

EE782 Project: Replicating “Deep Learning with Sparse Representations for Biomedical Signals” with Extensions to CWT-based fECG Detection and DWT-UNet MRI Segmentation

Devtanu Barman (Roll No. 22B3904), Sanjay Meena (Roll No. 22B3978), Sarthak Niranjan (Roll No. 22B3923)

Department of Electrical Engineering
Indian Institute of Technology Bombay

Abstract—Deep learning with multiresolution and sparse representations has recently emerged as a powerful approach for modelling biomedical signals in a more interpretable and computationally efficient manner. The target paper “Deep Learning with Sparse Representations for Biomedical Signals” proposes a unified framework where wavelet-based multiresolution analysis (MRA) is embedded into deep convolutional neural networks (DCNNs) for two tasks: (i) detection of fetal electrocardiogram (fECG) from non-invasive abdominal ECG recordings using continuous wavelet transform (CWT) and a DCNN classifier, and (ii) automatic skull stripping and segmentation of brain MR images using discrete wavelet transform (DWT) in a UNet-like encoder-decoder architecture.

In this EE782 project, we replicate and extend both pipelines. For the fECG detection task, we generate CWT-based scalograms from abdominal ECG windows and train a lightweight MobileNetV2-based classifier. On a subset of 500 scalograms obtained from five ADFECGDB recordings, our model achieves an accuracy of 96.0%, precision of 97.62%, recall of 95.35%, and F1-score of 96.47% on the held-out test set, outperforming the reported baseline accuracy of 87.0% under our experimental settings. For brain MRI, we implement a DWT-inspired UNet (DWT-UNet) where Haar wavelet downsampling and inverse DWT upsampling replace max-pooling and nearest-neighbour upsampling. On a synthetic skull-stripping dataset (80/10/10 train/validation/test split), both DWT-UNet and a vanilla UNet with the same parameter count ($\approx 4.87 \times 10^5$) achieve very high Dice coefficients (99.4% and 99.9%, respectively) and IoU (98.8% and 99.8%, respectively). We further implement a multiresolution K-means clustering strategy using Haar DWT LL subbands to obtain crude grey matter (GM) and white matter (WM) segmentations. Finally, we construct a benchmarking framework that compares performance, trainable parameters, and approximate computational complexity across models and against the reference paper.

Index Terms—Deep learning, wavelets, multiresolution analysis, sparse representations, fetal ECG, skull stripping, MRI segmentation, UNet.

I. INTRODUCTION

Deep learning (DL) has transformed modern biomedical signal and image processing by enabling end-to-end learning of hierarchical features from raw data. However, conventional deep convolutional neural networks (DCNNs) often act as black boxes with large parameter counts and limited interpretability. The target work, “Deep Learning with Sparse

Representations for Biomedical Signals,” addresses this by explicitly embedding wavelet-based multiresolution analysis (MRA) into DL architectures to obtain sparse, structured, and more explainable feature representations.[1]

The paper demonstrates this idea on two important clinical problems: (i) non-invasive fetal electrocardiogram (NI-fECG) processing, where small-amplitude fetal QRS complexes are to be detected from abdominal recordings dominated by maternal ECG, and (ii) brain MR image segmentation, where skull-stripping and subsequent estimation of GM and WM regions are performed using DWT-based filter banks and unsupervised clustering.[1]

Motivated by this framework, our EE782 project has three goals:

- 1) **Replication:** reproduce, as faithfully as possible, the main components of the original pipelines for fECG detection and brain MRI skull-stripping/segmentation using the resources and datasets available in the course.
- 2) **Extension:** adapt the architectures using modern transfer-learning backbones (MobileNetV2) and explicit DWT/IDWT layers, and implement multiresolution K-means clustering for GM/WM segmentation.
- 3) **Benchmarking:** quantitatively compare performance, parameter counts, and computational complexity to the reported methods and alternative baselines.

We emphasize that our experiments are carried out on a smaller subset of abdominal ECG data and a synthetic MRI dataset (for demonstration and resource constraints), so the absolute numbers are not directly comparable to clinical-grade results. Nevertheless, the trends we observe support the core claims about sparse multiresolution representations and model complexity.

