I.J. Information Technology and Computer Science, **2025, 3, 1-12**
Published Online on June 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2025.03.01

Spectrogram-based Deep Learning Approach for Anomaly Detection from Cough Sounds

Tugce Keles*
Digital Forensics Engineering, Firat University, Elazig, 23200, Turkiye
E-mail: tkeles@firat.edu.tr
ORCID iD: https://orcid.org/0000-0003-0131-2826
*Corresponding author

Sengul Dogan
Digital Forensics Engineering, Firat University, Elazig, 23200, Turkiye
E-mail: sdogan@firat.edu.tr
ORCID iD: https://orcid.org/0000-0001-9677-5684

Abdul-Hafeez Baig
University of Southern Queensland, School of Management and Enterprise, Toowoomba, QLD, Australia
E-mail: Abdul.Hafeez-Baig@usq.edu.au
ORCID iD: https://orcid.org/0000-0003-3848-8008

Turker Tuncer
Digital Forensics Engineering, Firat University, Elazig, 23200, Turkiye
E-mail: turkertuncer@firat.edu.tr
ORCID iD: https://orcid.org/0000-0002-5126-6445

Received: 04 March 2025; Revised: 18 April 2025; Accepted: 29 April 2025; Published: 08 June 2025

Abstract: Artificial intelligence is now applied in many fields beyond computer science. In healthcare, it enables early disease detection and improves patient outcomes. This study develops a model that uses AI to find abnormal patterns in cough sounds. A cough is a key symptom of asthma and other respiratory diseases. Previous research has focused on raw audio signals of coughs. In contrast, we analyze spectrogram images derived from these sounds to improve accuracy. We designed a new convolutional neural network (CNN) for this purpose and the recommended CNN is termed as TwoConvNeXt. To showcase the classification performance of the recommended TwoConvNeXt model, a cough sound dataset has been utilized and the recommended TwoConvNeXt achieved 99.66% classification test accuracy.

These results illustrate that the presented TwoConvNeXt CNN architecture can be useful in both research and clinical settings. This CNN model can be utilized for other image classification problems. It may aid in the early diagnosis of respiratory conditions. Future work will expand the dataset and test the model on larger, more diverse samples.

Index Terms: Asthma Detection, Deep Learning, Cough Sounds, Artificial Intelligence, Convolutional Neural Networks.

1. Introduction

Cough is a natural reflex that clears the airways. It is a key symptom in diagnosing respiratory conditions such as allergies, pneumonia, bronchitis, chronic obstructive pulmonary disease (COPD), and asthma [1]. Coughs are generally classified as dry or productive. This distinction offers key information about the underlying illness. Characteristics like how often a person coughs, how long each cough lasts, its strength, and when it occurs can all help clinicians during diagnosis [2].

Asthma is a chronic respiratory disease common around the world. It can greatly affect daily life. It causes cough, wheezing, shortness of breath, and chest tightness [3]. These symptoms result from inflammation in the airways. Early diagnosis of asthma is important to improve patients' quality of life and reduce death rates. However, asthma symptoms often resemble those of other respiratory diseases. This similarity can make a clear diagnosis difficult [4].

Nowadays, the use of artificial intelligence in healthcare continues to grow rapidly. This provides new opportunities in analyzing cough and similar symptoms [5]. With the developed artificial intelligence algorithms, the frequency, intensity and timing of cough sounds can be analyzed. In this way, it becomes possible to distinguish normal and abnormal coughs and detect diseases at an early stage [6]. Studies on sound signal analysis are quite common in the literature, this study focuses on the classification of spectrograms obtained from cough sounds. This study aims to detect abnormalities in cough sounds using an artificial intelligence-based model. As a result of the study, it is aimed to develop personal treatment strategies, contribute to medical studies with data analysis and achieve high accuracy rates.

Within the scope of the study, cough sounds were obtained from various datasets. These sound data will be trained with deep learning models and the performances of these models will be evaluated according to accuracy, precision and specificity metrics. In the study, an artificial intelligence-based early diagnosis system was developed to improve the treatment processes of chronic diseases such as asthma. This allows patients to better manage their health status and the process. In addition, analyses to be performed on the developed datasets in the future can provide important information about the causes, triggers and progression of asthma. This can also make important contributions to the literature.

2. Literature Review

Reviewing the existing literature reveals that research has predominantly focused on cough and similar sound signals, typically using pre-trained deep learning models for analysis. Many studies in the literature diagnose diseases based on the analysis of frequency, time, and intensity characteristics of sound signals. These studies are particularly supported by various neural network architectures and signal processing techniques. However, current approaches are primarily concentrated on raw sound data, with insufficient attention given to more advanced image-based analysis methods such as spectrograms.

Kilic et al. [7] aimed to detect asthma from cough sounds with artificial intelligence. They used a dataset consisting of 1428 participants in the study. The age range of these participants ranged from 5 to 18 and included 613 asthma patients and 815 healthy individuals. In the proposed automatic asthma detection model, global chaotic logistic pattern (GCLP) was used in the feature extraction phase. In signal separation, feature vectors were created at four different levels using Tunable Q-factor wavelet transform (TQWT). The proposed model has an accuracy rate of 99.44%, while the LOSO CV method reached an accuracy rate of 98.53%. Shruthi et al. [8] they aimed to use deep learning to increase the accuracy rate in the process of detecting lung diseases. They used normalization and data augmentation techniques in the pre-processing phase to increase the diversity of the data set. In the study, they used a dataset consisting of extensive lung images in the training and testing phases. Using deep learning models, features for these diseases were learned and their performance was evaluated. The results of the study show that deep learning models surpass traditional methods in the detection of lung diseases and achieve high accuracy rates. Wang et al. [9], they used cough sounds for COVID-19 detection. They used the AlCoVidvn 115 M dataset in the study. The dataset used has 4,068 cough sound recordings, 669 from COVID-19 positive patients and 3,399 from COVID-19 negative participants. To mitigate the data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was used to augment the COVID-19 patient samples. Features were extracted from the cough sounds using Mel spectrograms, Mel Frequency Cepstral Coefficients (MFCC), and VGGish embeddings. These networks take one-dimensional audio feature sequences as input to three separate Convolutional Neural Networks (CNNs) through different processes and combine the output data through fully connected layers for classification. The study compared the performance of the MHCNN by splitting it into three independent single-head CNNs named Head1 CNN, Head2 CNN, and Head3 CNN. The MHCNN model, using different features such as Mel spectrogram, MFCC, and VGG embeddings, outperformed the single-head CNNs, achieving an accuracy rate of 98.09%. Kuluozturk et al. [10], focused on disease detection using cough sound signals by collecting a large cough sound dataset consisting of categories such as COVID-19, heart failure, acute asthma, and healthy individuals. The created model includes several components: feature extraction utilizing the Directed Knight Pattern (DKP), four different pooling methods for signal decomposition, Iterative Neighbor Analysis (INCA) for selecting features, and a k-nearest neighbor (kNN) classifier for identification. The model generated 41 feature vectors and selected the best ten of these vectors. The model achieved an accuracy rate of 99.39% in multi-class classification based on cough sounds. Prajapati et al. [11], aimed to predict lung diseases using a multi-class classification method in their research. The central objective of the study is to accurately predict lung diseases using breath sounds. The study encompasses five different disease classes simultaneously: Healthy, Asthma, Chronic Obstructive Pulmonary Disease (COPD), Lung Fibrosis, and Pneumonia. An imbalanced dataset containing breath sounds obtained from DB-1 and DB-2 was balanced using data augmentation techniques. Data augmentation was performed using methods such as Pitch Shifting, Time Stretching, Dynamic Range Compression, and Background Noise. As a result, a total of 1964 files were obtained for 224 healthy, 198 asthma, 210 pneumonia and 216 COPD and lung fibrosis. In the obtained dataset, windowing was applied to the audio signals. Then, feature extraction was performed with FFT. In the study, they used five different machine learning algorithms: black trees, naive bayes, support vector machine, k-nearest neighbor and discriminant analysis. The study was used with GTCC, MFCC and basic spectral features, and 99.81% accuracy rate was obtained.

