Multiview Outlier Filtered Pediatric Heart Sound Classification

Sagnik Dakshit
Computer Science
The University of Texas at Tyler
Tyler, TX 75799

***Abstract*—The advancements in deep learning has generated a large-scale interest in development of black-box models for various use cases in different domains such as healthcare, in both at-home and critical setting for diagnosis and monitoring of various health conditions. The use of audio signals as a view for diagnosis is nascent and the success of deep learning models in ingesting multimedia data provides an opportunity for use as a diagnostic medium. For the widespread use of these decision support systems, it is prudent to develop high performing systems which require large quantities of data for training and low-cost method of data collection making it more accessible for developing regions of the world and general population. Data collected from low-cost collection especially wireless devices are prone to outliers and anomalies. The presence of outliers skews the hypothesis space of the model and leads to model drift on deployment. In this paper, we propose a multiview pipeline through interpretable outlier filtering on the small Mendeley Children Heart Sound dataset collected using wireless low-cost digital stethoscope. Our proposed pipeline explores and provides dimensionally reduced interpertable visualizations for functional understanding of the effect of various outlier filtering methods on deep learning model hypothesis space and fusion strategies for multiple views of heart sound data namely raw time-series signal and Mel Frequency Cepstrum Coefficients achieving 98.19% state-of-the-art testing accuracy.**

***Keywords*—Deep learning; outlier filtering; machine learning; ECG**

I. Introduction

Deep learning (DL), a subset of Artificial Intelligence (AI), has gained significant attention for its remarkable ability to analyze complex multimedia data, extracting meaningful patterns, and making predictions with unprecedented performance. In the context of healthcare informatics, deep learning is revolutionizing the way medical data is interpreted demonstrating remarkable success in a range of applications. In the ever-evolving landscape of healthcare informatics, audio signals have emerged as a valuable source of multimedia information that can contribute to enhanced health outcomes. Advancements in audio processing, coupled with the rise of artificial intelligence, have enabled healthcare professionals to extract meaningful insights from physiological audio sounds. The integration of audio signals into healthcare informatics showcases the versatility of data-driven technologies in improving patient care and holds the promise of more accurate diagnoses, personalized treatments, and innovative healthcare solutions.

One of the significant challenges that hinders the availability of large quantities of data required for training data hungry deep learning models is the cost of data collection, limiting the accessibility of deep learning based decision support systems to general public specially in developing regions of the world [14]. This has motivated the development of low-cost diagnostic devices such as wireless phone stethoscope [14], wireless OCT devices [25]. Data collection using these low-cost devices can be noisy and have out-of-distribution samples collectively termed as outliers. Outliers are data points that deviate significantly from the rest of the dataset and can distort statistical measures such as the mean and standard deviation, in turn affecting not only the predictive performance, generalizability and robustness but also skew the learned features by obscuring meaningful patterns. This shortcomings makes it prudent to filter out outliers from the training dataset.

The black-box nature of deep learning models exacerbates the above challenges and hinders the acceptance of automated decision support systems in critical healthcare tasks due to a lack of understanding of the effect of outliers on model learning and performance. In this work, we address the above research gap by using Uniform Manifold Approximation and Projection (UMAP) technique [24] to visualize the effect of various outlier filtering techniques on the learned deep learning model hypothesis space. UMAP provides interpretable insights in understanding how the removal of outliers affect model performance and also in identification of test outliers. Furthermore, in terms of acoustic heart sounds, to the best of the author's knowledge this is the first work leveraging multiple view of data for classification. Multiview data consists of different views or perspectives of the same data unlike multiview data which involved different types of data or entities. These views are usually derived from different sources, methodologies, or angles. Multiview learning involves leveraging these different views to improve the overall performance of the models by combining their outputs for more robust predictions. Parallel to multiview, We compare both early and late-fusion strategies using the raw time-series signals and Mel Frequency Cepstrum Coefficients (MFCC) as multiple views to achieve state-of-the-art performance on the small Mendeley Children Heart Sound dataset collected through low-cost wireless stethoscope.

A. Contributions

In this paper, we investigate the effect of popular outlier filtering methods through interpretable visualizations of learned model hypothesis spaces. We also investigate the feasibility of multiview early and late fusion strategies to achieve state-of-the-art binary classification on the Mendeley Children Heart Sound dataset collected from low-cost wireless stethoscopes.