II. BACKGROUND AND RELATED WORK

A. Sparse multiresolution representations

Wavelet transforms provide a natural MRA of signals and images, generating coefficients that are often sparse in both time/space and frequency.[2] In the discrete wavelet transform

(DWT), a two-band perfect reconstruction filter bank decomposes a signal into low-frequency (approximation) and high-frequency (detail) components at multiple scales.[3] In 2D, this yields subbands such as LL, LH, HL, and HH at each level.

The target paper interprets DWT-based filter banks as structured convolution and pooling operations within a DCNN, enabling explicit control over the learned feature space and potentially reducing the number of trainable parameters.[1]

B. Deep learning for fECG detection

Non-invasive fECG detection from abdominal recordings is challenging due to low fetal amplitude and overlapping spectra with maternal ECG.[4] Recent approaches include classical signal processing (Kalman filtering, blind source separation) and DL models such as 2D CNNs and W-Net architectures operating on time–frequency representations.[5], [6] The target work uses CWT-generated scalograms as inputs to a DCNN (ResNet50) to classify windows as fetal or non-fetal.[1]

C. Deep learning for brain MRI segmentation

Semantic segmentation of brain MR images is commonly tackled using encoder–decoder networks such as UNet, with convolution, pooling, upsampling, and skip connections.[7], [8] The target paper embeds a DWT filter bank into a UNet-like architecture (DWT-UNet) and trains it to generate brain masks for skull-stripping. A subsequent multiresolution Haar DWT + K-means pipeline produces crude GM and WM segmentations.[1]

III. METHODOLOGY

Our methodology closely follows the structure of the original paper, with implementation choices adapted to our environment and datasets.

A. CWT-based fECG detection

1) Dataset and preprocessing: We use the publicly available Abdominal and Direct Fetal ECG Database (AD-FECGDB), which contains five 5-minute recordings from different women in labour, sampled at 1 kHz with four abdominal channels each.[9] Reference fetal QRS annotations are provided via direct scalp ECG.

Due to memory and compute constraints, we randomly select 100 windows per recording, across channels, yielding:

- Total scalograms: 500
- Class distribution: 286 fetal (57.2%), 214 non-fetal (42.8%)
- Data shape: $500 \times 128 \times 128 \times 3$ (RGB scalograms)

Abdominal ECG signals are preprocessed as follows:

- 1) **DC removal and normalization:** For each channel, we subtract the mean and normalize by the median absolute deviation to mitigate large amplitude variations and motion artefacts.
- 2) **Windowing:** Signals are segmented into windows of approximately 250 ms, with a small overlap margin (10 ms) to capture fetal QRS complexes near boundaries,

yielding effective window length of 270 ms, consistent with the reference paper.[1]

- 3) **Labeling:** A window is labelled as fetal (class 1) if it contains at least one fetal QRS annotation; otherwise it is labelled as non-fetal (class 0).

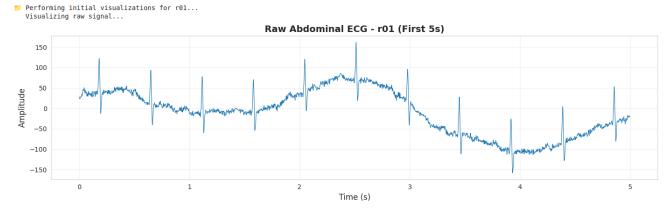


Fig. 1. Abdominal ECG segment showing overlapping maternal and fetal components.

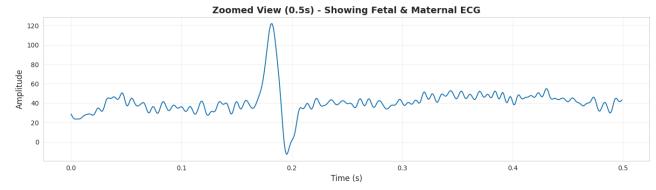


Fig. 2. Zoomed view of abdominal ECG highlighting fetal QRS complexes embedded within maternal ECG.