Iqbal et al. [12], used machine learning methods to detect asthma. They used the hunter search algorithm in the feature extraction and selection stages during the analysis of asthma and non-asthmatic voices. Then, they used the DQNN classifier to distinguish asthma, wheezing and normal speech. The results show that the proposed method achieves better results than traditional methods.

Barua et al. [13] proposed a feature extractor based on Attractive and Repulsive Center Symmetric Local Binary Pattern (1D-ARCSLBP) to detect asthma from cough sounds. The dataset used included cough sounds obtained from 511 asthma patients and 815 non-asthmatic participants. TQWT was used in feature extraction. In the feature selection stage, neighboring component analysis (NCA) and in the classification stage, SVM were used to detect asthma and non-asthmatic sounds. The proposed model achieved 98.24% accuracy in ten-fold cross-validation method and 96.91% accuracy in one participant-out cross-validation method. Topaloglu et al. [14] created training images using the mel spectrogram technique. They aimed to detect asthma by proposing a pre-trained model that extracts features using these. They used iterative neighboring component analysis (INCA) in the feature selection phase to reduce the data size. Heat maps were created to highlight the regions in the mel spectrogram images. After the pre-processing phase in the proposed model, they performed training with Attention Resnet18. Then, feature extraction, feature selection and classification phases were performed. The proposed model reached 99.73% accuracy rate with SVM. Kumar et al. [15] used deep learning to recognize lung diseases from cough spectrograms. The study used a dataset collected from 112 patients in a pediatric clinic in India. Deep CNN was proposed to classify eight different lung diseases. The proposed model performed 50% better than the traditional methods. Yue et al. [16] used cough sounds to detect childhood asthma and pneumonia. The dataset used in the study was obtained from hospitals and publicly available datasets. The dataset consists of voice recordings of asthma and pneumonia patients and consists of 850 cough recordings. MFCC was used to increase the classification accuracy. As a result, 90% accuracy rate was obtained with support vector machine (SVM). Naqvi et al. [17] aimed to classify wheezing as normal, asthma and pneumonia in order to analyze lung sounds. Data were collected from 1,000 asthma patients, 1,000 pneumonia patients and 1,000 healthy individuals at the Pakistan Institute of Medical Sciences. 10-second recordings were obtained from the participants' right lungs and nostrils using electronic microphones and stethoscopes. 96.70% accuracy rate was obtained with the SVM classifier. Porter et al. [18] aimed to detect common respiratory tract diseases in children based on cough sounds. The study was conducted on 585 children ranging in age from 29 days to 12 years. The children were reached from emergency services, inpatient services and outpatient units. The obtained dataset includes diseases such as asthma, bronchitis and pneumonia. Positive Percent Agreement (PPA) and Negative Percent Agreement (NPA) values of 97.91% for asthma, 87.85% for pneumonia, 83.82% for lower respiratory tract diseases and 84.81% for bronchiolitis were reached. Turkcetin et al. [19] used artificial neural networks to detect lung diseases. In the study, voice signals of 58 patients who applied to the university hospital from Afyonkarahisar city were used. 30 of the patients were male, 28 were female and their ages ranged from 35 to 85. The dataset consisted of 22 asthmas, 22 COPD and 14 pneumonia patients. Sounds such as cough, breathing and the letter /a/ were recorded using smartphones. Three datasets were created with these sounds in the study. Spectrogram images were obtained with MFCC and fast Fourier transform techniques. These images were trained with artificial neural networks and a high success rate was achieved. Sen et al. [20] aimed to develop a system that can analyze lung sounds in order to distinguish similar respiratory diseases such as COPD and asthma. A dataset consisting of 50 people, 30 asthma and 20 COPD patients, was used in the study. As a result of this study, 100% classification accuracy was achieved in asthma and 95% in COPD with the SVM classifier. Balamurali et al. [21] developed a sound-based classification model to distinguish healthy and asthmatic children. The dataset used in the study included cough sounds from 997 asthmatic and 1032 healthy children and "A" sounds from 80 asthmatic and 253 healthy children. The proposed model achieved 95.3% accuracy in the classification of cough sounds and 72.2% accuracy in the classification of /A:/ sounds.

2.1. Literature Gap

The identified literature gaps are:

- The studies were generally conducted on audio signals. In this study, spectrograms created instead of audio signals were used to eliminate this deficiency in the literature. A more detailed and rich analysis was made possible by using time, frequency and amplitude components.
- In this study, a new CNN model was proposed, unlike traditional CNNs. The performance of the proposed model was tested and a high accuracy rate was obtained compared to traditional methods.

2.2. Novelities

- In this study, a new CNN architecture is proposed, unlike traditional CNN models.
- In the study, spectrograms containing time, amplitude and frequency data of cough sounds were used instead of audio data.
- The developed model can be applied in the early diagnosis of respiratory diseases.

