis pagalbininkas esant vidutinio garso užterštumo lygmeniui.

PRAKTINĖS REKOMENDACIJOS

1. Visi gydytojai, slaugytojai, rezidentai ir medicinos studentai turėtų žinoti savo diagnostinį tikslumą pagal tam tikrą plaučių garsų klasę ir triukšmo lygį jų darbo aplinkoje.
2. Visi būsimi mašininio mokymosi modeliai turėtų būti kuriami ir vertinami aplinkos triukšmo sąlygoms.
3. Histgradient stiprinantysis mašininio mokymosi modelis turėtų būti toliau tyrinėjamas ir plėtojamas, siekiant padėti identifikuoti drėgnų karkalų grupei priskiriamus plaučių garsus, esant garso užterštumo sąlygoms.

4. Visi modeliai, naudojami kaip pagalbinė priemonė plaučių garsų diagnostikoje klinikiniame sveikatos priežiūros specialistų darbe, turėtų būti vertinami pagal skirtingus aplinkos triukšmo lygius.
5. Integruojant mašininio mokymosi modelį, kaip diagnostikos pagalbininką sveikatos priežiūros specialistui, atliekančiam auskultaciją, visuomet reikėtų įvertinti ir atsižvelgti į specialisto auskultacijos testo rezultatus ir jautrumą tam tikram konkrečiam garsui, esant tam tikroms aplinkos triukšmo sąlygoms. Taipogi reikėtų įvertinti konkretaus naudojamo mašininio mokymosi modelio darbo rezultatus, jo diagnostinį jautrumą tomis pačiomis sąlygomis, tuomet pritaikyti mašininį modelį turintį geresnį jautrumą kaip pagalbinę priemonę konkretaus sveikatos specialisto auskultacijų interpretavimo asistavime.

REFERENCES

1. René Théophile Hyacinthe Laënnec (1781–1826). Two hundred years of the stethoscope. A brief overview. Arch Argent Pediat [Internet]. 2020 Oct 1 [cited 2025 Jan 25];118(5). Available from: https://www.sap.org.ar/docs/publicaciones/archivosarg/2020/v118n5a12e.pdf
2. Drie MV, Harris A. The stethoscope goes digital: learning through attention, distraction and distortion. 2020 Nov 6 [cited 2025 Jan 27]. Available from: https://brill.com/view/journals/ges/77/1/article-p123_5.xml
3. Callahan D, Waugh J, Mathew GA, Granger WM. Stethoscopes: what are we hearing? Biomedical Instrumentation & Technology. 2007 Jul;41(4):318–23.
4. Thompson WR. In defence of auscultation: a glorious future? Heart Asia. 2017 Feb 1;9(1):44.
5. Goldsworthy S, Gomes P, Coimbra M, Patterson JD, Langille J, Perez G, et al. Do basic auscultation skills need to be resuscitated? A new strategy for improving competency among nursing students. Nurse Education Today. 2021 Feb 1;97:104722.
6. Abera Tessema B, Nemomssa HD, Lamesgin Simegn G. Acquisition and classification of lung sounds for improving the efficacy of auscultation diagnosis of pulmonary diseases. Medical Devices: Evidence and Research. 2022 Apr 7;15:89–102.
7. Viegi G, Maio S, Fasola S, Baldacci S. Global burden of chronic respiratory diseases. Journal of Aerosol Medicine and Pulmonary Drug Delivery. 2020 Aug;33(4):171–7.
8. Swarup S, Makaryus AN. Digital stethoscope: technology update. Medical Devices: Evidence and Research. 2018 Jan 4;11:29–36.
9. Leng S, Tan RS, Chai KTC, Wang C, Ghista D, Zhong L. The electronic stethoscope. BioMed Eng OnLine. 2015 Jul 10;14(1):66.
10. Gwo-Ching Chang, Yi-Ping Cheng. Investigation of noise effect on lung sound recognition. In: 2008 International Conference on Machine Learning and Cybernetics [Internet]. Kunming, China: IEEE; 2008 [cited 2024 Oct 9]. p. 1298–301. Available from: http://ieeexplore.ieee.org/document/4620605/
11. Kim KS, Kwon J, Ryu H, Kim C, Kim H, Lee EK, et al. The future of two-dimensional semiconductors beyond Moore's law. Nat Nanotechnol. 2024 Jul;19(7):895–906.
12. Diniz WJS, Canduri F. REVIEW-ARTICLE Bioinformatics: an overview and its applications. Genet Mol Res [Internet]. 2017 [cited 2025 Jan 25];16(1). Available from: http://www.funpecrp.com.br/gmr/year2017/vol16-1/pdf/gmr-16-01-gmr.16019645.pdf
13. Caballé-Cervigón N, Castillo-Sequera JL, Gómez-Pulido JA, Gómez-Pulido JM, Polo-Luque ML. Machine learning applied to diagnosis of human diseases: a systematic review. Applied Sciences. 2020 Jan;10(15):5135.
14. Jung SY, Liao CH, Wu YS, Yuan SM, Sun CT. Efficiently classifying lung sounds through depthwise separable CNN models with Fused STFT and MFCC features. Diagnostics. 2021 Apr;11(4):732.
15. Mayat U, Qureshi F, Ahmed S, Athavale Y, Krishnan S. Towards a low-cost point-of-care screening platform for electronic auscultation of vital body sounds. In: 2017 IEEE Canada International Humanitarian Technology Conference (IHTC) [Internet]. 2017 [cited 2025 Jan 24]. p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/8058166