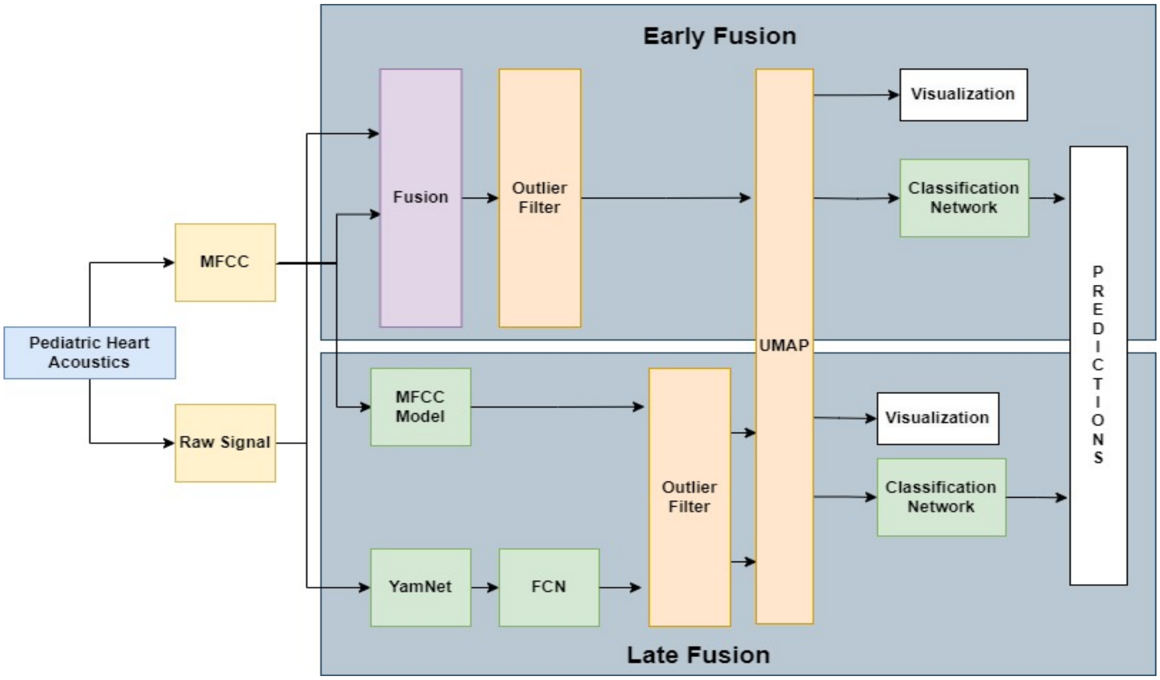

Fig. 1. Our proposed pipeline for outlier filtered MFCC and signal multiview integration for state of the art performance on Mendeley children heart sound dataset.

Our main contributions in this paper can be listed as follows:

- Our proposed pipeline achieve state-of-the-art accuracy for classification of normal and abnormal pediatric heart sounds through late-fusion of multiview outlier filtered Mendeley Children Heart Sound dataset.
- We investigate fusion strategies to improve classification accuracy of multiview pediatric heart sounds using a small number of samples.
- We provide interpretable visualization to understand the effect of outlier filtering on deep learning model hypothesis space.
- We provide a comparison of the effect on model performance of four popular outlier filtering strategies and their contamination hyperparameter.

The rest of the paper is organized as related works in Section II. We discuss our experimental methodology and proposed pipeline including our deep learning models, dataset and proposed pipeline in Section III. In Section IV, we discuss the results for both early and late fusion as well as various outlier filtering methods. Lastly in Section VI, we discuss the effect of outlier filtering on the hypothesis space of the best performing model through interpretable visualizations.

II. Related Works

Significant research progress has been made in developing decision support systems based on various machine learning and deep learning algorithms for heart disease diagnosis which poses a serious health concern for the general population. Electrocardiograms (ECG or EKG) [29], [22], [6], [23], [30], [8], Photoplethysmograph [27], [36], [9], [26], [13], and auscultations [34], [14], [31], [32] modalities have been used for automated decision making. Owing to large body of work on deep learning based heart disease classification, in this section we limit the scope of our discussion to the most relevant works to our approach of using UMAP and research on the Mendeley Children heart Sound dataset.

UMAP has been used extensively in all domains to reduce dimensionality for various purposes including interpretability, primarily for feature selection in context of ECG [28], [19], [15], [35] and heart sound classification [4], [5]. While various anomaly detection techniques have been used, there is limited use of Local Outlier Filtering (LOF) [37], [33], [16]. In our literature search, we found only a single work exploring anomaly detection with UMAP in an unsupervised approach unlike our supervised multiview approach [10] on non-healthcare acoustic scenes. Furthermore, none of the explored work provides a functional understanding of the effect of outlier removal on the deep learning model hypothesis space especially exploring different outlier filtering methods for comparison of both early and late-fusion of multiple views, making this the first novel investigation.

On the Mendeley Children heart Sound dataset, authors Islam et. al [14] in their work converted the signals into MFCC, engineered features and presented comparisons of various SVM kernels. Authors Rani et. al [31] denoised the dataset and removed noise artifacts followed by conversion to MFCC for Convolution Neural Network based classification. Our proposed approach uses MFCC as one of the views for deep learning models similar to the above approaches in conjugation with signals for fusion. Moreover, our approach performs outlier filtering and provides insight into how the outliers affect model performance and achieves state-of-the-art performance with significant improvement.

III. Experimental Methodology

A. Proposed Approach

Our developed pipeline as illustrated in Fig. 1 uses Mel Frequency Cepstrum Coefficients (MFCC) and raw time-series signals as multiple views of acoustic data for exploration of both early and late fusion strategies. The collection of coefficients that represent short-term log power spectrum of a signal through a linear cosine transform on a non-linear mel-frequency scale is referred to as MFCC. Our choice of modalities is inspired by the large body of work in literature which shows success in classification of audio signals with both modalities independently. For both fusion strategies, we investigate the effect of outlier filtering strategies on the learned model hypothesis space.

For late-fusion, we generate embeddings for each of the considered views which are passed through the outlier filtering module before fusion. The outlier samples can skew model learning so their removal can significantly improve model performance as demonstrated by Dakshit et. al [7]. While the embeddings for time-series auscultations are generated through pre-trained YamNet model, MFCC embeddings are generated using our deep learning MFCC model as illustrated in Section III-D. Following outlier filtering, the embeddings of the remaining samples are fused and passed through our late-fusion neural network architecture for classification. The hypothesis space of this trained model is investigated for functional understanding of the effect of outlier filtering through interpretable UMAP visualization of the learned hypothesis space. For comparative understanding, we also present visualizations of the hypothesis space of the model trained without passing the multiview late-fused data through the outlier filtering module.

For early-fusion, we concatenate the MFCC two-dimensional data view and one-dimensional raw time-series signal view which is then passed through outlier filtering modules and used to train the classification deep learning model as illustrated in Section III-D. The models trained by passing the training data views through the outlier filtering module as well as a control experiment without passing through the outlier filtering modules are visualized by UMAP for interpertable functional understanding of the effect of the outliers on the model hypothesis space.