2.3. Contributions

- In this study, a new CNN was designed to classify spectrogram data of cough sounds. The developed artificial intelligence-based model contributes to the early diagnosis of respiratory diseases such as asthma.
- The developed model was tested on cough sounds recorded from patients at Fırat University Hospital. The obtained dataset consists of voice recordings collected from 1192 patients.
- Studies in the literature have generally been conducted on voice signals. In this study, the performance of the use of spectrograms was analyzed.

3. The Proposed Models

In this research, two new models have been proposed for asthma sound classification and these models are deep learning model and deep feature engineering model. The detalis of these models have been explaied in below.

3.1. The presented TwoConvNeXt

The main purpose of this study is to introduce the newly designed TwoConvNeXt model. The proposed model aims to achieve high classification performance with a small number of parameters. The block diagram explaining the proposed TwoConvNeXt model is presented in Figure 1. Also, as seen in Figure 1, the proposed model consists of 4 parts: the stem block, main block, downsampling block and output sections.

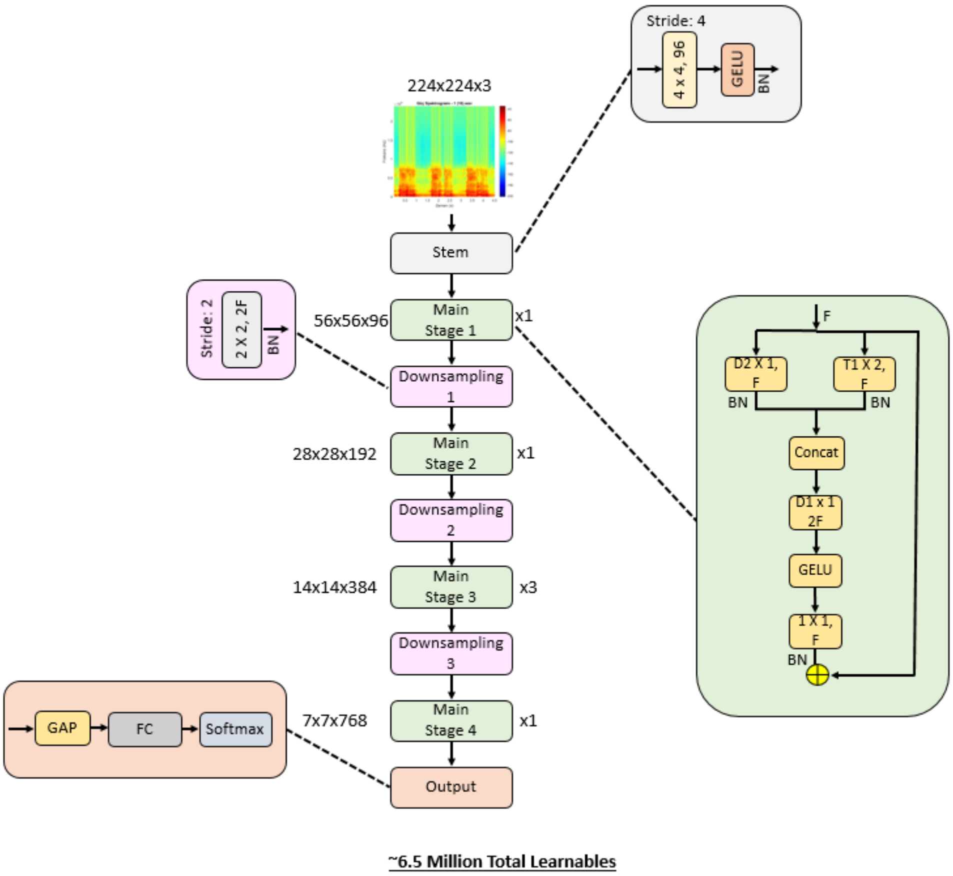

Fig.1. Diagram of the proposed TwoConvNeXt model

Figure 1 shows the stem, main, downsampling and output sections that make up the model. In the stem block, images with dimensions of 224×224×3 are taken as input. The convolution layer is set with a stride value of 4, allowing the patchify process to occur. At this stage, a 4×4 convolution, GeLU activation function, and batch normalization techniques are utilized. This process is illustrated in Figure 3. The structure of the proposed model is shown in the second step, the main block. Two different convolution blocks are used in this stage. D2×1 and T1×2 convolution operations are used to extract different feature maps from the input data. Then, by combining the outputs obtained from different convolution operations, the F-dimensional filter is increased to 2F-dimensional. Convolution with a 2F filter number of 1×1 dimensions is applied to the obtained data. In this way, the number of channels is expanded and the model can learn more features. Then, GELU activation function is used, F filter convolution is applied in 1×1 dimensions. This allows the number of channels expanded in the previous step to be narrowed. Then, batch normalization process is performed for the second time. In the last stage, residual connection is applied by adding the input data to the final output. In this way, the learning process is accelerated and the gradient loss is reduced. In the third stage, downsampling process is applied. In this block, a 2×2 Depthwise Convolution is used. The stride is set to 2, and the input of size F is transformed into a size of 2F. And in the last stage, the output stage, global average pooling (GAP), fully connected (FC) and Softmax operations are applied. The transition table of the proposed TwoConvNeXt model is shown in Table 1.

Table 1. Transition table of the proposed TwoConvNeXt model

Layer	Input	Operation	Output
Stem	224 × 224 × 3	4 × 4, 96, GELU, BN, stride: 4	56 × 56 × 96
Main 1	56 × 56 × 96	$\begin{bmatrix} Concat(2\times 1, 96, T1\times 2, 96) \\ 1\times 1, 192, \\ 1\times 1, 96 \end{bmatrix} \times 1$	56 × 56 × 96
Downsampling 1	56 × 56 × 96	GELU, 2 x 2, 192, stride: 2, BN	28 × 28 × 192
Main 2	28 × 28 × 192	$\begin{bmatrix} Concat(2\times 1, 192, T1\times 2, 192) \\ 1\times 1, 384, \\ 1\times 1, 192 \end{bmatrix} \times 1$	28 × 28 × 192
Downsampling 2	28 × 28 × 192	GELU, 2 x 2, 384, stride: 2, BN	14 × 14 × 384
Main 3	14 × 14 × 384	$\begin{bmatrix} Concat(2\times 1, 384, T1\times 2, 384) \\ 1\times 1, 768, \\ 1\times 1, 384, \end{bmatrix} \times 3$	14 × 14 × 384
Downsampling 3	14 × 14 × 384	GELU, 2 x 2, 768, stride: 2, BN	7 × 7 × 768
Main 4	7 × 7 × 768	$\begin{bmatrix} Concat(2\times 1, 768, T1\times 2, 768) \\ 1\times 1, 1536, \\ 1\times 1, 768 \end{bmatrix} \times 1$	7 × 7 × 768
Output size	7 × 7 × 768	GELU, GAP, FC, Softmax	Number of classes
Total learnable parameters			6.5 Million