16. Bachtiger P, Petri CF, Scott FE, Park SR, Kelshiker MA, Sahemey HK, et al. Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination in London, UK: a prospective, observational, multicentre study. The Lancet Digital Health. 2022 Feb 1;4(2):e117–25.
17. Yilmaz G, Rapin M, Pessoa D, Rocha BM, de Sousa AM, Rusconi R, et al. A Wearable stethoscope for long-term ambulatory respiratory health monitoring. Sensors. 2020 Jan;20(18):5124.
18. Bank I, Vliegen HW, Bruschke AVG. The 200th anniversary of the stethoscope: can this low-tech device survive in the high-tech 21st century? European Heart Journal. 2016 Dec 14;37(47):3536–43.
19. Nussbaumer M, Agarwal A. Stethoscope acoustics. Journal of Sound and Vibration. 2022 Oct 24;539:117194.
20. Brandt LJ. Use of the stethoscope to diagnose gastrointestinal and hepatic disorders. The American Journal of Gastroenterology 118(7):p 1113-1116, July 2023. Available from: https://doi.org/10.14309/ajg.0000000000002233.
21. Venkatram P. Auscultation: normal and abnormal heart sounds: aid in diagnosis and relationship to echocardiogram. In: Venkatram P, editor. Heart diseases and echocardiogram: principles in practice [Internet]. Cham: Springer Nature Switzerland; 2024 [cited 2025 Jan 26]. p. 55–62. Available from: https://doi.org/10.1007/978-3-031-59246-1_3
22. Kawamura R, Shimizu T. Auscultatory insights: unmasking atypical fibromuscular dysplasia in secondary hypertension. Journal of Hospital General Medicine. 2024;6(4): 103–5.
23. Brown SA, Carius BM, Monti JD, Robeck RS, Fritz DK. Combat medic-performed auscultation versus thoracic ultrasound image interpretation for pneumothorax detection: look or listen? Cureus. 16(9):e68657.
24. Reyes FM, Modi P, Le JK. Lung exam. In: StatPearls [Internet]. StatPearls Publishing; 2024 [cited 2025 Jan 26]. Available from: https://www.ncbi.nlm.nih.gov/sites/books/NBK459253/
25. BrJCardiol. Cardiac auscultation: an essential clinical skill in decline – The British Journal of Cardiology [Internet]. [cited 2025 Jan 24]. Available from: https://bjcardio.co.uk/2010/02/cardiac-auscultation-an-essential-clinical-skill-in-decline/
26. de Lima Andrade E, da Cunha e Silva DC, de Lima EA, de Oliveira RA, Zannin PHT, Martins ACG. Environmental noise in hospitals: a systematic review. Environ Sci Pollut Res. 2021 Apr 1;28(16):19629–42.
27. Nussbaumer M, Agarwal A. Stethoscope acoustics. Journal of Sound and Vibration. 2022 Oct 24;539:117194.
28. Arts L, Lim EHT, van de Ven PM, Heunks L, Tuinman PR. The diagnostic accuracy of lung auscultation in adult patients with acute pulmonary pathologies: a meta-analysis. Sci Rep. 2020 Apr 30;10(1):7347.
29. Pasterkamp H, Melbye H. Machines are learning chest auscultation. Will they also become our teachers? CHEST Pulmonary [Internet]. 2024 Dec 1 [cited 2025 Jan 24]; 2(4). Available from: https://www.chestpulmonary.org/article/S2949-7892(24)00045-X/fulltext
30. Emmanouilidou D, McCollum ED, Park DE, Elhilali M. Adaptive noise suppression of pediatric lung auscultations with real applications to noisy clinical settings in developing countries. IEEE Transactions on Biomedical Engineering. 2015 Sep;62(9): 2279–88.

31. Zun LS, Downey L. The effect of noise in the Emergency department. Academic Emergency Medicine. 2005;12(7):663–6.
32. Wallis R, Harris E, Lee H, Davies W, Astin F. Environmental noise levels in Hospital settings: a rapid review of measurement techniques and implementation in Hospital settings. Noise and Health. 2019 Oct;21(102):200.
33. Chang GC, Lai YF. Performance evaluation and enhancement of lung sound recognition system in two real noisy environments. Computer Methods and Programs in Biomedicine. 2010 Feb 1;97(2):141–50.
34. Zhang J, Wang HS, Zhou HY, Dong B, Zhang L, Zhang F, et al. Real-world verification of artificial intelligence algorithm-assisted auscultation of breath sounds in children. Front Pediatr [Internet]. 2021 Mar 23 [cited 2025 Jan 27];9. Available from: https://www.frontiersin.org/journals/pediatrics/articles/10.3389/fped.2021.627337/full
35. Narula J, Chandrashekhar Y, Braunwald E. Time to add a fifth pillar to bedside physical examination: inspection, palpation, percussion, auscultation, and insonation. JAMA Cardiology. 2018 Apr 1;3(4):346–50.
36. Jácome C, Aviles-Solis JC, Uhre ÅM, Pasterkamp H, Melbye H. Adventitious and normal lung sounds in the general population: comparison of standardized and spontaneous breathing. Respiratory Care. 2018 Nov 1;63(11):1379–87.
37. Torino C, Gargani L, Sicari R, Letachowicz K, Ekart R, Fliser D, et al. The agreement between auscultation and lung ultrasound in hemodialysis patients: the LUST study. Clin J Am Soc Nephrol. 2016 Nov 7;11(11):2005–11.
38. Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: a systematic review. PLOS ONE. 2017 May 26;12(5):e0177926.
39. Huang CH, Chen CH, Tzeng JT, Chang AY, Fan CY, Sung CW, et al. The unreliability of crackles: insights from a breath sound study using physicians and artificial intelligence. npj Prim Care Respir Med. 2024 Oct 15;34(1):1–7.
40. Haider NS, Behera AK. Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD). Biocybernetics and Biomedical Engineering. 2022 Jan 1;42(1):42–59.
41. Ponte DF, Moraes R, Hizume DC, Alencar AM. Characterization of crackles from patients with fibrosis, heart failure and pneumonia. Medical Engineering & Physics. 2013 Apr 1;35(4):448–56.
42. Lightfoot JT, Tuller B, Williams DF. Ambient noise interferes with auscultatory blood pressure measurement during exercise. Medicine & Science in Sports & Exercise. 1996 Apr;28(4):502.
43. Khan S, Smedt VD, Karsmakers P. Comparison of acoustic performance of different sensors for the purposes of on-body auscultation in noisy environments. IEEE Sensors Journal. 2023 Jul;23(13):14203–14.
44. Emmanouilidou D, McCollum ED, Park DE, Elhilali M. Computerized lung sound screening for pediatric auscultation in noisy field environments. IEEE Trans Biomed Eng. 2018 Jul;65(7):1564–74.
45. Heart and lung sound measurement using an esophageal stethoscope with adaptive noise cancellation – PMC [Internet]. [cited 2025 Jan 23]. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8540990/
46. Ye P, Li Q, Jian W, Liu S, Tan L, Chen W, et al. Regularity and mechanism of fake crackle noise in an electronic stethoscope. Front Physiol [Internet]. 2022 Dec 12 [cited 2025 Jan 23];13. Available from: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2022.1079468/full