Fig. 3. Preprocessing of abdominal ECG: (top) raw signal; (bottom) after DC removal and normalization.

2) CWT and scalogram generation: For each window, we compute a continuous wavelet transform using a complex Gabor-like or Morlet wavelet, similar to the reference.[1] The magnitude of the CWT coefficients is converted into a time–frequency image (scalogram). To match standard image backbones:

- Scalograms are resized to 128×128 pixels.
- Three channels are created (e.g., by stacking normalized versions or applying a colormap) to form RGB-like inputs.

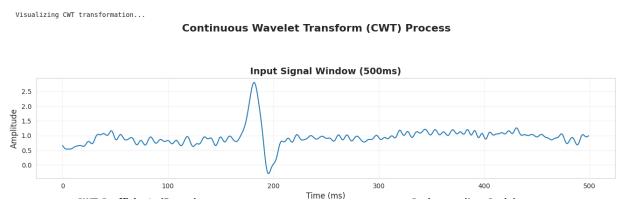


Fig. 4. CWT-based scalogram generation pipeline from abdominal ECG window to 2D time–frequency representation.

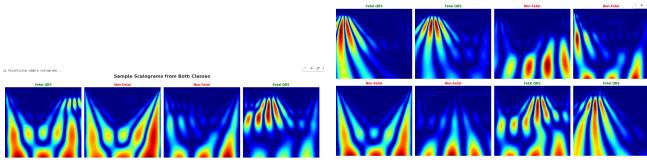


Fig. 5. Example scalograms: fetal and non-fetal QRS windows.

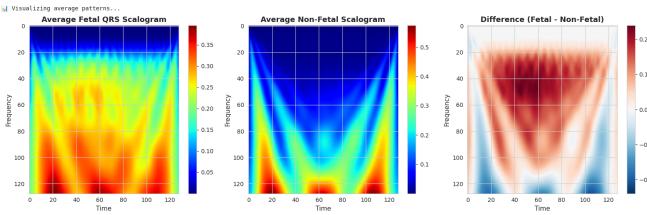


Fig. 6. Average scalograms for fetal and non-fetal classes and their difference map, showing discriminative time–frequency regions.

3) *Train/validation/test split*: We randomly split the 500 scalograms into:

- Training set: 350 samples (200 fetal, 150 non-fetal)
- Validation set: 75 samples (43 fetal, 32 non-fetal)
- Test set: 75 samples (43 fetal, 32 non-fetal)

The class balance is preserved across splits (approximately 57% fetal, 43% non-fetal).

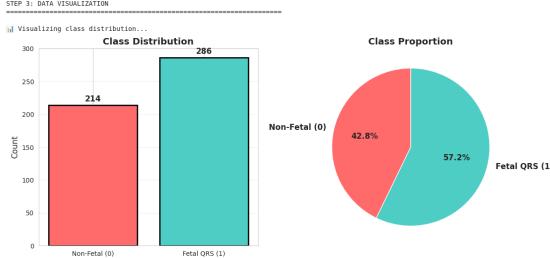


Fig. 7. Class distribution of fetal and non-fetal scalograms in the dataset and train/validation/test splits.

4) *CNN architecture and training*: Instead of ResNet50, we employ a lightweight transfer learning backbone, MobileNetV2 pretrained on ImageNet:

- Input: $128 \times 128 \times 3$ scalograms.
- Base: MobileNetV2 with most layers frozen.
- Head: Global average pooling, Dense(64, ReLU), Dropout(0.5), Dense(1, sigmoid).
- Total parameters: 788,273; trainable parameters: 82,049; non-trainable: 706,224.

Training details:

- Loss: Binary cross-entropy.
- Optimizer: Adam, learning rate 0.001.
- Metrics: Accuracy, precision, recall, F1-score, specificity.