TwoConvNeXt shown in Table 1 contains approximately 6.5 million parameters. And its general equation is as follows:

$$\text{TwoConvNeXt} = F: (96,192,384,768), R: (1,1,3,1) \tag{1}$$

In the mathematical definition of TwoConvNeXt, F represents the number of filters and R represents the number of repetitions. The steps of the proposed model are as follows:

Step 1: Apply stem step to each image data of size 224×224×3.

$$final_1 = BN\left(G\left(C^{96}_{4\times 4, Stride:4}(Im)\right)\right) \tag{2}$$

Here *final₁:* Output of Stem Block, *BN:* Batch Normalization, *G:* GELU function, *C:* Convolution and Im: Image.

In this step, a tensor of size 56×56×96 is obtained.

Step 2: After obtaining the tensor of size 56×56×96, the first step of the main stage is applied. The main block passes the incoming input in size F through 2×1 Depthwise Convolution and 1×2 Transpose Convolution operations, respectively. Batch normalization is applied after each convolution operation. The outputs obtained from these two operations are combined in the Concat layer. After the Concat layer, a 1×1 Depthwise Convolution using the GeLU activation function is applied. Finally, a 1×1 Convolution operation is performed after the batch normalization operation. The obtained output is combined with the input data in the summation layer. The mathematical equation is as follows:

$$final_2 = BN\left(C^{96}_{1\times 1}\left(G\left(C^{96}_{1\times 1}\left(Concat\left(C^{96}_{2\times 1}(final_1), C^{96}_{T1\times 2}(final_1)\right)\right)\right)\right)\right) + final_1 \tag{3}$$

At this stage, the Main 1 step is defined. Here $final_2$: is the output of the Main 1 block.

Step 3: Apply the first downsampling step and convert the tensor size to 28×28×192.

$$final_3 = BN\left(C^{192}_{2\times 2, Stride:2}\left(G(final_{2)}\right)\right) \tag{4}$$

Here $final_3$: represents the tensor produced in size 28×28×192.

Step 4: Apply the second step of the main step.

$$final_4 = BN\left(C^{192}_{1\times 1}\left(G\left(C^{192}_{1\times 1}\left(Concat\left(C^{192}_{2\times 1}(final_3), C^{192}_{T1\times 2}(final_3)\right)\right)\right)\right)\right) + final_3 \tag{5}$$

Here, a feature map of size 28×28×192 is created.

Step 5: Apply the second downsampling step and convert the tensor size to 14×14×384.

$$final_5 = BN\left(C^{384}_{2\times 2, Stride:2}\left(G(final_{4)}\right)\right) \tag{6}$$

Here $final_5$ represents the 14×14×384 tensor.

Step 6: Apply the third step of the main step to create a new tensor. Repeat this step 3 times.

$$final_6 = \left(BN\left(C_{1\times1}^{384}\left(G\left(C_{1\times1}^{384}\left(Concat\left(C_{2\times1}^{384}(final_5), C_{T1\times2}^{384}(final_5)\right)\right)\right)\right)\right) + final_5\right)^3 \tag{7}$$

Here, a feature map of size 14×14×384 is created. This steps repated three times to create a robust feature map.

Step 7: Apply the final downsampling step and convert the tensor size to size 7×7×768.

$$final_7 = BN\left(C_{2x2,Stride:2}^{768}\left(G(final_{6)})\right)\right) \tag{8}$$

Here final7 represents the tensor of size 7×7×768.

Step 8: Apply the final, fourth step of the main step.

$$final_8 = BN(C_{1x1}^{768}\left(G(C_{1x1}^{1536}\left(Concat\left(BN\left(C_{2x1}^{768}(final_7)\right), BN\left(C_{T1x2}^{768}(final_7)\right)\right)\right)\right)\right) \tag{9}$$

Here, a feature map of 7×7×768 size is created.

Step 9: Finally, apply the output stage to obtain the classification outputs. In this stage, Global Average Pooling (GAP), Fully Connected Layer (FC) and Softmax layers are used. The mathematical equation is as follows:

$$Out = Softmax\big(FC(GAP)\big) \tag{10}$$

Here is *out:* classification output.

The TwoConvNeXt model, all stages of which are defined with the equations above, has 6.5 million parameters and is considered a lightweight model.

3.2. The proposed TwoConvNeXt-based Deep Feature Engineering Model

A. Material

In this study, the dataset used in the study conducted by Barua et al. [13] was used. The dataset consists of 1,192 cough sounds obtained from a total of 1,192 patients, including 541 asthma patients and 651 individuals without respiratory diseases, who were admitted to the Pulmonology Clinic of Fırat University Hospital. The resulting dataset contains two classes: asthma and control. The sounds were recorded in clinical settings using different models of mobile phones, with durations ranging from 2 to 8 seconds, and the sampling frequencies of the phones were either 44.1 kHz or 48 kHz. The study was approved by the institutional review board, and informed consent was obtained from every participant.

The dataset used in the study was obtained from Fırat University Hospital. This situation may limit the performance of the model under normal conditions. However, the hospital serves as a regional center. It provides service to many patients coming from neighboring provinces other than Elazig. Therefore, although the data was collected from a single center, the dataset includes individuals with various demographic characteristics. The dataset created includes different gender and age groups, and the age range varies between 18 and 70. This situation also provides a limited variety. Since a sufficient number of asthma patients were reached at Fırat University Hospital in the region where the study was conducted, the studies were limited to this institution. In future studies, it is aimed to increase the generalizability of the model by collecting data from different regions and institutions.

B. Deep Feature Engineering Method Based on TwoConvNeXt

In this study, it is aimed to distinguish between asthmatic and healthy individuals using the obtained sound data. In this study, a feature engineering method based on a new model, TwoConvNeXt, was developed. The aim of this study is to increase the classification performance of the developed model and to contribute to deep feature engineering. In the study, firstly, the dataset was trained on MobileNetV2 and the developed TwoConvNeXt models. Then, feature extraction process is performed with global average pooling (GAP) and dropout layers on MobileNetV2 and previously trained TwoConvNeXt models. Then, the most significant features are selected using neighboring component analysis (NCA) and the classification process is performed in the last stage. When the results of the two convolutional neural networks (CNN) are compared, it is seen that the proposed TwoConvNeXt model gives significantly higher results. The model proposed in the study is shown in Figure 2.