B. Interpretable Outlier Filtering

In this section, we discuss our strategy for interpretable outlier filtering from training samples. We use UMAP to reduce dimensions for interpretable outlier filtering. UMAP (Uniform Manifold Approximation and Projection) [24], is a dimensionality reduction technique widely used by high-dimensional data to provide a more effective visualization of complex datasets. Unlike comparative methods such as t-SNE (t-distributed stochastic neighbor embedding) [21], UMAP offers scalability and preserves both global and local structures within the data. UMAP operates by mapping data points from a high-dimensional space to a lower-dimensional one, making it easier to visualize and interpret patterns. However, it is a non-deterministic algorithm leading to slight variation in results with the same data and the same parameters each time, primarily due to random initiation and stochastic optimization. We project each embedding in 2D using UMAP for Local Outlier Filtering (LOF). LOF provides interpretability in terms of probabilities. The interpretations and explanations of these methods do not provide or allow functional understanding on how the selection and filtering of outliers affect the deep learning model's learned hypothesis space as discussed earlier in Section II. We address this challenge using UMAP in this paper and demonstrate the effect of outlier removal on the children heart sounds dataset.

C. Dataset

We select the Mendeley Children heart sound dataset of normal and abnormal pediatric heart sounds from the rural areas of Bangladesh [14] collected from 60 subjects using their developed wireless electronic stethoscope. The collected dataset of 1657 samples, is preprocessed by re-sampling to 44100 Hz, normalized and denoised using Discrete Wavelet Transform. This dataset poses challenges in terms of its small number of samples to train deep learning models and quality of collection with low-cost wireless devices. Furthermore, the data has been recently published and not adequately tested in research but provides the opportunity to investigate the feasibility of developing high-performing deep learning models on wireless and cost-effective data collection devices for developing sections of the world.

D. Deep Learning Architectures

In this section, we illustrate our three deep networks for early and late-fusion of multiview learning for subtasks of 1) Raw Signal Embeddings, 2) MFCC Embedding, and 3) Fusion Classification Network and 4) Early-fusion strategy .

- MFCC Model: Our network used to generate MFCC embeddings is shown in Fig. 2. Our architecture has two 2D convolution layers with ReLu activation (64 and 32 filters of size $3 * 3$), each followed by Batch Normalization with an interwined Max Pooling layer and batch normalization layer. These layers are followed by a Global Average Pooling layer and a fully connected layer of 128 dense nodes and a sigmoid classification layer.
- Raw Signal Embeddings Model: YAMNet is a lightweight deep network trained on AudioSet data by Google to classify on devices with limited computational resources. It is trained on a large dataset containing thousands of different sound events, enabling it to identify and categorize various acoustic patterns. We leverage the pre-trained YAMNet as our encoder backbone which generates embeddings of length 1024. We fine-tuned the pre-trained YamNet for generating pediatric heart sound embeddings using two Fully Connected Layers (FCN) as shown in Fig. 1. Our FCN has dense layers of 256 and 128 nodes and reduces the embedding dimensions to match MFCC and signal embedding dimensions.
- Fusion Classification Model: For our fusion classification network, we use three dense layers with 512, 128 and 64 nodes with ReLu activation and followed by a sigmoid layer for binary classification. All models are

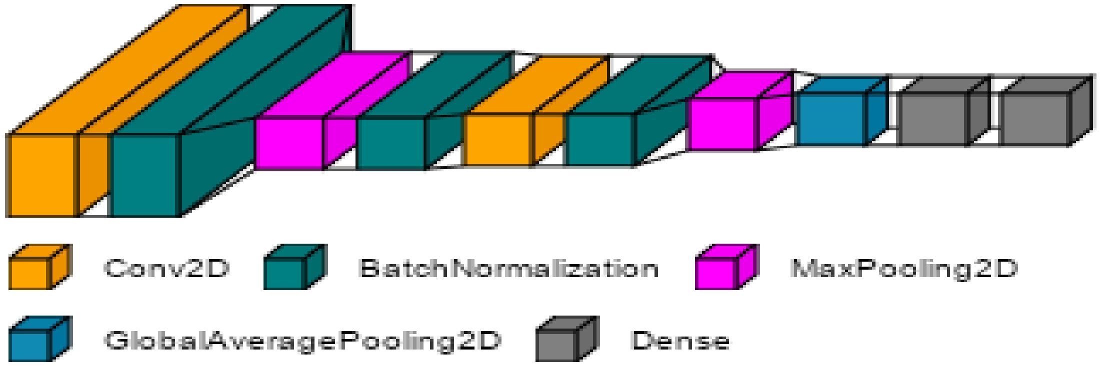

Fig. 2. MFCC Model deep learning architecture.

trained with a learning rate 0.001 and a decay rate of 0.0001 and binary crossentropy loss.

- Early-Fusion Model: We tried various architectures for early fusion based on 2D convolutional layers but most architectures lead to underfitting or extensive overfitting. We report results on EfficientNet B7 [17] which obtained the best results as reported in Section IV.

IV. Experimental Results

In this section, we discuss the observed results in terms of performance of baseline models for both view and improvement using outlier filtered fusion to achieve state-of-the-art test performance. We trained DL models for 100 epochs each with hyperparameters as discussed in Section III-D and save the best model in terms of validation accuracy.

A. Baseline Results

In this section, we discuss the observed results in terms of performance of baseline models for both views and improvement through outlier filtering We trained deep learning models for 100 epochs each with hyperparameters as discussed in Section III-D and save the best model in terms of validation accuracy. For outlier filtering we present only the results for Local Outlier Filtering in this section as LOF is more suited towards this task and data. One-Class SVM, Isolation Forest, and Elliptical Envelope outlier filtering methods are more suitable for datasets where outliers are well-separated from the majority of the data and where outliers are in sparsely populated regions of the feature space. Given the task of binary classification, our feature space is not expected to have sparsely populated regions. The effectiveness of LOF in detecting outliers where the density of the data points varies across different regions makes it a better theoretical choice. We empirically show the results for all four considered outlier filtering methods on fusion in section V-A.