47. Osterwalder J, Polyzogopoulou E, Hoffmann B. Point-of-care ultrasound – history, current and evolving clinical concepts in Emergency Medicine. Medicina (Kaunas). 2023 Dec 15;59(12):2179.
48. Sekiguchi H. Tools of the trade: point-of-care ultrasonography as a stethoscope. Seminars in Respiratory and Critical Care Medicine. 2016 Feb 4;37:68–87.
49. Lehrer S. Understanding lung sounds: third edition. Steven Lehrer; 2018. 157 p.
50. Afdi TLAS, Indriani SI. The art of diagnosis from breath sound: a literature review. 1. 2024;8(3):4136–44.
51. Naves R, Barbosa BHG, Ferreira DD. Classification of lung sounds using higher-order statistics: a divide-and-conquer approach. Computer Methods and Programs in Biomedicine. 2016 Jun 1;129:12–20.
52. Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. Annals of Thoracic Medicine. 2015 Sep;10(3):158.
53. Francis NA, Melbye H, Kelly MJ, Cals JWL, Hopstaken RM, Coenen S, et al. Variation in family physicians' recording of auscultation abnormalities in patients with acute cough is not explained by case mix. A study from 12 European networks. European Journal of General Practice. 2013 Jun;19(2):77–84.
54. Melbye H, Solis JCA, Jácome C, Pasterkamp H. Inspiratory crackles – early and late – revisited: identifying COPD by crackle characteristics. BMJ Open Resp Res [Internet]. 2021 Mar 5 [cited 2025 Jan 23];8(1). Available from: https://bmjopenrespres.bmj.com/content/8/1/e000852
55. Singh S. Respiratory symptoms and signs. Medicine. 2020 Apr 1;48(4):225–33.
56. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. Behavior Therapy. 2020 Sep 1;51(5):675–87.
57. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022 Jan;23(1):40–55.
58. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Markets. 2021 Sep 1;31(3):685–95.
59. Long Z, Zhuang L, Killick G, McCreadie R, Aragon-Camarasa G, Henderson P. Understanding and mitigating human-labelling errors in supervised contrastive learning. In: Leonardis A, Ricci E, Roth S, Russakovsky O, Sattler T, Varol G, editors. Computer vision – ECCV 2024. Cham: Springer Nature Switzerland; 2025. p. 435–54.
60. Kotsiantis SB, Kanellopoulos D, Pintelas PE. Data preprocessing for supervised leaning. 2006;1(1).
61. Haider NS, Singh BK, Periyasamy R, Behera AK. Respiratory sound based classification of chronic obstructive pulmonary disease: a risk stratification approach in machine learning paradigm. J Med Syst. 2019 Jun 28;43(8):255.
62. Palaniappan R, Sundaraj K. Respiratory sound classification using cepstral features and Support Vector Machines. In: 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS) [Internet]. 2013 [cited 2025 Jan 23]. p. 132–6. Available from: https://ieeexplore.ieee.org/abstract/document/6745460
63. Cinyol F, Baysal U, Köksal D, Babaoğlu E, Ulaşlı SS. Incorporating Support Vector Machines to the classification of respiratory sounds by Convolutional Neural Network. Biomedical Signal Processing and Control. 2023 Jan 1;79:104093.
64. Chen CH, Huang WT, Tan TH, Chang CC, Chang YJ. Using K-nearest neighbor classification to diagnose abnormal lung sounds. Sensors. 2015 Jun;15(6):13132–58.
65. Jaber MM, Abd SK, Shakeel PM, Burhanuddin MA, Mohammed MA, Yussof S. A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. Measurement. 2020 Oct 1;162:107883.