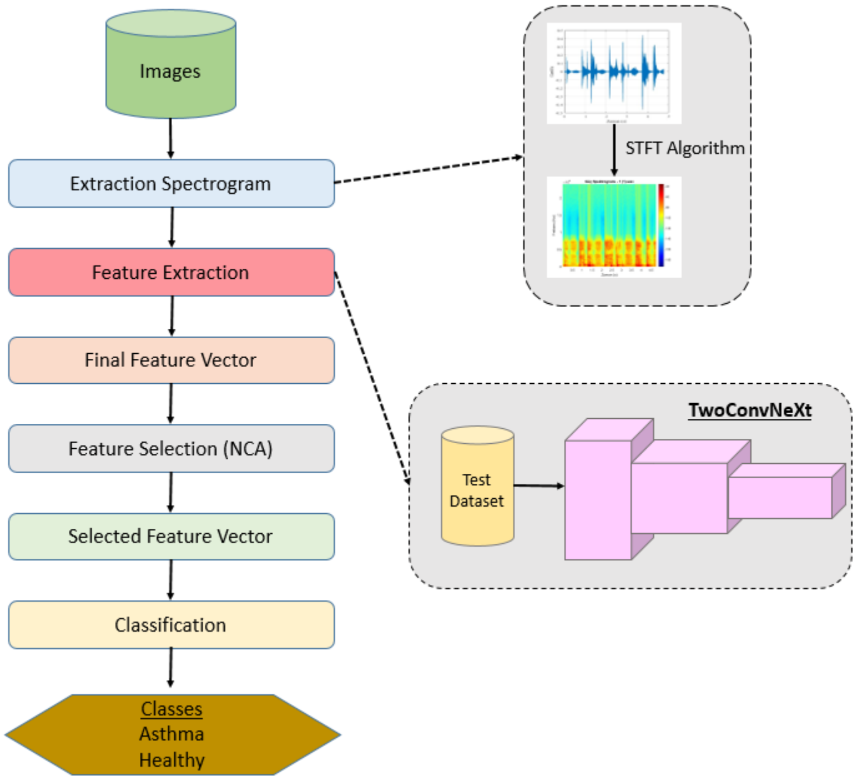

Fig.2. Proposed model

First, spectrograms showing the frequency components of the sounds over time were obtained. The short-time fourier transform (STFT) algorithm was used to obtain the spectrograms. STFT is an efficient method used to see how the frequency of the sound signal changes over time [22]. The parameters used in the spectrogram generation phase affect the time-frequency resolution of the signals. A careful selection was made considering this situation. The window size used in the study is 256 and 50% overlap is frequently encountered in the literature. These parameters provide balance in terms of time frequency and preserve frequency components. It has been observed that choosing a high window size reduces the time resolution. As a result of the preliminary experiments carried out to maintain this balance, the window size was determined as 256 and the overlap rate as 128. It is seen that positive results are obtained in the classification performance of the model by using these parameters.

The obtained spectrograms were tested on CNNs widely used in the literature and generally accuracy rates of 99% were achieved [23]. Based on this, a new CNN model was proposed and its performance was tested. It was observed that the proposed CNN reached a high accuracy rate of 99.6%. This shows that the proposed model exhibits higher performance compared to traditional methods.

Step 1: Firstly, spectrograms were created by applying short-time fourier transform (STFT) to the sounds in the dataset. As a result of this process, a visually-based dataset was formed. Spectrograms are a type of graph that shows the change of a signal over time. In the spectrogram creation stage, the parameters were determined as window:256, noverlap:128, nfft:512.

Step 2: The newly created image-based dataset is used for feature extraction with the TwoConvNeXt CNN model. In the second section, all the blocks of the proposed TwoConvNeXt architecture are explained in detail.

Step 3: After the feature extraction process, the final feature vector arranged in a format that the model can process is obtained.

Step 4: The most meaningful features are selected from the extracted features. At this stage, the NCA algorithm is used. The NCA algorithm is an algorithm that is used to select the most appropriate features in order to improve the classification result. Here, unnecessary features are eliminated and only important features are used in order to make a more accurate decision.

Step 5: After the model is trained and optimized, the classification process is performed. In this step, each input is classified as asthma and healthy as the decision mechanism of the model.

4. Experimental Results

In this study, the TwoConvNeXt model was developed with MATLAB (version 2023a) Deep Network Designer. The developed CNN has 82 layers and 93 connections. In addition, the training parameters used are as shown in Table 2.

Table 2. Training parameters

Solver	SGDM (Stochastic Gradient Descent Momentum)
Initial learning rate	0.01
Training and validation ratio	80:20 randomized
Mini-batch size	64
Maximum epochs	22

The developed model was first trained on the dataset obtained from the cough sound spectrogram images. Then, feature selection was performed with neighboring component analysis (NCA) using machine learning techniques. Then, the classification process was performed.

In this study, the dataset is divided into training and testing. A separation ratio of 80% is determined for training and 20% for testing. The traininig and validation curves have been illustrated in Figure 3.

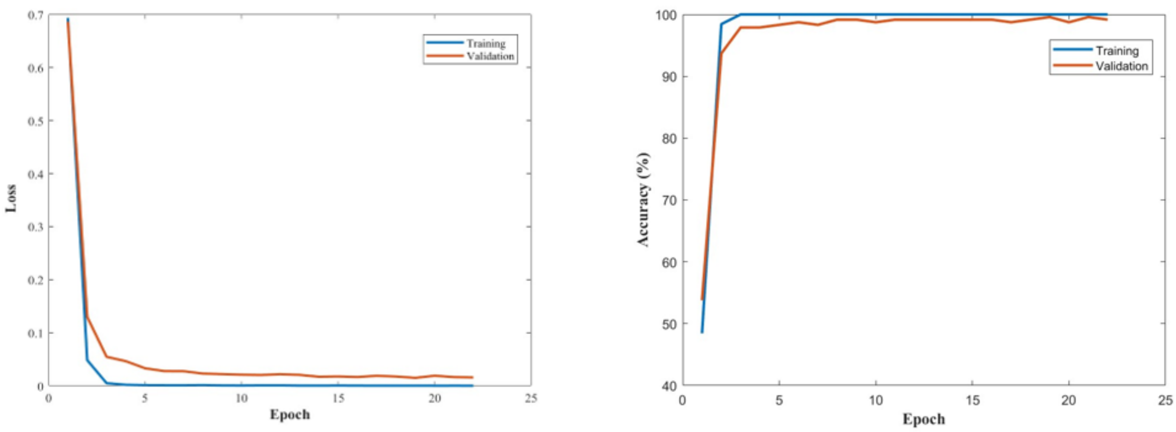

Fig.3. Training and validation curves

Figure 3 illustrated the training and validation curves on cough dataset used. These curves provide information about the progress of the model in the learning process and its ability to generalize on the test data. This shows the robustness and reliability of the proposed model. Here, the final validation loss value was obtained as 0.016 and the final validation accuracy value was obtained as 99.15%.

