E3S Web of Conferences **723**, 01012 (2026) https://doi.org/10.1051/e3sconf/202672301012
AIEI 2026

An Application in Health Monitoring Systems: A Deep Learning-based Flu Detection model

Khoa Tran[1,*], *Ho-Si-Hung* Nguyen[2,**], *Hung-Cuong* Trinh[3,***], *Vy-Rin* Nguyen[4,****], and *Thi Thanh Quynh* Nguyen[2,†]

[1]Data Science Laboratory, Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam
[2]Faculty of Electrical engineering, University of Science and Technology, The University of Danang, Da Nang 50000, Vietnam
[3]Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam
[4]Department of Semiconductor and Microelectronics, FPT University, Da Nang 50000, Vietnam

Abstract. Monitoring the health status of students within confined spaces is critical for preventing flu outbreaks and ensuring a healthy educational environment. This proposed Flu Detection model is an innovative approach designed to swiftly identify instances of flu among students. Which is a supervised learning model enhanced with Mixture augmentation classifies various flu types with high accuracy, leveraging a robust training dataset. This comprehensive model yields strong experimental results on flu datasets, demonstrating effectiveness in identifying flu cases, showcasing its potential for health monitoring in educational facilities.

1 Introduction

The outbreak of pneumonia caused by the novel coronavirus was detected in Wuhan, China, in December 2019 [1], and the virus was later named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the WHO. SARS-CoV-2 exhibits a higher transmission rate compared to SARS-CoV. As of January 31, 2021, over 101.9 million cases and 2.2 million deaths had been reported globally. The rapid spread of COVID-19 has accelerated the development of deep learning-based healthcare and surveillance techniques aimed at mitigating the virus's spread.

In educational institutions, where large populations interact closely, the risk of disease transmission is significantly high. Therefore, developing effective monitoring systems to detect flu outbreaks in real time is crucial. This paper introduces a Deep Learning (DL)-based Flu Detection model designed to monitor and manage flu outbreaks in educational settings. By leveraging advanced deep learning algorithms, the proposed system aims to provide accurate and timely insights into flu transmission patterns, enabling proactive measures to mitigate disease spread.

Recent advancements in health monitoring systems have focused on various methodologies. An approach for using audio inputs to classify cough sounds as indicative or non-indicative of illness was introduced in the FluSense project [2]. They transformed 1-second audio segments into spectrograms with a bin size of 257, which were then processed using a Convolutional Neural Network (CNN) [3]. Another notable contribution is the framework developed by [4], which classifies cough sounds from COVID-19-positive and COVID-19-negative cases. Their approach involves converting cough sounds into Log-Mel spectrograms using the Wavegram-Log-Mel-CNN model.

Despite these advancements, challenges remain in the deployment of deep learning-based health monitoring systems. There is a need for a model with high accuracy and a low number of parameters that is suitable for real-world applications, such as being embedded in edge devices. Therefore, we propose a deep learning-based flu detection model to address this issue.

The paper is structured as follows: Section 2 describes the methodology and architecture of our Flu Detection model, detailing each phase and the algorithms used. Section 4 evaluates the system's performance through experimentation, presenting results and discussing their implications. Finally, Section 5 summarizes the findings, discusses practical implications, and suggests directions for future research.

2 Proposed Model

As illustrated in Figure 1, our proposed DL-based Flu Detection model is a supervised learning model employed for flu detection. The model comprises our proposed GroupNet and a stack output layers serving for classification task include Linear (128 output units), ReLU, Dropout (with a rate of 0.2), Linear (4 output units), and Softmax layer arranged in sequential order. The final output is a one-hot vector representing probabilities.