In Table I, we report the results of the best baseline models in terms of traditional deep learning metrics of Precision, Recall, AUC, and Accuracy. From these baseline model results, it can be observed that both modalities achieve high performance on the small dataset, with MFCC having superior performance in terms of all evaluated traditional deep learning

TABLE I. Baseline Model and Local Outlier Filtered Deep Learning Model Test Performance on MFCC and Signal Views

Model	*Test Accuracy*	*Precision*	*Recall*	*AUC*
Signal	0.90	0.884	0.897	0.96
MFCC	0.96	0.97	0.95	0.99
OF-Signal	0.928	0.92	0.96	0.924
OF-MFCC	0.976	0.974	0.994	0.987

metrics. Moreover, it should be noted that the raw signal view leverages pre-trained YamNet model while MFCC is trained from scratch. We do not report the results for training raw signal view from scratch as comparable accuracy could not be achieved even with state-of-the-art architectures. To demonstrate the effect of outlier filtering, we retrain the baseline architectures without changing the hyperparameters on the outlier filtered training data and these models are represented with the prefix OF in Table I. We record a significant improvement of 2.8%, 8.6%, 5.3%, and 3% in test accuracy, precision, recall, and AUC respectively for raw signal view. Similar observations are recorded for MFCC view with a 1.6%, 0.4%, 4.7% improvement in test accuracy, precision, and recall respectively. After outlier filtering it is observed that MFCC holds its superior performance in classifying pediatric heart sounds.

B. Fusion Results

TABLE II. Local Outlier Filtered Early and Late Fusion

Model	*Test Accuracy*	*Precision*	*Recall*	*AUC*
OF-Late Fusion	0.9819	0.984	0.996	0.987
OF-Early Fusion	0.5693	0.5696	0.9626	0.5273

Following our proposed pipeline as shown in Fig. 1, for both early and late-fusion, we report our observed results in Table II. It can be observed that the late-fusion of the generated embeddings of the two views of MFCC and signal significantly outperforms the early-fusion approach. As recorded in Table II, Late-Fusion of the outlier filtered embeddings of both views yields an improvement of 0.59% in test accuracy and 1% for precision and 0.2% recall. The fusion strategy yields significantly better results over both the individual view models. The above is observed for both with and without

TABLE III. CONTAMINATION FACTOR HYPERPARAMETER STUDY FOR LOCAL OUTLIER FILTERING ON LATE-FUSION DEEP LEARNING MODEL TEST PERFORMANCE

Contamination Factor	**Training Sample Count**	***Test Accuracy***	***Precision***	***Recall***	***AUC***
0.01	1311	0.9819	0.984	0.996	0.99
0.05	1258	0.9789	0.9737	0.9893	0.984
0.10	1192	0.9789	0.9737	0.9893	0.9796
0.50	663	0.9729	0.9734	0.9786	0.9934

TABLE IV. CONTAMINATION FACTOR HYPERPARAMETER STUDY FOR ISOLATION FOREST FILTERING ON LATE-FUSION DEEP LEARNING MODEL TEST PERFORMANCE

Contamination Factor	**Training Sample Count**	***Test Accuracy***	***Precision***	***Recall***	***AUC***
0.01	1311	0.9819	0.9738	0.995	0.99
0.05	1258	0.9367	0.9029	0.99	0.9863
0.10	1192	0.9337	0.8986	0.9947	0.9413
0.50	663	0.9729	0.9684	0.9840	0.9743

TABLE V. CONTAMINATION FACTOR HYPERPARAMETER STUDY FOR COVARIANCE ELLIPTIC ENVELOPE FILTERING ON LATE-FUSION DEEP LEARNING MODEL TEST PERFORMANCE

Contamination Factor	**Training Sample Count**	***Test Accuracy***	***Precision***	***Recall***	***AUC***
0.01	1311	0.9819	0.9738	0.995	0.99
0.05	1258	0.9819	0.9738	0.99	0.9897
0.10	1192	0.9819	0.9738	0.9947	0.9922
0.50	663	0.9518	0.9340	0.9840	0.9569

TABLE VI. CONTAMINATION FACTOR HYPERPARAMETER STUDY FOR ONE-CLASS-SVM FILTERING ON LATE-FUSION DEEP LEARNING MODEL TEST PERFORMANCE

Contamination Factor	**Training Sample Count**	***Test Accuracy***	***Precision***	***Recall***	***AUC***
0.01	1310	0.9819	0.9738	0.995	0.98
0.05	1256	0.9819	0.9738	0.99	0.9796
0.10	1192	0.9819	0.9738	0.9947	0.98
0.50	663	0.9581	0.9340	0.9840	0.9887
0.75	331	0.9488	0.9620	0.9465	0.9645
1.0	150	0.9398	0.9372	0.9572	0.9867

outlier filtering achieving state-of-the-art classification performance. The observed superiority of late-fusion performance can be primarily attributed to the embeddings having learned meaningful representations of the multiple data views, which leads to more efficient fusion of features and consequently better classification outcomes.

V. COMPARATIVE DISCUSSION

A. Comparison of Outlier Filtering Methods

We investigate the feasibility of outlier filtering with four popular inherently non-interpretable methods namely Local Outlier Filtering, Isolation Forest, Elliptical Envelope, One-class SVM.

- Local Outlier Filtering (LOF) [3]: It is a data-driven approach to identifying outliers in a dataset by assessing the local neighborhood of each data point. This method operates on the premise that outliers are data points that deviate significantly from their local surroundings. By comparing the distance or density of a point to its nearest neighbors, the local outlier filtering method can effectively detect data points that are inconsistent with the patterns observed in their immediate vicinity. LOF works in 3 steps namely local density estimation followed by comparison to neighbors and lastly outlier detection. Points with LOF scores significantly higher than 1 are considered outliers.