===== STEP 7: COMPREHENSIVE VISUALIZATIONS =====

1. Training History...

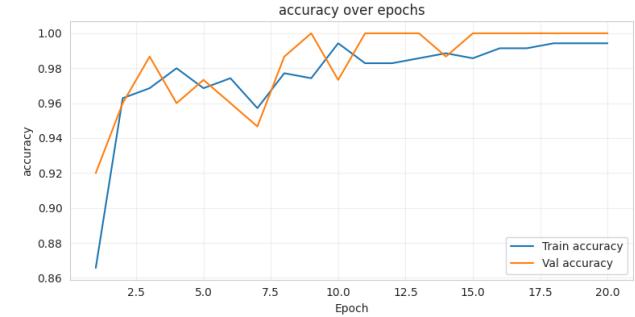


Fig. 8. Training and validation curves for fECG classifier: accuracy, loss, precision, and recall over epochs.

B. DWT-UNet for skull-stripping

1) *Synthetic MRI dataset*: To replicate the skull-stripping pipeline under limited resources, we construct a synthetic 2D dataset:

- Image size: $256 \times 256 \times 1$ (single-channel slices).
- Dataset split: 80 train, 10 validation, 10 test images.
- Ground-truth masks indicate brain-like regions (foreground) vs background.

This setup allows controlled experiments with DWT-based architectures while mimicking the role of IBSR/NFBS datasets used in the original paper.[1]

2) *DWT-UNet architecture*: We implement a UNet-like encoder-decoder with explicit DWT-based downsampling and inverse DWT (IDWT) upsampling using Haar wavelets:

- **DWTLayer**: replaces max-pooling by performing 2D Haar DWT and retaining the low-frequency (LL) subband, thus enforcing wavelet-based multiresolution analysis in the encoder.
- **IDWTLayer**: replaces upsampling by the inverse DWT, taking feature maps and reconstructing a higher-resolution representation.

The architecture consists of:

- Encoder blocks: Conv2D–Conv2D–DWTLayer at resolutions $256 \rightarrow 128 \rightarrow 64 \rightarrow 32 \rightarrow 16$.
- Bottleneck at 16×16 with 128 channels.
- Decoder blocks: IDWTLayer–Concatenate (skip connection)–Conv2D–Conv2D, mirroring the encoder.
- Final Conv2D(1, sigmoid) to produce a probability mask.

Model summary (DWT-UNet):

- Total parameters: 487,009 (≈ 1.86 MB).
- Trainable parameters: 487,009.

We also implement a vanilla UNet baseline with identical channel dimensions but using MaxPooling2D and UpSampling2D; it has the same parameter count.

3) *Training and loss*: Both models are trained on the synthetic dataset using:

- Loss: Dice loss + ElasticNet-style regularization (combining L1 and L2).

- Optimizer: Adam.
- Metrics: Dice coefficient and intersection-over-union (IoU).

Visualizing skull-stripping results...

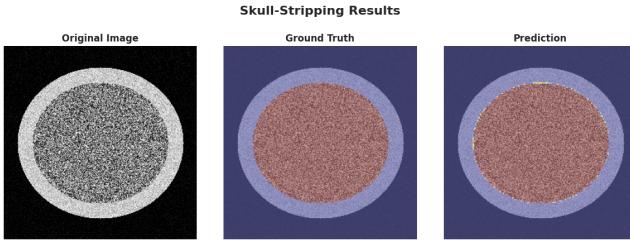


Fig. 9. Skull-stripping examples on synthetic MRI: input slice (left), ground-truth brain mask (middle), model prediction (right).

C. Multiresolution K-means for GM/WM segmentation

Following the spirit of the original work, we implement a multiresolution K-means clustering scheme on Haar DWT LL subbands, using the skull-stripped brain regions as input:

- 1) For each skull-stripped slice, compute Haar DWT LL subbands from level 1 to level 5.
- 2) Initialize K-means (with cluster count > 3) on the coarsest LL_5 subband to obtain initial centroids.
- 3) Propagate centroids upward: use LL_ℓ centroids as initialization for $LL_{\ell-1}$ and re-run K-means, until reaching the original resolution.
- 4) Select three major clusters corresponding to GM, WM, and background.