The proposed GroupNet architecture is inspired by the MobileNetV2 [5] framework and is designed to achieve a

*e-mail: trandinhkhoa@tdtu.edu.vn
**Corresponding author: e-mail: nhshung@dut.udn.vn
***e-mail: trinhhungcuong@tdtu.edu.vn
****e-mail: RinNV@fe.edu.vn
†e-mail: nttquynh@dut.udn.vn

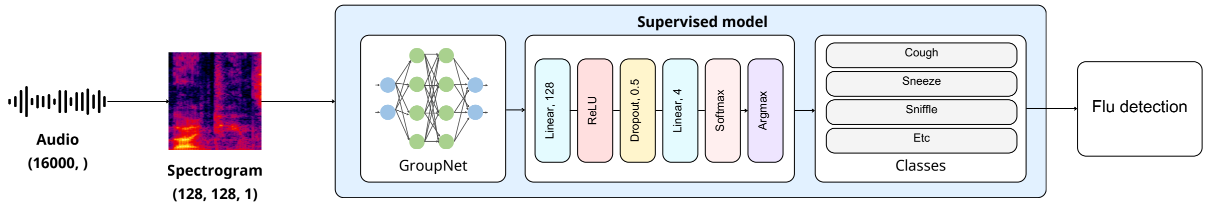

Figure 1. The architecture of the DL-based Flu Detection model

lower parameter count while maintaining competitive performance for flu detection. The main modifications introduced in this network are as follows: (1) the adoption of lightweight Bottleneck blocks with an increasing number of filters to improve hierarchical feature extraction; (2) the integration of Group Normalization (GN) layers after convolutional operations to improve training stability; and (3) the use of a Global Average Pooling 2D layer to reduce spatial dimensions before the classification layer, thereby decreasing model complexity and improving computational efficiency.

Looking in detail at our proposed GroupNet architecture, Conv2d, GN, ReLU, DepthwiseCon2D are represented for 2D convolution layer, GN layer, Rectified Linear Unit (ReLU) activation layer, Depthwise 2D convolution, respectively. In layers, f, g, a indicate filters, groups, alpha. The notation 3×3 or 1×1 following the layer name is used to indicate the kernel size (for example, 3×3 means kernel size is $(3, 3)$). ReLU6 sets the maximum value of ReLU activation to 6.

Residual connections are used to combine the outputs from ReLU6 and the bottneck block, aiming to alleviate the degradation problem, improve gradient flow and enhance learning capacity as well. Moreover, in the botteneck block, GN layers are placed immediately after the Depthwise 2D convolution layer and subsequent 2D convolution layer to stabilize the training process, and reduce internal covariate shift.

The Softmax layer calculates the predicted probabilities for specific groups (Cough, Sneeze, Sniffle, etc). We employ the Cross-Entropy loss function, represented by Equation (1), to compare these predicted probabilities (denoted as $\hat{\mathbf{y}}$) with the true labels $\mathbf{y}$.

$$\mathcal{L}_{entropy} = -\sum_i y_i \cdot \log(\hat{y}_i) \quad (1)$$

To enhance the training of the Flu Detection model, the Mixup algorithm [6] is employed as an augmentation technique applied to spectrograms. The Mixup algorithm operates as follows: Given a pair of input spectrogram samples (x_i, y_i) and (x_j, y_j), where x_i and x_j denote the input spectrograms, and y_i and y_j represent their corresponding labels, the algorithm generates a synthetic sample $(\tilde{x}, \tilde{y})$ through the following operations:

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j$$
$$\tilde{y} = \lambda y_i + (1-\lambda)y_j$$

Here, λ is a mixing coefficient randomly sampled from a Beta distribution with parameters α, i.e., $\lambda \sim \mathrm{Beta}(\alpha, \alpha)$.

2.1 Hyperparameter Optimization

To fine-tune the hyperparameters of our deep learning model, we employed Bayesian Optimization. This technique constructs a probabilistic model of the objective function and uses it to iteratively select promising hyperparameter configurations. Specifically, Bayesian Optimization aims to minimize the validation loss by optimizing the following acquisition function:

$$a(\mathbf{x}) = \mu(\mathbf{x}) + \lambda\sigma(\mathbf{x}), \quad (2)$$

where $\mu(\mathbf{x})$ is the mean prediction of the objective function at hyperparameter configuration $\mathbf{x}$, $\sigma(\mathbf{x})$ is the standard deviation (uncertainty) of the prediction, and λ is a parameter that balances exploration versus exploitation.

For our purposes, we utilized the BayesianOptimization tuner to minimize the validation loss. The tuning process was constrained to a maximum of 50 trials, with each trial executed three times to ensure reliability. The search was performed within a designated directory, with results organized by project name. Bayesian Optimization capitalizes on previous evaluation results to build a probabilistic model, which guides the efficient exploration of the hyperparameter space and aids in pinpointing the optimal model configuration.