- Isolation Forest [20]: It identifies anomalies by constructing decision trees to isolate individual data points. Unlike LOF that measures the distance or density of data points, Isolation Forest focuses on how quickly a data point can be isolated from the rest of the data. The approach involves randomly selecting a feature and a split value, then recursively partitioning the data into two subsets based on this split. Anomalous data points are expected to be isolated in fewer partitions (or trees) because they are different from the majority of the data.

- Elliptical Envelope [2]: The Elliptic Envelope method is a statistical approach to outlier detection that models the data distribution using a multivariate Gaussian (normal) distribution and identifies outliers as data points that deviate significantly from this distribution. By fitting an elliptical envelope around the data points, this method estimates the mean and covariance of the data and calculates the Mahalanobis distance for each point. Data points that fall outside a certain threshold of the distance metric are classified as outliers.

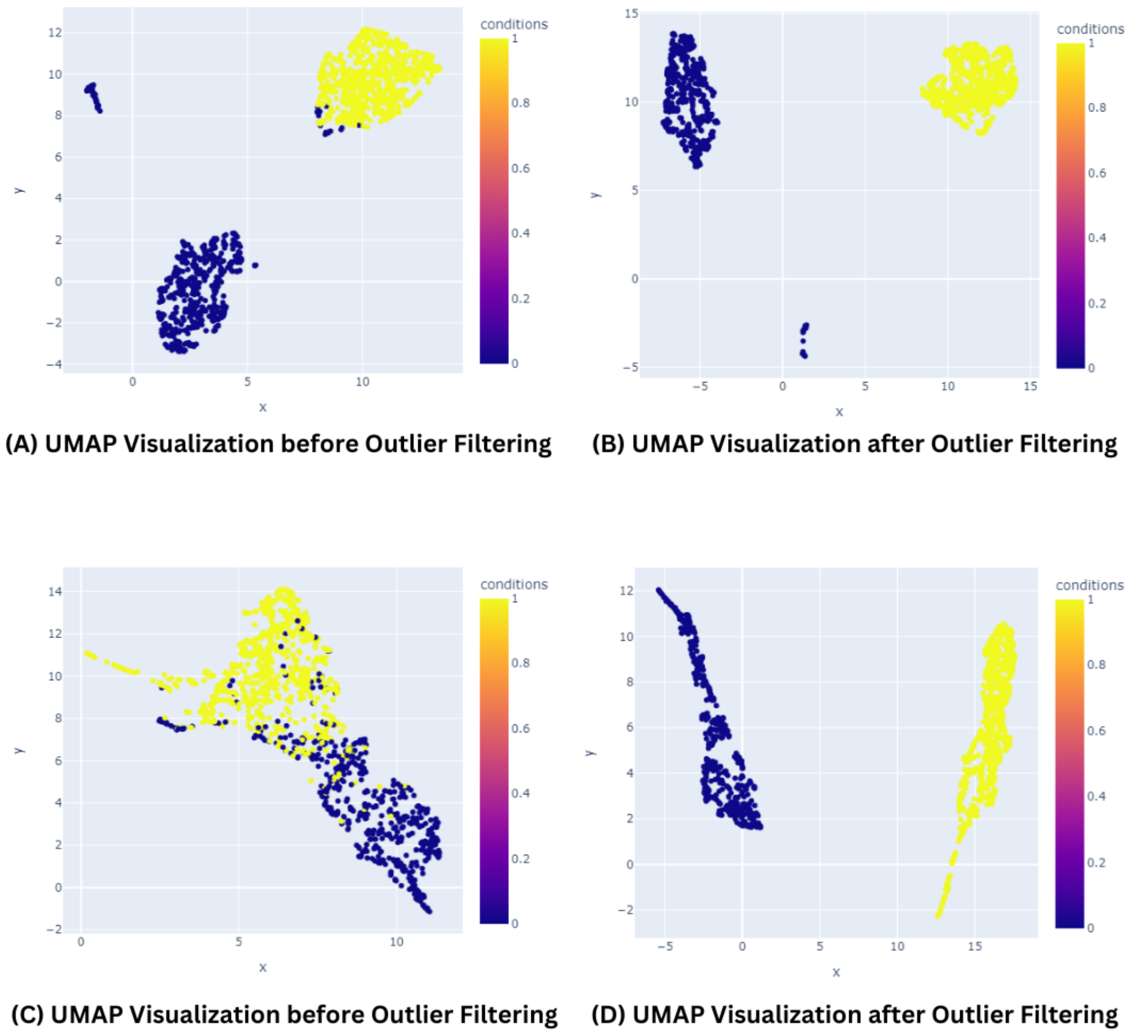

Fig. 3. Local outlier filtering: UMAP Visualization of hypothesis space before (LEFT) and after filtering with 0.01 contamination (RIGHT) outlier filtering for MFCC (TOP) and raw signal (BOTTOM) views.

- One-Class Support Vector Machines (One-Class SVM) [1]: It is a machine learning-based approach for outlier detection that aims to identify anomalies by constructing a decision boundary that encapsulates the normal data points in a dataset. This method operates in a similar way to traditional support vector machines but is adapted for unsupervised learning and outlier detection. By training on a dataset consisting predominantly of normal data distribution, One-Class SVM learns a hyperplane that separates the data from the origin. Points that fall within this boundary are considered normal, while points outside the boundary are deemed outliers.

These methods provide robust outlier detection capabilities to handle large datasets of high-dimensionality, each with its own advantages and disadvantages. Owing to the large superiority of late-fusion approach, in Tables III, IV, V, and VI, we report the results for each of the considered outlier filtering methods including range of values for the contamination hyperparameter for only late-fusion strategy. We observe from the presented results that as the contamination value is increased, there is a significant drop in model performances. For all the compared methods, the best performance is observed for the lowest contamination value of 0.01, with LOF method as the best performing model globally with comparatively similar performances for the other outlier filtering methods.