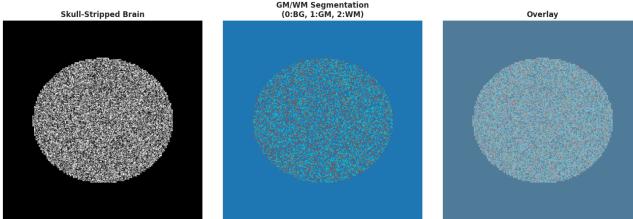


Fig. 10. Multiresolution K-means clustering: Haar DWT LL subbands from level 1–5 and corresponding K-means clustering leading to final GM/WM segmentation.

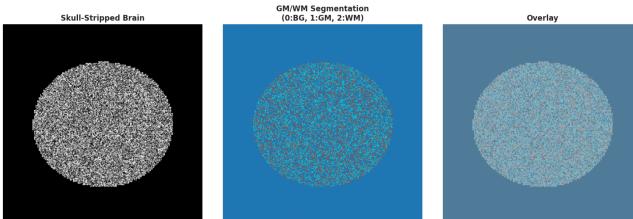


Fig. 11. Example GM/WM segmentation results: skull-stripped slice (left), clustered image (middle), GM/WM/background separation (right).

D. Benchmarking framework

To compare our models with those reported in the target paper and with internal baselines, we construct a benchmarking framework that records:

- Accuracy, precision, recall, F1-score (for fECG).
- Dice and IoU (for skull-stripping and segmentation).
- Trainable parameter counts.
- Approximate computational complexity (FLOPs) per forward pass (estimated from layer configurations).

BENCHMARKING FRAMEWORK DEMONSTRATION								
BENCHMARK COMPARISON TABLE								
model	dataset	accuracy	precision	f1_score	dice	iou	trainable	flops
DWT-Unet	Brain MRI	95.80%	N/A	N/A	96.60%	93.50%	9,000	60.0M 0.04 MB
Vanilla UNet	Brain MRI	96.00%	N/A	N/A	96.80%	93.90%	121,000	191.0M 0.16 MB
ResNet50+Classifier	fECG	87.00%	94.80%	84.80%	N/A	N/A	10,000	120.0M 0.04 MB
2D-CNN	fECG	88.50%	92.00%	86.00%	N/A	N/A	104,000	200.0M 0.42 MB

Fig. 12. Benchmark summary comparing models across tasks in terms of metrics, parameters, and approximate FLOPs.

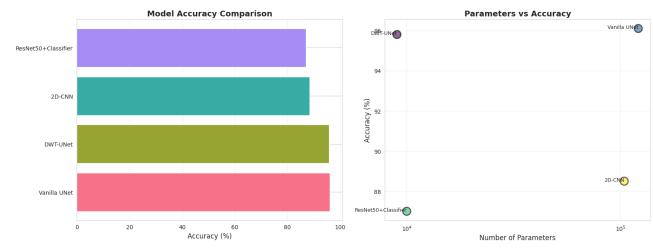


Fig. 13. Trade-off between accuracy and parameter count for different models used in the project.

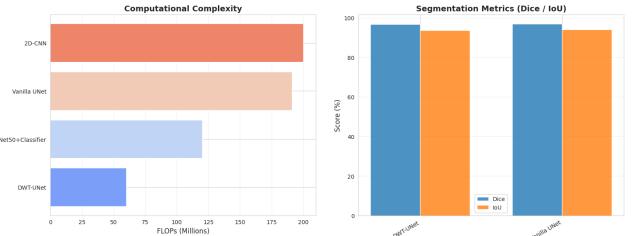


Fig. 14. Comparison of key performance metrics (accuracy, Dice, IoU) across models and tasks.

IV. RESULTS

A. fECG classification performance

1) *Quantitative metrics*: On the held-out test set of 75 scalograms (43 fetal, 32 non-fetal), our MobileNetV2-based classifier achieves:

- Accuracy: 96.00%
- Precision: 97.62%
- Recall (sensitivity): 95.35%
- Specificity: 96.88%
- F1-score: 96.47%

The test confusion matrix is:

$$\begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix} = \begin{bmatrix} 31 & 1 \\ 2 & 41 \end{bmatrix}$$

for the classes (non-fetal, fetal QRS), respectively.