3 Dataset

ICBHI 2017 Respiratory Sound Dataset:

The ICBHI 2017 Respiratory Sound dataset [7] is a comprehensive collection of respiratory sound recordings gathered from various environments. It contains 6,898 respiratory cycles from 126 subjects, comprising a total of 920 annotated recordings. These recordings were sampled at rates of 4 kHz, 10 kHz, and 44.1 kHz. Each recording is annotated with labels indicating four respiratory conditions: normal breathing, crackles, wheezes, or both crackles and wheezes, with the annotations provided by medical experts.

FluSense dataset:

The FluSense [2] dataset was compiled from 21,000,000 non-speech audio snippets collected over a 7-month clinical deployment study. These snippets were recorded in

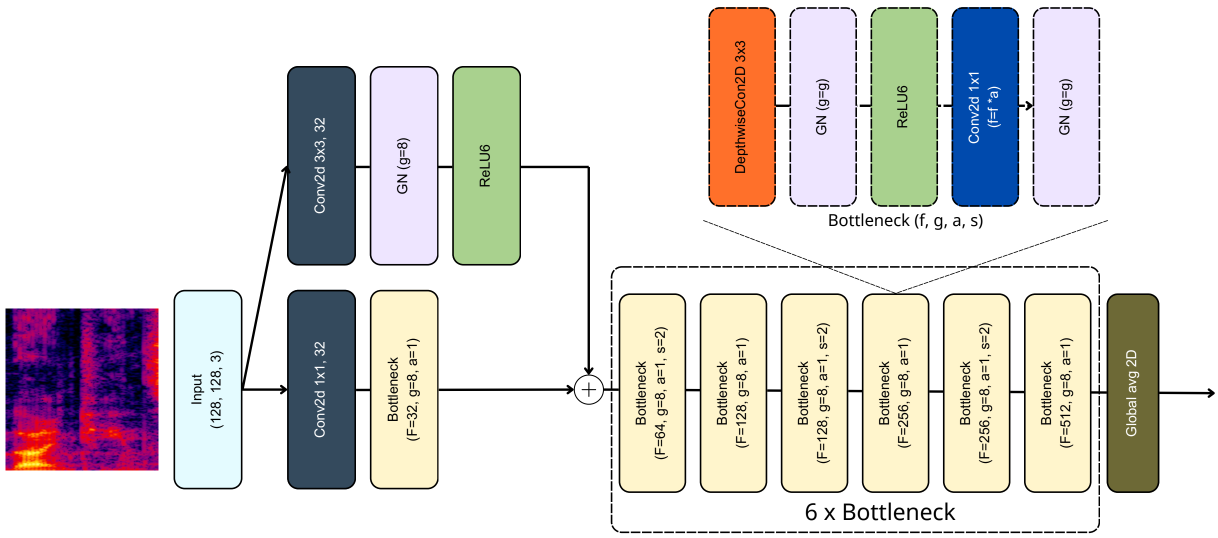

Figure 2. The GroupNet architecture

four hospital waiting rooms and include actual instances of coughing. Additionally, the dataset contains precisely 45,550 seconds of audio from the Audioset dataset, featuring labels including speech, cough, sneeze, sniffle, gasp, silence, and background noise.

4 Performance evaluation

In this section, we detail the experimental setup and the datasets used to train our proposed architectures. We utilize three datasets: The ICBHI 2017 Respiratory Sound dataset [8] to evaluate our proposed supervised model and compare it with state-of-the-art methods, assessing the performance of our backbone architecture, GroupNet.

4.1 Experimental Setup

In Table 1, the parameters used for the training process are presented. The Adam optimizer is utilized to optimize both the supervised and unsupervised models. The learning rate is set to 10^{-3}, and the epoch and batch size are configured as shown in Table 1. We will validate the performance of flu detection and occupancy estimation using various metrics, including Accuracy, Precision, Recall, F1 score, and Specificity.

Table 1. Training Configuration.