B. Comparison on Same Dataset

The nascent nature of the dataset reduces the possibility of comparison with existing works. Authors Islam et. al in their work proposed the dataset [14] and achieved 94.12% test accuracy, 88.89% specificity, and 100% sensitivity values using RBF SVM kernel on engineered MFCC features. In [31] 93.76% test accuracy was achieved using Convolution

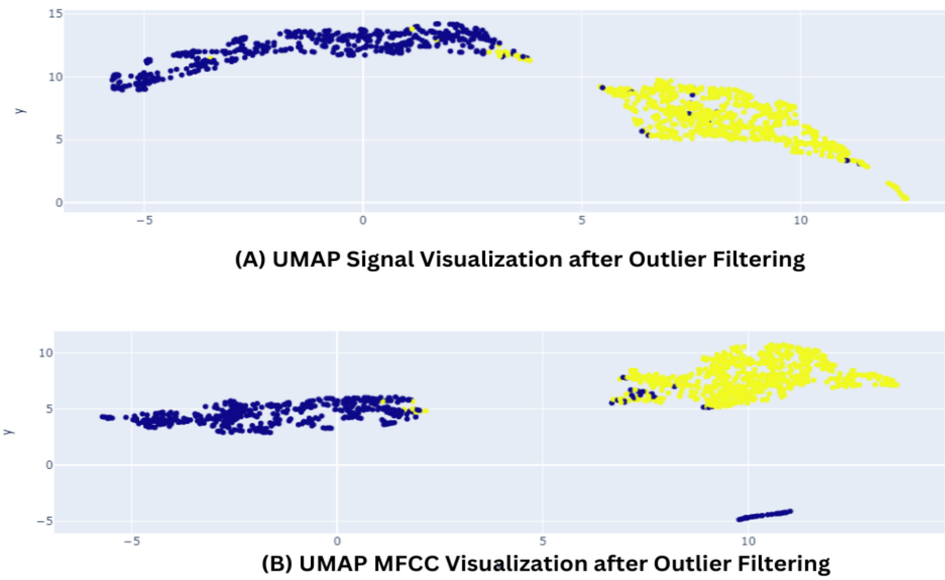

Fig. 4. One-Class-SVM Outlier filtering: UMAP visualization of hypothesis space after with filtering with 0.01 contamination for MFCC (BOTTOM) and raw signal(TOP) views.

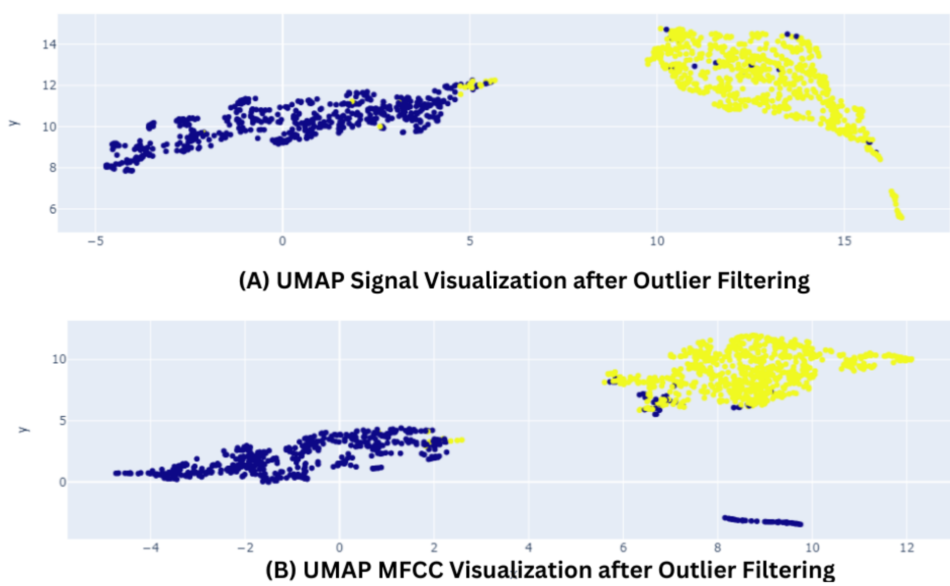

Fig. 5. Elliptical envelope outlier filtering: UMAP visualization of hypothesis space after with filtering with 0.01 contamination for MFCC (BOTTOM) and raw signal(TOP) views.

Neural Network using MFCC. The authors in this approach strategically removed artifacts and denoised the data as pre-processing steps. The comparison demonstrated that not only is our proposed approach able to achieve superior performance over existing approaches in terms of testing accuracy but also maintains comparable specificity and sensitivity of 99.6% and 99.1% while being an efficient way of removing and visualizing outliers.

VI. Interpretable Visualization

In this section, we discuss the effect of outliers through interpretable visualization. Neural Networks learn high dimensional embeddings from multiview data. Interpretability in the field of Explainable AI (XAI) is defined loosely as understanding what the model did or could have done [12]. Visualization methods for high dimensional embeddings is one of the established ways for providing interpretability [18], [11]. We use UMAP to reduce dimensionality of the sample embeddings allowing their visualization in two dimensions. Our dimensionally reduced 2D interpretable visualizations of the training dataset for both raw time-series signal and MFCC views are presented in Fig. 3, 4, 5, and 6.