In this study, the performance of the TwoConvNeXt model developed for sound data was thoroughly examined. To evaluate the classification performance of the model, a confusion matrix was used, and additional performance metrics such as accuracy, F1 score, precision, and recall were calculated.

By deploying the obtained trained CNN, we have computed the test results (see Figure 2), the computed confusion matrix has been demonstrated in Figure 4.

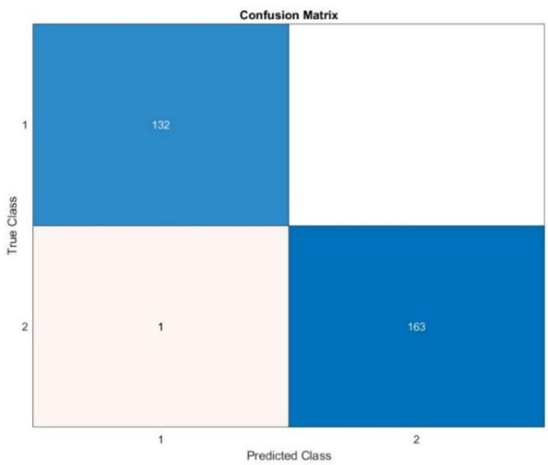

Fig.4. Confusion matrix

Per Figure 4, we have computed four performance evaluation metrics and these metrics are (i) accuracy, (ii) precision, (iii) recall and (iv) F1-score. The calculated test results have been listed Table 3.

Table 3. Performance metrics achieved by the model

Metric	Accuracy (%)
Accuracy	99.66
Precision	100
Recall	99.25
F1 Score	99.62

The performance metrics obtained by the developed CNN model show that the model has achieved a high rate of success. It is shown in Table 3 that the model has achieved very successful results. The developed model has an accuracy rate of 99.6%. As a result, this study shows that the developed CNN model achieves high and reliable results compared to existing architectures. In future studies, this model can be tested on different datasets and the generalization ability of the model can be evaluated comprehensively.

In this study, the results obtained using MobileNetv2 and the proposed TwoConvNeXt model were compared. The accuracy rates obtained in different classifiers of Mobilenetv2 and TwoConvNeXt models are shown graphically in the Figure 5.

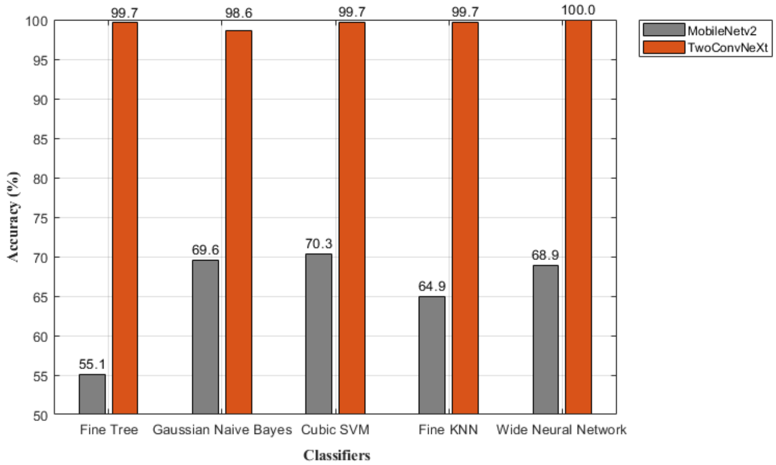

Fig.5. Comparison of mobilenetv2 and TwoConvNeXt models

5. Discussions

This study aims to analyze the performance of the TwoConvNeXt model. Spectrograms obtained from cough sounds were used in the study. After training the model on the dataset, performance metrics such as accuracy, F1 score, sensitivity, and recall were calculated. The developed model reached 99.6% accuracy. The highest performance was achieved with the Wide neural network classifier, with 100% accuracy.

The developed model was tested on only one dataset. This has limitations in terms of generalizability. Considering this situation, the developed model will be tested on different datasets in our future studies. And cross-validation methods will also be applied. In addition, in this study, a separation of training and validation was applied in order to prevent overfitting. Thus, it was tried to prevent the model from showing excessive learning. The dataset was divided into training and test data in a balanced manner. Here, the success rate of the model in validation is 99.15%. This result shows that the model provides good performance not only on training data but also on data it has not seen before.

In the study, the neighbor component analysis (NCA) method was preferred in the feature selection phase. The reason for this is to increase the classification performance. NCA reshapes the feature space using class labels. Thus, it highlights the features that indicate the maximum separation between the classes. This is advantageous in the separation of two classes, asthmatic and healthy individuals. Another alternative method, PCA, performs dimension reduction independently of the class labels. As a result, it does not guarantee the selection of the most informative features. In addition, LASSO may require more aparameter adjustments depending on the complexity of the model. In future studies, it will be evaluated as a comparison of different methods on different datasets and classifiers.

The obtained spectrograms were tested on CNNs widely used in the literature and generally accuracy rates 99% were achieved. Based on this, a new CNN model was proposed and its performance was tested. It was observed that the proposed CNN reached a high accuracy rate of 99.6%. This shows that the proposed model exhibits higher performance compared to traditional methods. Table 4 presents comparative studies in the literature.