	Epoch	Batch size	Optimizer
Flu detection	300	32	Adam

The performance metrics are defined as follows:

- **Accuracy** is the ratio of correctly predicted instances to the total instances:

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

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

- **Precision** is the ratio of correctly predicted positive instances to the total predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall** (or Sensitivity) is the ratio of correctly predicted positive instances to all instances in the actual class:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1 Score** is the harmonic mean of Precision and Recall:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Specificity** is the ratio of correctly predicted negative instances to the total actual negatives:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

4.2 Validation of flu detection

In this section, we first evaluate our proposed supervised model using the benchmark dataset (ICBHI 2017 Respiratory Sound Dataset) to compare its performance against other state-of-the-art models, with a particular focus on validating our proposed GroupNet. Subsequently, we utilize the pretrained supervised model to train and test on the FluSense dataset.

The results in Table 2 highlight the performance of our Supervised model with and without the Mixup data augmentation technique, using Mel spectrogram and STFT as preprocessing methods. Among the evaluated configurations, our Supervised model with Mixup and STFT preprocessing demonstrates the highest overall performance metrics: Accuracy (60.71%), Recall (41.21%), and Specificity (77.85%). This combination consistently outperforms other configurations in Accuracy, Recall, and AS.

Therefore, our Supervised model with Mixup and STFT is selected as the optimal approach for achieving superior classification performance and will be used in subsequent experiments.

Table 2. Performance comparison of our Deep Learning- based Flu Detection mode model with and without the Mixup data augmentation technique.

CNNs	Score	Mel spectrogram	STFT
Our Deep Learning-based Flu Detection without Mixup	Accuracy	52.25	**59.35**
	Recall	28.65	**36.95**
	Specificity	72.85	**80.55**
Our Deep Learning-based Flu Detection with Mixup	Accuracy	58.57	**60.71**
	Recall	35.87	**41.21**
	Specificity	79.59	77.85

Table 3. Performance comparison of our Deep Learning-based Flu Detection model with other state-of-the-art models using ICBHI 2017 Respiratory Sound Dataset.

	Reference	Recall	Specificity	Accuracy
Official, 4-class	[9]	-	-	39.56
	[10]	20.81	78.05	49.43
	[11]	-	-	49.86
	[12]	31.12	69.20	50.16
	[13]	41.32	63.20	52.26
	[14]	26.00	68.00	47.00
	[15]	36.36	71.44	53.90
	Our Supervised model	**43.52**	76.55	**55.75**
Official, 2-class	[10]	33.15	78.05	55.60
	[12]	47.70	69.20	58.45
	[14]	-	-	54.10
	[15]	51.40	71.44	61.42
	Our Supervised model	**53.32**	76.12	**64.04**
Random, 4-class	[16]	58.43	73.00	65.70
	[17]	48.63	84.14	66.38
	[13]	63.69	64.73	64.21
	[15]	54.69	80.45	67.57
	Our Supervised model	61.32	83.23	**70.12**

Table 4. Performance comparison of GroupNet and MobileNetV2 in terms of model size and accuracy on the FluSense dataset.

	GroupNet	MobileNet V2
Size	1.28 MB	9.24 MB
Acc	88.21	86.58

The results in Table 3 demonstrate the effectiveness of our proposed supervised model with Mixup and STFT preprocessing compared to other state-of-the-art methods using the ICBHI 2017 Respiratory Sound Dataset. Following the data splitting methodology of [15], our model achieves superior performance metrics across all classes. Specifically, our model shows improvements in recall, specificity, and accuracy, indicating a more robust and reliable classification performance. These findings validate the efficacy of incorporating Mixup and STFT techniques in enhancing model accuracy and generalization.

We utilize the FluSense dataset to compare our proposed GroupNet model with the state-of-the-art MobileNetV2 [5], which is a popular choice for mobile and microcontroller applications. As shown in Table 4, both models were trained for 20 epochs. The results indicate that our GroupNet, despite its smaller size, achieves higher prediction accuracy compared to MobileNetV2. This makes GroupNet more suitable for real-time precise predictions on resource-constrained devices

5 Conclusion

In summary, our deep learning-based flu detection model plays a pivotal role in monitoring people's health. We employ supervised learning to classify various types of flu based on spectrogram analysis. We propose the GroupNet architecture, which combines large-scale and small-scale features (kernel sizes of 1 and 3) to achieve strong performance. The model is significantly lighter than MobileNetV2 while delivering higher accuracy. Techniques such as mixup augmentation enhance the model's robustness. Overall, our model demonstrates effectiveness in rapidly identifying flu cases. Future work will focus on optimizing the model for real-time processing and integrating it with Internet of Things (IoT) devices for continuous monitoring.