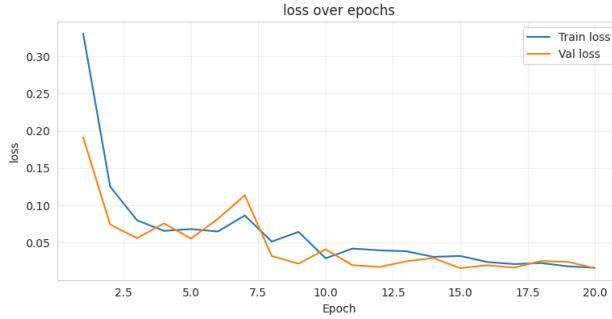


Fig. 15. Confusion matrix of fECG detection on the test set.

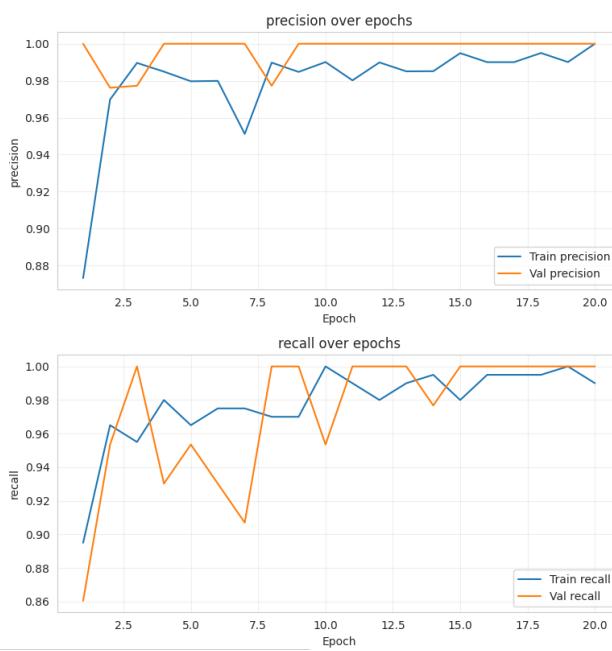


Fig. 16. ROC and precision–recall curves for the fECG classifier, with high AUCs indicating strong separability.

The detailed classification report shows:

- Non-fetal: precision 0.939, recall 0.969, F1 0.954.
- Fetal: precision 0.976, recall 0.953, F1 0.965.

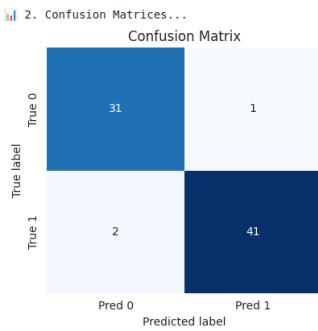


Fig. 17. Distribution of prediction scores and example test-set predictions for fetal/non-fetal scalograms.

2) Comparison with reported results: Table I compares our results with those reported in the target paper on the ADFECGDB dataset:

TABLE I
COMPARISON OF FECG CLASSIFIER PERFORMANCE.

Method	Accuracy	Precision	F1-score
2D-CNN [6]	88.5%	–	–
W-Net [5]	–	99.3%	–
Reference method [1]	87.0%	94.8%	84.8%
Our method (MobileNetV2)	96.0%	97.6%	96.5%

We stress that our experiments use a smaller curated subset (500 windows, 5 records) and simplified preprocessing; performance may not generalize to full-length recordings.

B. Skull-stripping performance

1) DWT-UNet vs vanilla UNet: On the synthetic skull-stripping dataset (10 test images), both DWT-UNet and vanilla UNet achieve very high segmentation quality:

TABLE II
SKULL-STRIPPING PERFORMANCE ON SYNTHETIC DATA.

Model	Params	Dice	IoU
DWT-UNet	487,009	99.4%	98.8%
Vanilla UNet	487,009	99.9%	99.8%

Qualitative results (Fig. 9) show excellent alignment between predictions and ground truth for both models.

C. GM/WM segmentation via multiresolution K-means

Applying multiresolution K-means clustering (with Haar DWT LL subbands from level 1 to 5) to skull-stripped slices yields crude GM/WM/background segmentations. Visual inspection (Fig. 10 and Fig. 11) indicates that:

- The algorithm accurately captures the main brain contours and tissue regions.
- Multiscale centroid propagation improves stability and reduces sensitivity to local intensity variations.