66. Ahuja R, Solanki V, Khullar V, Kumar L. Classification of non-speech sound signals: an approach of machine learning with MFCC feature extraction. In: 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT) [Internet]. 2024 [cited 2025 Mar 1]. p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/10738971
67. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev. 2021 Mar 1;54(3):1937–67.
68. Zhao X, Shao Y, Mai J, Yin A, Xu S. Respiratory sound classification based on BiGRU-attention network with XGBoost. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) [Internet]. 2020 [cited 2025 Jan 23]. p. 915–20. Available from: https://ieeexplore.ieee.org/abstract/document/9313506
69. Shokouhmand S, Rahman MM, Faezipour M, Bhatt S. Adventitious pulmonary sound detection: leveraging SHAP explanations and Gradient Boosting Insights. In: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. 2024 [cited 2025 Mar 1]. p. 1–4. Available from: https://ieeexplore.ieee.org/abstract/document/10782292
70. Sahin EK. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and Random Forest. SN Appl Sci. 2020 Jun 30;2(7):1308.
71. Lung Sound Recognition Based on Pre-Trained Convolutional Neural Network. AJETS [Internet]. 2022 [cited 2025 Mar 1];5(11). Available from: https://francis-press.com/papers/8179
72. Fraiwan L, Hassanin O, Fraiwan M, Khassawneh B, Ibnian AM, Alkhodari M. Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. Biocybernetics and Biomedical Engineering. 2021 Jan 1;41(1):1–14.
73. Jaber MM, Abd SK, Shakeel PM, Burhanuddin MA, Mohammed MA, Yussof S. A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. Measurement. 2020 Oct 1;162:107883.
74. Ira NT, Rahman MO. An efficient speech emotion recognition using ensemble method of supervised classifiers. In: 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE) [Internet]. 2020 [cited 2025 Mar 1]. p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/9350913
75. Mukherjee H, Sreerama P, Dhar A, Obaidullah SkMd, Roy K, Mahmud M, et al. Automatic lung health screening using respiratory sounds. J Med Syst. 2021 Jan 11;45(2):19.
76. Emmanouilidou D, Elhilal M. Characterization of noise contaminations in lung sound recordings. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. Osaka: IEEE; 2013 [cited 2025 Jan 24]. p. 2551–4. Available from: http://ieeexplore.ieee.org/document/6610060/
77. Chang GC, Cheng YP. Investigation of noise effect on lung sound recognition. In: 2008 International Conference on Machine Learning and Cybernetics [Internet]. 2008 [cited 2025 Jan 23]. p. 1298–301. Available from: https://ieeexplore.ieee.org/abstract/document/4620605
78. Emmanouilidou D, Elhilal M. Characterization of noise contaminations in lung sound recordings. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. 2013 [cited 2025 Jan 24]. p. 2551–4. Available from: https://ieeexplore.ieee.org/abstract/document/6610060

79. Razvadauskas H, Vaičiukynas E, Buškus K, Arlauskas L, Nowaczyk S, Sadauskas S, et al. Exploring classical machine learning for identification of pathological lung auscultations. Computers in Biology and Medicine. 2024 Jan 1;168:107784.
80. Nasifoglu H, Erogul O. Obstructive sleep apnea prediction from electrocardiogram scalograms and spectrograms using convolutional neural networks. Physiol Meas. 2021 Jun;42(6):065010.
81. Scarpiniti M, Parisi R, Lee YC. A scalogram-based CNN approach for audio classification in construction sites. Applied Sciences. 2024 Jan;14(1):90.
82. Fraiwan M, Fraiwan L, Alkhodari M, Hassanin O. Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. J Ambient Intell Human Comput. 2022 Oct 1;13(10):4759–71.
83. Kim Y, Hyon Y, Lee S, Woo SD, Ha T, Chung C. The coming era of a new auscultation system for analyzing respiratory sounds. BMC Pulm Med. 2022 Mar 31;22(1):119.
84. Aykanat M, Kılıç Ö, Kurt B, Saryal S. Classification of lung sounds using convolutional neural networks. J Image Video Proc. 2017 Sep 11;2017(1):65.
85. Nelson G, Rajamani R, Erdman A. Noise control challenges for auscultation on medical evacuation helicopters. Applied Acoustics. 2014 Jun 1;80:68–78.
86. Ye P, Li Q, Jian W, Liu S, Tan L, Chen W, et al. Regularity and mechanism of fake crackle noise in an electronic stethoscope. Front Physiol [Internet]. 2022 Dec 12 [cited 2025 Jan 24];13. Available from: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2022.1079468/full
87. Design and comparative performance of a robust lung auscultation system for noisy clinical settings | IEEE Journals & Magazine | IEEE Xplore [Internet]. [cited 2025 Jan 24]. Available from: https://ieeexplore.ieee.org/abstract/document/9345968
88. Nowak LJ, Nowak KM. Sound differences between electronic and acoustic stethoscopes. BioMed Eng OnLine. 2018 Aug 3;17(1):104.
89. Rabinowitz P, Taiwo O, Sircar K, Aliyu O, Slade M. Physician hearing loss. American Journal of Otolaryngology. 2006 Jan 1;27(1):18–23.
90. Hafke-Dys H, Bręborowicz A, Kleka P, Kociński J, Biniakowski A. The accuracy of lung auscultation in the practice of physicians and medical students. PLOS ONE. 2019 Aug 12;14(8):e0220606.
91. Seah JJ, Zhao J, Wang DY, Lee HP. Review on the advancements of stethoscope types in chest auscultation. Diagnostics. 2023 Jan;13(9):1545.
92. Gaydos S. Clinical auscultation in noisy environments. Journal of Emergency Medicine. 2012 Sep 1;43(3):492–3.
93. Leuppi JD, Dieterle T, Koch G, Martina B, Tamm M, Perruchoud AP, et al. Diagnostic value of lung auscultation in an emergency room setting. Swiss Med Wkly. 2005 Sep 3;135(35–36):520–4.
94. Ye P, Li Q, Jian W, Liu S, Tan L, Chen W, et al. Regularity and mechanism of fake crackle noise in an electronic stethoscope. Front Physiol [Internet]. 2022 Dec 12 [cited 2025 Jan 24];13. Available from: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2022.1079468/full
95. Aliberti S, Cruz CSD, Amati F, Sotgiu G, Restrepo MI. Community-acquired pneumonia. The Lancet. 2021 Sep 4;398(10303):906–19.
96. Olson G, Davis AM. Diagnosis and treatment of adults with community-acquired pneumonia. JAMA. 2020 Mar 3;323(9):885–6.
97. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. European Heart Journal. 2021 Sep 21;42(36):3599–726.

98. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. European Heart Journal. 2016 Jul 14;37(27):2129–2200m.
99. Louis R, Satia I, Ojanguren I, Schleich F, Bonini M, Tonia T, et al. European Respiratory Society guidelines for the diagnosis of asthma in adults. European Respiratory Journal [Internet]. 2022 Sep 7 [cited 2025 Jan 26];60(3). Available from: https://publications.ersnet.org/content/erj/60/3/2101585
100. MacLeod M, Papi A, Contoli M, Beghé B, Celli BR, Wedzicha JA, et al. Chronic obstructive pulmonary disease exacerbation fundamentals: diagnosis, treatment, prevention and disease impact. Respirology. 2021;26(6):532–51.
101. Canepa M, Franssen FME, Olschewski H, Lainscak M, B öhm M, Tavazzi L, et al. Diagnostic and therapeutic gaps in patients with heart failure and chronic obstructive pulmonary disease. JACC: Heart Failure. 2019 Oct;7(10):823–33.
102. Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management: a review. JAMA. 2019 Oct 1;322(13):1294–304.
103. Corsonello A, Fabbietti P, Formiga F, Moreno-Gonzalez R, Tap L, Mattace-Raso F, et al. Chronic kidney disease in the context of multimorbidity patterns: the role of physical performance. BMC Geriatr. 2020 Oct 2;20(1):350.
104. Pippard B, Bhatnagar M, McNeill L, Donnelly M, Frew K, Aujayeb A. Hepatic Hydrothorax: a narrative review. Pulm Ther. 2022 Sep 1;8(3):241–54.
105. About inpatient services provided at LSMU Kaunas Hospital [Internet]. LSMU Kauno ligoninė. [cited 2025 Jan 26]. Available from: https://kaunoligonine.lt/en/stacionarines-asmens-sveikatos-prieziuros-paslaugos/
106. KoreaMed Synapse [Internet]. [cited 2025 Jan 25]. Available from: https://synapse.koreamed.org/articles/1149215
107. Das S, Mitra K, Mandal M. Sample size calculation: basic principles. Indian Journal of Anaesthesia. 2016 Sep;60(9):652.
108. Ari S, Sensharma K, Saha G. DSP implementation of a heart valve disorder detection system from a phonocardiogram signal. Journal of Medical Engineering & Technology. 2008 Jan 1;32(2):122–32.
109. Higashiyama S, Tamakoshi K, Yamauchi T. Effectiveness of a new interactive web teaching material for improving lung auscultation skills: randomized controlled trial for clinical nurses. Nagoya J Med Sci. 2022 Aug;84(3):526–38.
110. Strauss-Blasche G, Moser M, Voica M, McLeod D, Klammer N, Marktl W. Relative timing of inspiration and expiration affects respiratory sinus arrhythmia. Clinical and Experimental Pharmacology and Physiology. 2000;27(8):601–6.
111. Sharma G, Umapathy K, Krishnan S. Trends in audio signal feature extraction methods. Applied Acoustics. 2020 Jan 15;158:107020.
112. Rácz A, Bajusz D, Héberger K. Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification. Molecules. 2021 Jan;26(4):1111.
113. Sadaiyandi J, Arumugam P, Sangaiah AK, Zhang C. Stratified sampling-based deep learning approach to increase prediction accuracy of unbalanced dataset. Electronics. 2023 Jan;12(21):4423.
114. Erickson BJ, Kitamura F. Magician's corner: 9. Performance metrics for machine learning models. Radiology: Artificial Intelligence. 2021 May;3(3):e200126.
115. Loey M, Mirjalili S. COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models. Computers in Biology and Medicine. 2021 Dec 1;139:105020.