Acknowledgment

This work was supported by University of Science and Technology, The University of Danang, code number of Project: T2025-02-19.

References

[1] A. Deslandes, V. Berti, Y. Tandjaoui-Lambotte, C. Alloui, E. Carbonnelle, J.R. Zahar, S. Brichler, Y. Cohen, Sars-cov-2 was already spreading in france in late december 2019, International journal of antimicrobial agents **55**, 106006 (2020).

[2] F. Al Hossain, A.A. Lover, G.A. Corey, N.G. Reich, T. Rahman, Flusense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **4**, 1 (2020).

[3] M.S. Alam, M. Alam, M. Tufail, M.U. Khan, A. Güneş, B. Salah, F.E. Nasir, W. Saleem, M.T. Khan, Tobset: A new tobacco crop and weeds image dataset and its utilization for vision-based spraying by agricultural robots, Applied Sciences **12**, 1308 (2022).

[4] L.H. Nguyen, N.T. Pham, L.T. Nguyen, T.T. Nguyen, H. Nguyen, N.D. Nguyen, T.T. Nguyen, S.D. Nguyen, A. Bhatti, C.P. Lim et al., Fruit-cov: An efficient vision-based framework for speedy detection and diagnosis of sars-cov-2 infections through recorded cough sounds, Expert Systems with Applications **213**, 119212 (2023).

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 4510–4520

[6] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, mixup: Beyond empirical risk minimization, arXiv preprint arXiv:1710.09412 (2017).

[7] B. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques et al., A respiratory sound database for the development of automated classification, in *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18-21 November 2017* (Springer, 2018), pp. 33–37

[8] B.M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y.P. Kahya, N. Jakovljevic, T.L. Turukalo, I.M. Vogiatzis, E. Perantoni et al., An open access database for the evaluation of respiratory sound classification algorithms, Physiological measurement **40**, 035001 (2019).

[9] N. Jakovljević, T. Lončar-Turukalo, Hidden markov model based respiratory sound classification, in *International Conference on Biomedical and Health Informatics* (Springer, 2017), pp. 39–43

[10] G. Chambres, P. Hanna, M. Desainte-Catherine, Automatic detection of patient with respiratory diseases using lung sound analysis, in *2018 International Conference on Content-Based Multimedia Indexing (CBMI)* (IEEE, 2018), pp. 1–6

[11] G. Serbes, S. Ulukaya, Y.P. Kahya, An automated lung sound preprocessing and classification system based onspectral analysis methods, in *International Conference on Biomedical and Health Informatics* (Springer, 2017), pp. 45–49

[12] Y. Ma, X. Xu, Q. Yu, Y. Zhang, Y. Li, J. Zhao, G. Wang, LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm, in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (IEEE, 2019), pp. 1–4

[13] Y. Ma, X. Xu, Y. Li, LungRN+ NL: An Improved Adventitious Lung Sound Classification Using Non-Local Block ResNet Neural Network with Mixup Data Augmentation., in *Interspeech* (2020), pp. 2902–2906

[14] L. Pham, H. Phan, R. Palaniappan, A. Mertins, I. McLoughlin, Cnn-moe based framework for classification of respiratory anomalies and lung disease detection, IEEE journal of biomedical and health informatics **25**, 2938 (2021).

[15] J. Li, J. Yuan, H. Wang, S. Liu, Q. Guo, Y. Ma, Y. Li, L. Zhao, G. Wang, Lungattn: advanced lung sound classification using attention mechanism with dual tqwt and triple stft spectrogram, Physiological Measurement **42**, 105006 (2021).

[16] K. Kochetov, E. Putin, M. Balashov, A. Filchenkov, A. Shalyto, Noise masking recurrent neural network for respiratory sound classification, in *International Conference on Artificial Neural Networks* (Springer, 2018), pp. 208–217

[17] J. Acharya, A. Basu, Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning, IEEE transactions on biomedical circuits and systems **14**, 535 (2020).