Table 1. Comparative studies in the literature

Study	Method(s)	Dataset	Results (%)
Topaloglu et al. [14]	ResNet18 based model	203 participants	Accuracy: 99.7%
Garcia-Ordas et al. [24]	Variational Convolutional Autocoder + CNN	126 participants, 920 audio recordings	Precision: 98.8 %
Kuo et al. [25]	Radial basis function neural network (RBFNN)	93 participants	Accuracy: 96.8%
Yahyaoui et al. [26]	Deep Neural Network (DNN) and K-nearest Neighbors (KNN)	212 participants	Accuracy: 95%
Tariq et al. [27]	2D CNN model	11960 audio samples	Accuracy: 97%
Jayalakshmy et al. [28]	CNN + Decision tree	596 audio samples	Accuracy: 96.7%
Fraiwan et al. [29]	CNN + BDLSTM 213 patients	1483 audio recordings	Accuracy: 99.6%
Varshni et al. [30]	DenseNet-169	ChestX-ray14	AUC: 80%
Roy et al. [31]	AsthmaSCELNet	Chest wall lung sound database (CWLSD)	Accuracy: 98.5%
Our proposed framework	TwoConvNeXt	1192 participants	Accuracy: 99.6%

5.1. Advantages

- The developed model shows an effective performance with a high accuracy rate of 99.56%. This increases its usability in the health field.
- In the study, a spectrogram was used instead of a sound signal, which allowed for more detailed analysis of time, frequency and amplitude components.
- Due to the high accuracy obtained, early diagnosis of respiratory diseases such as asthma is possible.

5.2. Future Directions

- The system will be tested with additional healthcare providers to evaluate usability across various clinical roles.
- Explanation methods will be improved to clarify model decisions through enhanced visualization techniques and textual summaries.
- Multi-center clinical trials will be conducted to confirm performance in diverse healthcare settings.
- The tool will be integrated with electronic health record systems to enable seamless workflow support.
- User feedback on system clarity and ease of use will be collected through structured questionnaires and semi-structured interviews.
- We will optimize the model architecture to reduce computational load and enable real-time spectrogram analysis.
- Real-time spectrogram generation and classification will be implemented for point-of-care applications.
- The dataset will be expanded with a wider range of cough recordings to improve model generalizability.
- Regulatory compliance and data privacy frameworks will be reviewed to ensure clinical readiness and patient safety.
- Training materials and user manuals will be developed to facilitate adoption by healthcare staff.
- Performance monitoring dashboards will be created to track model accuracy and user engagement post-deployment.

5.3. Limitations

- The age and gender distribution in the dataset used is limited. Therefore, no comparison was made between different age ranges and genders.
- In the study, only cough sounds were used to detect asthma. Symptoms such as wheezing and breathing were not included.

6. Conclusions

Within the scope of this study, an artificial intelligence-based model was developed for the purpose of detecting anomalies from cough sounds. The performance evaluation of the presented new CNN model was performed on the obtained dataset. The accuracy rate of the model, which reached 99.56%, proves the reliability of the presented new CNN model. The obtained results show that deep learning models are effective in the detection and early diagnosis of respiratory diseases. Early diagnosis of asthma and similar respiratory diseases is of great importance in terms of minimizing the risks related to the disease and increasing the quality of life of patients. This study also contributes to the literature by showing the applicability of artificial intelligence in the field of health. Considering real world applications, noisy environments and physiological variations among patients can be observed. Therefore, the developed model is expected to show high performance despite such negative variations. The cough sounds used in this study were obtained from individuals of different genders and different age groups. In addition, obtaining the sound recordings from different mobile devices created differences in the recording quality. This increases the robustness of the proposed model against variations.

In future studies, the developed new CNN model can be tested on different datasets. In this way, the generalization ability and applicability of the proposed model can be analyzed in more detail. In addition, it can be tested on different diseases, not just asthma. The application of the model in different fields, not only limited to the field of health, can also enable the performance of the system to be analyzed in the same way. As a result, the proposed deep learning model aims to detect anomalies in cough sounds. This also forms a basis for future studies. The results obtained show that the proposed model can be used not only in academic but also in clinical fields.

References

[1] D. Mutolo, L. Iovino, E. Cinelli, F. Bongianni, and T. Pantaleo, "Physiology of the Cough Reflex: Sensory and Mechanical Features," *Cough: Pathophysiology, Diagnosis and Treatment,* pp. 3-21, 2020.

[2] K. F. Chung *et al.*, "Cough hypersensitivity and chronic cough," *Nature Reviews Disease Primers,* vol. 8, no. 1, p. 45, 2022.

[3] U. Frey and B. Suki, "Complexity of chronic asthma and chronic obstructive pulmonary disease: implications for risk assessment, and disease progression and control," *The Lancet,* vol. 372, no. 9643, pp. 1088-1099, 2008.

[4] J. Kavanagh, D. J. Jackson, and B. D. Kent, "Over-and under-diagnosis in asthma," *Breathe,* vol. 15, no. 1, pp. e20-e27, 2019.

[5] A. Ijaz *et al.*, "Towards using cough for respiratory disease diagnosis by leveraging Artificial Intelligence: A survey," *Informatics in Medicine Unlocked,* vol. 29, p. 100832, 2022.

[6] S. Preethi, A. Revathi, and M. Murugan, "Exploration of cough recognition technologies grounded on sensors and artificial intelligence," *Internet of Medical Things for Smart Healthcare: Covid-19 Pandemic,* pp. 193-214, 2020.

[7] M. Kilic *et al.*, "GCLP: An automated asthma detection model based on global chaotic logistic pattern using cough sounds," *Engineering Applications of Artificial Intelligence,* vol. 127, p. 107184, 2024.

[8] G. Shruthi, K. R. PM, A. S. Naidu, A. Kumari, C. Sravanti, and P. Gayathri, "Detection of Lung Disease using Deep Learning Approaches," in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, 2023: IEEE, pp. 1-8.

[9] W. Wang, Q. Shang, and H. Lu, "Automatic COVID-19 detection from cough sounds using multi-headed convolutional neural networks," *Applied Sciences,* vol. 13, no. 12, p. 6976, 2023.

[10] M. Kuluozturk *et al.*, "DKPNet41: Directed knight pattern network-based cough sound classification model for automatic disease diagnosis," *Medical Engineering & Physics,* vol. 110, p. 103870, 2022.

[11] S. K. Prajapati, T. S. Choudhary, and S. Mishra, "Early detection of lung disease using multi-class classifiers," in *2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON)*, 2023: IEEE, pp. 01-06.

[12] M. A. Iqbal, K. Devarajan, and S. M. Ahmed, "Real time detection and forecasting technique for asthma disease using speech signal and DENN classifier," *Biomedical Signal Processing and Control,* vol. 76, p. 103637, 2022.