116. Al-Tawfiq JA, Memish ZA. Diagnosis of SARS-CoV-2 infection based on CT scan vs RT-PCR: reflecting on experience from MERS-CoV. Journal of Hospital Infection. 2020 Jun;105(2):154–5.
117. Yu X, Lou B, Shi B, Winkel D, Arrahmane N, Diallo M, et al. False positive reduction using multiscale contextual features for prostate cancer detection in multi-parametric MRI scans. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) [Internet]. 2020 [cited 2025 Jan 24]. p. 1355–9. Available from: https://ieeexplore.ieee.org/abstract/document/9098338
118. Ballabio D, Grisoni F, Todeschini R. Multivariate comparison of classification performance measures. Chemometrics and Intelligent Laboratory Systems. 2018 Mar 15; 174:33–44.
119. Halimu C, Kasem A, Newaz SHS. Empirical comparison of area under ROC curve (AUC) and Mathew Correlation Coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. In: Proceedings of the 3rd International Conference on Machine Learning and Soft Computing [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2025 Jan 24]. p. 1–6. (ICMLSC '19). Available from: https://doi.org/10.1145/3310986.3311023
120. Carrington AM, Fieguth PW, Qazi H, Holzinger A, Chen HH, Mayr F, et al. A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. BMC Med Inform Decis Mak. 2020 Jan 6;20(1):4.
121. Dinga R, Penninx BWJH, Veltman DJ, Schmaal L, Marquand AF. Beyond accuracy: Measures for assessing machine learning models, pitfalls and guidelines [Internet]. bioRxiv; 2019 [cited 2025 Jan 24]. p. 743138. Available from: https://www.biorxiv.org/content/10.1101/743138v1
122. Wong HB, Lim GH. Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. Proceedings of Singapore Healthcare. 2011 Dec 1;20(4):316–8.
123. Humphrey A, Kuberski W, Bialek J, Perrakis N, Cools W, Nuyttens N, et al. Machine learning classification of astronomical sources: estimating F1-score in the absence of ground truth. Monthly Notices of the Royal Astronomical Society: Letters. 2022 Nov 21;517(1):L116–20.
124. Warrens MJ. Five ways to look at Cohen's kappa. Journal of Psychology & Psychotherapy. 2015 Jul 28;5.
125. Meghanathan N. Assortativity analysis of real-world Network Graphs based on Centrality Metrics. CIS. 2016 Jul 3;9(3):7.
126. Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen J, et al. JASP: Graphical statistical software for common statistical designs. Journal of Statistical Software. 2019 Jan 29;88:1–17.
127. Ullah A, Khan MS, Khan MU, Mujahid F. Automatic classification of lung sounds using machine learning algorithms. In: 2021 International Conference on Frontiers of Information Technology (FIT) [Internet]. 2021 [cited 2025 Feb 17]. p. 131–6. Available from: https://ieeexplore.ieee.org/document/9701411
128. Ira NT, Rahman MO. An efficient speech emotion recognition using ensemble method of supervised classifiers. In: 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE) [Internet]. 2020 [cited 2025 Feb 17]. p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/9350913
129. Qin Y, Wu J, Xiao W, Wang K, Huang A, Liu B, et al. Machine learning models for data-driven prediction of diabetes by lifestyle type. International Journal of Environmental Research and Public Health. 2022 Jan;19(22):15027.

130. Zaman SR, Sadekeen D, Alfaz MA, Shahriyar R. One source to detect them all: gender, age, and emotion detection from voice. In: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC) [Internet]. 2021 [cited 2025 Mar 2]. p. 338–43. Available from: https://ieeexplore.ieee.org/abstract/document/9529731
131. Lee JA, Kwak KC. Heart sound classification using wavelet analysis approaches and ensemble of deep learning models. Applied Sciences. 2023 Jan;13(21):11942.
132. Soni PN, Shi S, Sriram PR, Ng AY, Rajpurkar P. Contrastive learning of heart and lung sounds for label-efficient diagnosis. Patterns (N Y). 2022 Jan 14;3(1):100400.
133. Addison PS. The Illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance, second edition. 2nd ed. Boca Raton: CRC Press; 2017. 464 p.
134. Wasano K, Kaga K, Ogawa K. Patterns of hearing changes in women and men from denarians to nonagenarians. The Lancet Regional Health – Western Pacific [Internet]. 2021 Apr 1 [cited 2025 Mar 2];9. Available from: https://www.thelancet.com/journals/lanwpc/article/PIIS2666-6065(21)00040-7/fulltext
135. Hafke-Dys H, Bręborowicz A, Kleka P, Kociński J, Biniakowski A. The accuracy of lung auscultation in the practice of physicians and medical students. PLOS ONE. 2019 Aug 12;14(8):e0220606.
136. Ye P, Li Q, Jian W, Liu S, Tan L, Chen W, et al. Regularity and mechanism of fake crackle noise in an electronic stethoscope. Front Physiol [Internet]. 2022 Dec 12 [cited 2025 Mar 2];13. Available from: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2022.1079468/full
137. Park JS, Kim K, Kim JH, Choi YJ, Kim K, Suh DI. A machine learning approach to the development and prospective evaluation of a pediatric lung sound classification model. Sci Rep. 2023 Jan 23;13(1):1289.
138. Moriki D, Koumpagioti D, Kalogiannis M, Sardeli O, Galani A, Priftis KN, et al. Physicians' ability to recognize adventitious lung sounds. Pediatric Pulmonology. 2023;58(3):866–70.

LIST OF ARTICLES IN WHICH THE RESULTS OF THE DISSERTATION RESEARCH HAVE BEEN PUBLISHED

1. Razvadauskas H, Razvadauskienė J, Aliulis M, Aliulytė R, Naudžiūnas A, Paukštaitienė R, Sadauskas S. Influence of Gaussian White Noise on Medical Students' Capacity to Accurately Identify Pulmonary Sounds. Noise & Health 26(123):p 474-482, October-December 2024. doi: 10.4103/nah.nah_98_24
2. Razvadauskas H, Vaičiukynas E, Buškus K, Arlauskas L, Nowaczyk S, Sadauskas S, Naudžiūnas A. Exploring classical machine learning for identification of pathological lung auscultations. Computers in Biology and Medicine. 2024 Jan 1;168:107784–4. doi: 10.1016/j.compbiomed. 2023.107784