In Fig. VI, the images on the top are for MFCC view and bottom for signal view with the left images representing the set before outlier filtering and right representing after outlier filtering. Our hyperparameter of contamination for LOF

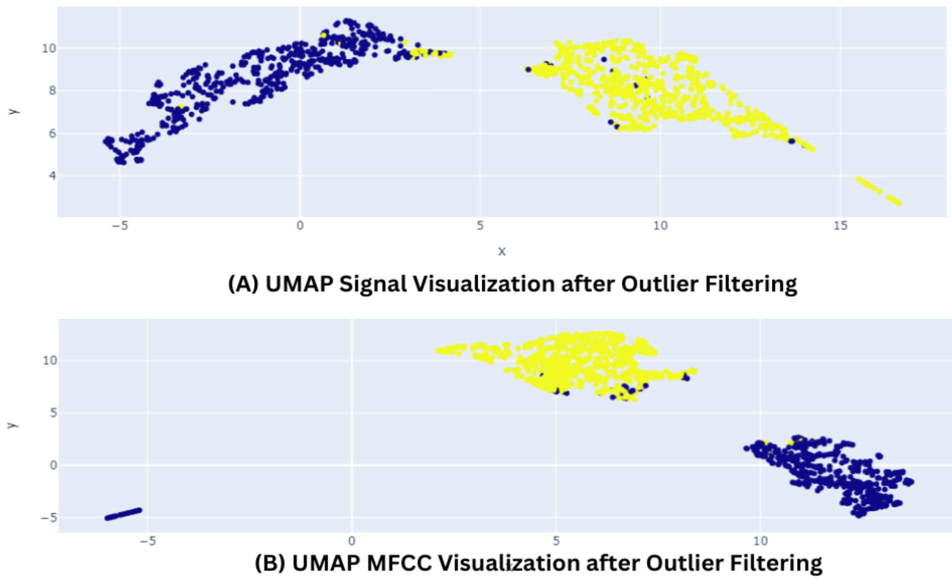

Fig. 6. Isolation forest outlier filtering: UMAP visualization of hypothesis space after with filtering with 0.01 contamination for MFCC (BOTTOM) and raw signal(TOP) views.

detection was set to 0.01 based on grid search which is reported in Section V-A. We can observe from the images on the left for both views of MFCC and signal, that the embeddings are clustered naturally into two groups with some overlap between classes for signal view. On outlier filtering with LOF, we observe significantly improved clusters without any intersection. As observed from the images on the right, our method allows removal of any potential outliers that could skew model learning leading to model drift in real world. The difference in position of the classes in the UMAP interpretable visualization before and after outlier filtering is primarily due to the stochastic and random initiation nature of the algorithm as discussed above and does not have any significance on the detection of outliers or contamination. The interpretable visualizations coupled with the model testing results reported in Table I and Table II explains the difference in performance. Visualization of the two views for outlier filtering methods of Isolation Forest (Fig. 6), Elliptical Envelope (Fig. 5), and One-Class-SVM (Fig. 4) with the best contamination hyper-parameter values are presented. It can be observed from the visualizations that the One-Class SVM, Elliptical Envelope, and Isolation Forest methods have a greater number of outliers present for both views in comparison to Local outlier Filtering (Fig. 3) explaining not only the better performance but also the superiority of LOF over the other methods for our case. These interpretable visualizations allow us to understand which samples have been removed based on the position of the samples in the embedding space.

VII. Conclusion

In healthcare, there has been nascent interest in the one-dimensional modality of auscultation, which represent sounds from physiological functions for diagnosis and monitoring of various health conditions. Despite the success of deep learning, the cost of quality data collection and at-home monitoring devices makes it an accessibility challenge for the developing parts of the world. Outliers or anomalies are frequently present in data collected using these low-cost devices which can skew model learning and consequently lead to model drift on deployment. In this paper, we developed a pipeline to filter outliers that have a derogatory effect on model performance and achieve 98.19% state-of-the-art testing accuracy through multiview fusion on the public Mendeley Children Heart Sound dataset collected through wireless low-cost stethoscope. To the best of the author's knowledge, this is the first work on the feasibility of both early and late-fusion approached for multiview heart sounds with late-fusion on generated view embeddings, demonstrating significantly better results. Our approach also investigated the effect of outliers on deep learning model hypothesis space through interpretable visualizations for a functional understanding. Outlier Filtering using reduced dimensions by UMAP not only allows for superior performance but also interpretable visualization of the effect of outliers on model performance. We compared the effect of four popular outlier filtering methods on the model hypothesis space demonstrating the importance of selection of appropriate method and interpretable functional understanding of the same. As future work, we would investigate other modalities and views with larger datasets to understand generalizability of outlier filtering.

References

[1] Amer, M., Goldstein, M., Abdennadher, S.: Enhancing one-class support vector machines for unsupervised anomaly detection. In: Proceedings of the ACM SIGKDD workshop on outlier detection and description. pp. 8–15 (2013)

[2] Ashrafuzzaman, M., Das, S., Jillepalli, A.A., Chakhchoukh, Y., Sheldon, F.T.: Elliptic envelope based detection of stealthy false data injection attacks in smart grid control systems. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI). pp. 1131–1137. IEEE (2020)

[3] Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: Lof: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data. pp. 93–104 (2000)

[4] Chen, X.J., Collins, L.M., Patel, P.A., Karra, R., Mainsah, B.O.: Heart sound analysis in individuals supported with left ventricular assist device: A first look. In: 2020 Computing in Cardiology. pp. 1–4. IEEE (2020)

[5] Chen, X.J., LaPorte, E.T., Olsen, C., Collins, L.M., Patel, P., Karra, R., Mainsah, B.O.: Heart sound analysis in individuals supported with left ventricular assist devices. IEEE Transactions on Biomedical Engineering **68**(10), 3009–3018 (2021)

[6] Dakshit, S., Maweu, B.M., Dakshit, S., Prabhakaran, B.: Core-set selection using metrics-based explanations (csume) for multiclass ecg. In: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI). pp. 217–225. IEEE (2022)

[7] Dakshit, S., Maweu, B.M., Dakshit, S., Prabhakaran, B.: Core-set selection using metrics-based explanations (csume) for multiclass ecg. In: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI). pp. 217–225 (2022). https://doi.org/10.1109/ICHI54592.2022.00041

[8] Dakshit, S., Prabhakaran, B.: Twelve lead double stacked generalization for ecg classification. In: 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI). pp. 245–251. IEEE (2023)