[13] P. D. Barua *et al.*, "Automated asthma detection in a 1326-subject cohort using a one-dimensional attractive-and-repulsive center-symmetric local binary pattern technique with cough sounds," *Neural Computing and Applications,* pp. 1-15, 2024.

[14] I. Topaloglu *et al.*, "Explainable attention ResNet18-based model for asthma detection using stethoscope lung sounds," *Engineering Applications of Artificial Intelligence,* vol. 126, p. 106887, 2023.

[15] A. Kumar, K. Abhishek, C. Chakraborty, and N. Kryvinska, "Deep learning and internet of things based lung ailment recognition through coughing spectrograms," *IEEE Access,* vol. 9, pp. 95938-95948, 2021.

[16] L. Yue and W. Xu, "Automatic classification of childhood asthma and pneumonia based on cough sound analysis," in *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, 2021: IEEE, pp. 779-783.

[17] S. Z. H. Naqvi, M. Arooj, S. Aziz, M. U. Khan, and M. A. Choudhary, "Spectral analysis of lungs sounds for classification of asthma and pneumonia wheezing," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 2020: IEEE, pp. 1-6.

[18] P. Porter *et al.*, "A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children," *Respiratory research,* vol. 20, pp. 1-10, 2019.

[19] A. Ö. Türkçetin, T. Koç, and Ş. Çilekar, "The Use of ANN in the Sound Detection of Lung Diseases: Example of COPD, Asthma, Pneumonia," in *2023 31st Signal Processing and Communications Applications Conference (SIU)*, 2023: IEEE, pp. 1-4.

[20] I. Sen, M. Saraclar, and Y. P. Kahya, "Differential diagnosis of asthma and COPD based on multivariate pulmonary sounds analysis," *IEEE Transactions on Biomedical Engineering,* vol. 68, no. 5, pp. 1601-1610, 2021.

[21] B. BT *et al.*, "Asthmatic versus healthy child classification based on cough and vocalised/ɑ:/sounds," *The Journal of the Acoustical Society of America,* vol. 148, no. 3, pp. EL253-EL259, 2020.

[22] H. Jeon, Y. Jung, S. Lee, and Y. Jung, "Area-efficient short-time fourier transform processor for time–frequency analysis of non-stationary signals," *Applied Sciences,* vol. 10, no. 20, p. 7208, 2020.

[23] J. Amoh and K. Odame, "Deep neural networks for identifying cough sounds," *IEEE transactions on biomedical circuits and systems,* vol. 10, no. 5, pp. 1003-1011, 2016.

[24] M. T. García-Ordás, J. A. Benítez-Andrades, I. García-Rodríguez, C. Benavides, and H. Alaiz-Moretón, "Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data," *Sensors,* vol. 20, no. 4, p. 1214, 2020.

[25] H.-C. Kuo, B.-S. Lin, Y.-D. Wang, and B.-S. Lin, "Development of automatic wheeze detection algorithm for children with asthma," *IEEE Access,* vol. 9, pp. 126882-126890, 2021.

[26] A. Yahyaoui and N. Yumuşak, "Deep and machine learning towards pneumonia and asthma detection," in *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2021: IEEE, pp. 494-497.

[27] Z. Tariq, S. K. Shah, and Y. Lee, "Lung disease classification using deep convolutional neural network," in *2019 IEEE international conference on bioinformatics and biomedicine (BIBM)*, 2019: IEEE, pp. 732-735.

[28] S. Jayalakshmy, B. L. Priya, and N. Kavya, "CNN based Categorization of respiratory sounds using spectral descriptors," in *2020 International Conference on Communication, Computing and Industry 4.0 (C2I4)*, 2020: IEEE, pp. 1-5.

[29] M. Fraiwan, L. Fraiwan, M. Alkhodari, and O. Hassanin, "Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory," *Journal of Ambient Intelligence and Humanized Computing,* pp. 1-13, 2022.
[30] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia detection using CNN based feature extraction," in *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)*, 2019: IEEE, pp. 1-7.
[31] A. Roy and U. Satija, "AsthmaSCELNet: A lightweight supervised contrastive embedding learning framework for asthma classification using lung sounds," *entropy,* vol. 1282, p. 100, 2023.

Authors' Profiles

Tugce Keles received her bachelor's degree in digital forensics engineering from Elazig, Fırat University in 2016. Then, she continued her studies in the same department and received her master's degree in 2023. She has been working as a research assistant in the department of digital forensics engineering Engineering at Firat University's Faculty of Technology since 2022. Her main research interests include signal processing and image processing.

Sengul Dogan received the master's degree in bioengineering the Ph.D. degree in electrical and electronics engineering from the Firat University, Elazig, Turkey, in 2007 and 2011, respectively. She is currently a Professor with the Digital Forensics Engineering, Technology Faculty, Firat University. Her main research interests include computer forensics, mobile forensics, image processing, and signal processing. She has been working actively on developing algorithms in machine learning for biomedical data.

Abdul-Hafeez Baig received the master's degree in MIS and MBA from Griffith University, Brisbane, Australia, and the Ph.D. degree in the domain of healthcare and information systems from USQ. Since joining USQ, in January 2004, he has published several refereed publications and has numerous research grants in the domains of health and education regarding technology management. He is very conversant with wireless technology as well as emerging technologies and learning management systems. He teaches information systems concepts to both undergraduate and postgraduate students, including MBA students. He also has numerous publications in academic and scholarly journals and has a vast array of scholarly conference papers, all of which have been focused on the domain of information systems. He has extensive experience in the area of information systems, especially related to the healthcare sector. He is quite interested in wireless healthcare applications, systems analysis and design, adoption, the infusion and diffusion of information technology, m-learning and e-commerce, outsourcing, networking, healthcare and information technology, and the re-engineering of business processes.

Turker Tuncer received the master's degree in electronics and computer sciences and the Ph.D. degree in software engineering from Firat University, Elazig, Turkey, in 2011 and 2016, respectively. He is currently an Associate Professor with the Digital Forensics Engineering, Technology Faculty, Firat University. His main research interests include feature engineering, image processing, signal processing, information security, and pattern recognition. He has been working actively on developing algorithms in machine learning applied to visual surveillance and biomedical data.

How to cite this paper: Tugce Keles, Sengul Dogan, Abdul-Hafeez Baig, Turker Tuncer, "Spectrogram-based Deep Learning Approach for Anomaly Detection from Cough Sounds", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.3, pp.1-12, 2025. DOI:10.5815/ijitcs.2025.03.01