LIST OF SCIENTIFIC CONFERENCES WHERE THE RESULTS OF THE DISSERTATION RESEARCH HAVE BEEN PUBLISHED

1. Razvadauskas H, Vaičiukynas E, Danėnas, P, Sadauskas, S, Naudžiūnas, A. Comparison of spectrograms and scalograms influence on machine learning model accuracy in identifying pathological lung sounds. International Health Sciences Conference for All (IHSC for All) "Precision Medicine": March 25-26, 2024, Kaunas: abstract book 2024 / p. 487-489
2. Razvadauskas H, Vaičiukynas E, Buškus K, Razvadauskienė J, Vaičiukynas E, Buškus K, Aliulis M, Aliulytė R, Sadauskas S, Naudžiūnas A, Paukštaitienė R. Comparison of machine learning model with human subjects to correctly identify adventitious lung sounds. Medicina: Abstracts of the International Scientific Conferences on Medicine & Public Health Research Week 2023 (RW2023) : March 29-31, 2023, Riga, Latvia, 2023-06-10, vol. 59, no. Suppl. 2, p. 9-9
3. Razvadauskas H, Sadauskas S, Naudžiūnas A, Razvadauskienė, J, Aliulis M, Aliulytė R, Paukštaitienė R. The Effect of white noise on medical students' ability to identify pathological lung sounds. International scientific conference "Actualities in the treatment and diagnosis of internal diseases and syndromes, scientific research news": April 26, 2023 : abstract book / Department of Internal Medicine, Medical Academy, Lithuanian University of Health Sciences Kaunas. International Laser Center CVTI, Bratislava, Slovak Republic. Immunology and Pathology University of Siena, Dpt. of Molecular and Developmental Medicine, Italy. Kaunas : UAB „Medicinos spaudos namai", 2023. ISBN 9786098113150., 2023-04-26, p. 13-14.

APPENDIX

Annex 1

PACIENTO IŠTYRIMO ANKETA

Data			
Suteiktasis tiriamojo kodas			
Tiriamąjį apibūdinantys dokumentiniai duomenys			
Lytis	☐ Vyras ☐ Moteris	Amžius (m.)	
Ūgis (cm)		Svoris (kg)	
Diagnozė (TLK-10-AM kodas)		Pagrindinė	
		Gretutinė	
		Komplikacijos	
Anamnezės ypatumai (turintys įtakos plaučių auskultaciniams garsams)			
Persirgtos ligos	☐ Plaučių _____ _____ ☐ Širdies _____ _____	Žalingi įpročiai	☐ Rūko ☐ Nerūko ☐ Neberūko _____ metų
Objektyvus ištyrimas			
Dusulys	☐ Taip ☐ Ne	Kvėpavimo dažnis (k./min.)	
Karkalai	☐ Sausi ☐ Drėgni ☐ Mišrūs	Periferinės edemos	☐ Yra ☐ Nėra
SpO_2 (proc.)		AKS (mmHg)	
		ŠSD (k./min.)	
Instrumentinių ir laboratorinių tyrimų rezultatai			
BKT	☐ Anemija ☐ Uždegiminis procesas (bakterinis) ☐ Kita _____ _____	CRB	

Plaučių rentgenologiniai pakitimai	☐ Norma ☐ Pneumonija _____ ☐ Bronchitas ☐ Peribronchiniai pakitimai ☐ Kita _____ _____	D-dimerai (jei atlikta)	
		Plaučių KT rezultatai (jei atlikta)	
Plaučių garsų įrašai (auskultuoti garsai ir pastabos)			
Pirmas taškas		Antras taškas	
Trečias taškas		Ketvirtas taškas	
Penktas taškas		Šeštas taškas	

Annex 2

KAUNO REGIONINIS BIOMEDICININIŲ TYRIMŲ ETIKOS KOMITETAS

Lietuvos sveikatos mokslų universitetas, A. Mickevičiaus g. 9, LT 44307 Kaunas, tel. (+370) 37 32 68 89;el.paštas: kaunorbtek@lsmuni.lt

LEIDIMAS ATLIKTI BIOMEDICININĮ TYRIMĄ

2021-05-11 Nr. BE-2-57

Biomedicininio tyrimo pavadinimas: „Plaučių ir širdies auskultacijų patologinių garsų vertinimas, taikant dirbtinio intelekto analizę“	
Protokolo Nr.:	**1.3**
Data:	**2021-04-25**
Versija:	**1.2**
Asmens informavimo forma:	**Versija 1.3, data: 2021-04-13**
Pagrindinis tyrėjas: Biomedicininio tyrimo vieta: Įstaigos pavadinimas: Adresas:	**Gyd. Haroldas Razvadauskas** Lietuvos sveikatos mokslų universiteto Kauno ligoninė, Vidaus ligų klinika Josvainių g. 2, LT-47141, Kaunas

Išvada:

Kauno regioninio biomedicininių tyrimų etikos komiteto posėdžio, įvykusio **2021 m. gegužės mėn. 4 d.** (protokolo Nr. 2021-BE10-0005) sprendimu pritarta biomedicininio tyrimo vykdymui.

Mokslinio eksperimento vykdytojai įsipareigoja: (1) nedelsiant informuoti Kauno Regioninį biomedicininių Tyrimų Etikos komitetą apie visus nenumatytus atvejus, susijusius su studijos vykdymu, (2) iki sausio 15 dienos – pateikti metinį studijos vykdymo apibendrinimą bei, (3) per mėnesį po studijos užbaigimo, pateikti galutinį pranešimą apie eksperimentą.

Kauno regioninio biomedicininių tyrimų etikos komiteto nariai			
Nr.	Vardas, Pavardė	Veiklos sritis	Dalyvavo posėdyje
1.	Doc. dr. Gintautas Gumbrevičius	Klinikinė farmakologija	Taip
2.	Prof. dr. Kęstutis Petrikonis	Neurologija	Taip
3.	Dr. Saulius Raugelė	Chirurgija	Ne
4.	Dr. Lina Jankauskaitė	Pediatrija	Taip
5.	Prof. dr. Džilda Veličkienė	Endokrinologija	Ne
6.	Doc. dr. Eimantas Peičius	Visuomenės sveikata	Taip
7.	Aušra Degutytė	Visuomenės sveikata	Taip
8.	Dr. Žydrūnė Luneckaitė	Visuomenės sveikata	Taip
9.	Viktorija Bučinskaitė	Teisė	Taip

Kauno regioninis biomedicininių tyrimų etikos komitetas dirba vadovaudamasis etikos principais nustatytais biomedicininių tyrimų Etikos įstatyme, Helsinkio deklaracijoje, vaistų tyrinėjimo Geros klinikinės praktikos taisyklėmis.