[9] El-Dahshan, E.S.A., Bassiouni, M.M., Khare, S.K., Tan, R.S., Acharya, U.R.: Exhyptnet: An explainable diagnosis of hypertension using efficientnet with ppg signals. Expert Systems with Applications **239**, 122388 (2024)

[10] Fernandez, A., Plumbley, M.D.: Using umap to inspect audio data for unsupervised anomaly detection under domain-shift conditions. arXiv preprint arXiv:2107.10880 (2021)

[11] Freris, N.M., Ajalloeian, A., Vlachos, M.: Interpretable embedding and visualization of compressed data. ACM Transactions on Knowledge Discovery from Data **17**(2), 1–22 (2023)

[12] Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L.: Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA). pp. 80–89. IEEE (2018)

[13] Hettiarachchi, C., Chitraranjan, C.: A machine learning approach to predict diabetes using short recorded photoplethysmography and physiological characteristics. In: Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17. pp. 322–327. Springer (2019)

[14] Islam, M.R., Hassan, M.M., Raihan, M., Datto, S.K., Shahriar, A., More, A.: A wireless electronic stethoscope to classify children heart sound abnormalities. In: 2019 22nd International Conference on Computer and Information Technology (ICCIT). pp. 1–6 (2019). https://doi.org/10.1109/ICCIT48885.2019.9038406

[15] Jain, R., Betrabet, P.R., Rao, B.A., Reddy, N.S.: Classification of cardiac arrhythmia using improved feature selection methods and ensemble classifiers. In: Journal of Physics: Conference Series. vol. 2161, p. 012003. IOP Publishing (2022)

[16] Karasmanoglou, A., Antonakakis, M., Zervakis, M.: Ecg-based semi-supervised anomaly detection for early detection and monitoring of epileptic seizures. International Journal of Environmental Research and Public Health **20**(6), 5000 (2023)

[17] Koonce, B., Koonce, B.: Efficientnet. Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization pp. 109–123 (2021)

[18] Lal, V., Ma, A., Aflalo, E., Howard, P., Simoes, A., Korat, D., Pereg, O., Singer, G., Wasserblat, M.: Interpret: An interactive visualization tool for interpreting transformers. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations. pp. 135–142 (2021)

[19] Li, M., Si, Y., Yang, W., Yu, Y.: Et-umap integration feature for ecg biometrics using stacking. Biomedical Signal Processing and Control **71**, 103159 (2022)

[20] Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data (TKDD) **6**(1), 1–39 (2012)

[21] Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)

[22] Mathews, S.M., Kambhamettu, C., Barner, K.E.: A novel application of deep learning for single-lead ecg classification. Computers in biology and medicine **99**, 53–62 (2018)

[23] Maweu, B.M., Dakshit, S., Shamsuddin, R., Prabhakaran, B.: Cefes: a cnn explainable framework for ecg signals. Artificial Intelligence in Medicine **115**, 102059 (2021)

[24] McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)

[25] Mehta, R., Nankivil, D., Zielinski, D.J., Waterman, G., Keller, B., Limkakeng, A.T., Kopper, R., Izatt, J.A., Kuo, A.N.: Wireless, web-based interactive control of optical coherence tomography with mobile devices. Translational Vision Science & Technology **6**(1), 5–5 (2017)

[26] Neha, Sardana, H., Kanwade, R., Tewary, S.: Arrhythmia detection and classification using ecg and ppg techniques: A review. Physical and Engineering Sciences in Medicine **44**(4), 1027–1048 (2021)

[27] Pachori, D., Tripathy, R.K., Jain, T.K.: Detection of atrial fibrillation from ppg sensor data using variational mode decomposition. IEEE Sensors Letters (2024)

[28] Pan, X.: Deep Transfer Learning Applied to Time-series Classification for Predicting Heart Failure Worsening Using Electrocardiography. Ph.D. thesis, Worcester Polytechnic Institute (2020)

[29] Pyakillya, B., Kazachenko, N., Mikhailovsky, N.: Deep learning for ecg classification. In: Journal of physics: conference series. vol. 913, p. 012004. IOP Publishing (2017)

[30] Qiu, C., Li, H., Qi, C., Li, B.: Enhancing ecg classification with continuous wavelet transform and multi-branch transformer. Heliyon (2024)

[31] Rani, S., Dutta, M.K.: Heart anomaly classification using convolutional neural network. In: Proceedings of International Conference on Data Science and Applications: ICDSA 2021, Volume 1. pp. 541–550. Springer (2022)

[32] Ren, Z., Cummins, N., Pandit, V., Han, J., Qian, K., Schuller, B.: Learning image-based representations for heart sound classification. In: Proceedings of the 2018 international conference on digital health. pp. 143–147 (2018)

[33] Shan, L., Li, Y., Jiang, H., Yu, H., Chang, S.: Abnormal ecg detection based on an adversarial autoencoder. Frontiers in Physiology **13**, 961724 (2022)

[34] Tschannen, M., Kramer, T., Marti, G., Heinzmann, M., Wiatowski, T.: Heart sound classification using deep structured features. In: 2016 Computing in Cardiology Conference (CinC). pp. 565–568. IEEE (2016)

[35] Vadillo-Valderrama, A., Chaquet-Ulldemolins, J., Goya-Esteban, R., Caulier-Cisterna, R., Sánchez-Muñoz, J.J., García-Alberola, A., Rojo-Alvarez, J.L.: Exploring cardiac rhythms and improving ecg beat classification through latent spaces. IEEE Access (2024)

[36] Wu, Y., Tang, Q., Zhan, W., Li, S., Chen, Z.: Res-bianet: A hybrid deep learning model for arrhythmia detection based on ppg signal. Electronics **13**(3), 665 (2024)

[37] Yuan, J., Fang, S., Huang, X., Cao, N.: Ecglens: Interactive ecg classification and exploration