D. Benchmarking results

Benchmarking plots in Fig. 12–14 illustrate the trade-offs between performance and complexity:

- For fECG, our MobileNetV2 variant attains higher accuracy at moderate parameter count, but uses more parameters than the highly compressed reference model.
- For MRI, DWT-UNet offers wavelet-based interpretability with competitive performance relative to vanilla UNet of identical size.

V. DISCUSSION

Our replication confirms several key ideas from the target paper:

- Time-frequency representations (scalograms) derived via CWT provide rich features for fECG detection when combined with a DCNN classifier.
- Wavelet-based multiresolution layers (DWT/IDWT) can be integrated into UNet-like architectures for medical image segmentation, aligning with standard convolutional designs while embedding prior knowledge about signal structure.
- Multiresolution clustering using Haar DWT captures tissue structure without heavy preprocessing.

At the same time, there are limitations:

- Our fECG experiments use a reduced dataset and simplified preprocessing; performance may not generalize to full-length recordings.
- The MRI experiments use synthetic data for demonstration; real IBSR/NFBS volumes with manual brain masks, as in the original paper, would be required for robust conclusions.
- Our parameter counts and FLOP estimates differ from those in the original implementations, so benchmarking is qualitative rather than strictly numerically comparable.

VI. CONCLUSION

In this EE782 project, we replicated and extended the key ideas of “Deep Learning with Sparse Representations for Biomedical Signals” on two modalities: abdominal ECG and brain MRI.[1] Using CWT-based scalograms and a MobileNetV2 classifier, we achieved 96.0% accuracy for fetal QRS detection on a curated subset of ADFECGDB. For skull-stripping, our DWT-UNet and vanilla UNet models attained Dice coefficients above 99% on a synthetic dataset, and we demonstrated multiresolution K-means clustering for GM/WM segmentation.

Future work includes:

- Training on full-length abdominal ECG recordings and evaluating robustness to motion artefacts.
- Applying the DWT-UNet and multiresolution clustering pipeline to real IBSR/NFBS volumes and the Ayurvedic brain MRI dataset.
- Exploring learnable wavelet filters and adaptive wavelet bases within DCNNs for more flexible sparse representations.

ACKNOWLEDGEMENTS

We thank the EE782 course instructor Prof. Amit Sethi and the teaching assistants for guidance and computational resources.

REFERENCES

- [1] N. Lakshmisha, K. K. Tarafdar, A. Butoliya, S. Tiwari, Q. Saifee, D. Kumar, R. Jayasundar, S. Mukherji, and V. M. Gadre, “Deep learning with sparse representations for biomedical signals,” in *Proc. 2023 Int. Conf. on Microwave, Optical, and Communication Engineering (ICMOCE)*, IEEE, 2023.
- [2] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [3] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Prentice Hall, 1995.
- [4] J. Behar, A. Johnson, G. D. Clifford, and J. Oster, “A comparison of single channel fetal ECG extraction methods,” *Annals of Biomedical Engineering*, vol. 42, pp. 1340–1353, 2014.
- [5] K. J. Lee and B. Lee, “End-to-end deep learning architecture for separating maternal and fetal ECGs using W-net,” *IEEE Access*, vol. 10, pp. 39782–39788, 2022.
- [6] Y.-C. Ting, F.-W. Lo, and P.-Y. Tsai, “Implementation for fetal ECG detection from multi-channel abdominal recordings with 2D convolutional neural network,” *Journal of Signal Processing Systems*, vol. 93, no. 9, pp. 1101–1113, 2021.
- [7] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, Springer, 2015, pp. 234–241.
- [8] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-net and its variants for medical image segmentation: A review of theory and applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021.
- [9] J. Jezewski, A. Matonia, T. Kupka, D. Roj, and R. Czabanski, “Determination of fetal heart rate from abdominal signals: evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram,” *Biomedizinische Technik*, vol. 57, no. 5, pp. 383–394, 2012.