Kauno RBTEK pirmininkas Doc. dr. Gintautas Gumbrevičius

LIETUVOS SVEIKATOS MOKSLŲ UNIVERSITETAS * KAUNO REGIONINIS BIOMEDICININIŲ TYRIMŲ ETIKOS KOMITETAS

CURRICULUM VITAE

Name, Surname:	**Haroldas Razvadauskas**
Address:	LSMU Kaunas Hospital Internal Medicine Department Josvainių 2, LT-47144 Kaunas, Lithuania
E-mail:	h.razvadauskas@protonmail.com
Phone:	+370 612 52277
Web:	linkedin.com/in/haroldas-razvadauskas-md-26aba35a

Education:

06/2020–present	Doctor of Philosophy Medicine Lithuanian University of Health Science (LSMU), Kaunas, Lithuania
09/2014–06/2020	Internal Medicine Physician specialisation LSMU, Kaunas, Lithuania
09/2008–06/2014	Doctor of Medicine LSMU, Kaunas, Lithuania
09/2004–07/2007	Bachelor of Sciences in Genetics (Hons.) 2–1, Aberystwyth University of Wales, Aberystwyth, United Kingdom

Work experience:

12/2023–present	Biomedical Signal Processing and Control Journal Reviewer
09/2020–present	Assistant lecture at Internal Medicine Department, LSMU Kaunas Hospital, Kaunas, Lithuania
03/2023–11/2023	Principle Researcher in Artificial Intelligence Application to Pulmonary Sounds, Lithuanian University of Health Science (LSMU), Kaunas, Lithuania
06/2020–10/2022	Accident and Emergency Doctor, LSMU Kaunas Hospital, Kaunas, Lithuania
08/2014–06/2020	Resident Doctor, LSMU Kaunas Hospital, Kaunas, Lithuania
08/2011–06/2014	Principle Researcher Investigating in International Projected with 3M company, 3M Company, Kaunas, Lithuania
06/2011–08/2011	Research Assistant Internship, Abbaltis Ltd., Kent, United Kingdom
06/2012–08/2012	Research Assistant Internship, Abbaltis Ltd., Kent, United Kingdom

Awards:

03/2023	€20,000 won for a multidisciplinary team for a project: "Identification of pathological lung sounds using artificial intelligence techniques (DITA)", LSMU and KTU Innovation Funds
12/2020	Participated in a competition and won a grant for my project: "A prospective study on the effect of white noise on medical students' ability to identify three different classes of lung sounds", LSMU Education Funds

Skills:

Expertise in writing a study protocol.
Experience in managing international research projects.
Strong expertise in creating and collaborating within a multidisciplinary team to a set goal.
In-depth knowledge of diagnostics and therapeutics in healthcare settings.
Robust comprehension of bioethics.
Data cleaning, feature extraction and machine learning model application via Python on Jupyter Notebook.

Languages:

English: native
Lithuanian: native
Norwegian (Bokmål): beginner

ACKNOWLEDGMENTS

I want to express my sincere gratitude to my supervisor, Prof. Saulius Sadauskas, whose unwavering support has been instrumental in my academic journey. Thank you for encouraging me to pursue my goals in this challenging path.

I am grateful to the Clinic of Internal Medicine staff at LSMU: Prof. Albinas Naudžiūnas, who has provided perceptive scientific insights, helped my research, and always encouraged and pushed me during difficult times.

I would like to acknowledge all my colleagues, especially Lec. Andrius Ališauskas, for his generosity with his time and advice.

During the research, I worked with Prof. Evaldas Vaičiukynas and Prof. Renata Paukštaitienė, from whom I learned a great deal about machine learning and statistical analysis, respectively, which I could apply to my research project. It is also important to mention Prof. Virgilijus Ulozas provided constructive criticism to improve and refine the draft work.

I want to thank the heads of the Internal Medicine and Cardiology Departments at Kaunas Hospital of Lithuanian University of Health Sciences, Irena Šidiškienė and Laima Jankauskienė, for creating suitable conditions for recording lung sounds for the database. I would also like to express my gratitude to the nurses and physicians who created conducive circumstances for me to work in.

In addition, I would like to acknowledge the following persons and companies for their letters of support: James Schofield of TopMD Precision Medicine Ltd., Giedrė Brandao of Abbaltis Ltd., and Stéphane Favier of eKuore Chip Ideas Electronics S.L. Mobility Insights Manager Fransua Razvadauskas at Euromonitor for reviewing the article for readability.

A great part of the project depended on LSMU Science and Education staff providing a grant for this research. This created an opportunity for KTU and LSMU collaboration, from which a highly impactful article was published, and future articles are still being generated for publication.

I must thank my whole family, especially my parents (Aleksandras and Giedrė Razvadauskai), my parents-in-law, my brother Marijus Razvadauskas, my sister Ruta Astani, and my dear friend Mindaugas Kukis for their moral support.

Finally, this project would be impossible without my fantastic wife, Jurgita Razvadauskienė, who fully supported this endeavour, and provided time for me to write this thesis, and even offered to read through the thesis draft whilst looking after two of our children: Dorotėja and Motiejus.

I am very grateful to all of you. This work would not be possible without your help and contributions.