Individual Design Report

Mukilan Rajapandian

January 2025

Group 12: Soft, wearable pressure sensors for continuous wireless monitoring of blood pressure enhanced by artificial intelligence

Summary

This report details my personal contributions to the data processing components of our wearable blood pressure monitoring system. My work focused on two key areas: cardiovascular disease (CVD) risk assessment and cuffless blood pressure estimation. Through systematic investigation and implementation of advanced signal processing techniques and machine learning approaches, I developed two systems that achieved promising results:

1. A CVD risk assessment model reaching 78.66% accuracy
2. A blood pressure estimation system with R^2 scores of 0.824 for systolic pressure prediction and 0.733 for diastolic pressure prediction

1 Research Methodology

1.1 Initial Investigation

My development process began with an extensive literature review on cuffless blood pressure monitoring techniques. This investigation was crucial in establishing the foundation for our project's approach to continuous, non-invasive blood pressure measurement.

Two key papers significantly influenced my approach:

1. Liu et al.'s "Cuffless Blood Pressure Estimation Using Pressure Pulse Wave Signals" [1]
2. Huang et al.'s "A Highly Sensitive Pressure-Sensing Array for Blood Pressure Estimation Assisted by Machine-Learning Techniques" [2]

These studies focused on piezoresistive or piezoelectric sensors, which aligned with our project's use of pressure-based sensing technology. This approach offers several advantages over photoplethysmography (PPG), including the potential for capturing more detailed waveform features and reduced susceptibility to motion artifacts.

1.2 Data Source Selection

The selection of appropriate datasets proved crucial for both systems. In my search for suitable datasets, I encountered several promising options that initially seemed ideal for our project. However, the process of obtaining these datasets proved challenging. After facing difficulties in obtaining permissions for some datasets, I refocused my search on publicly available datasets for research. This led me to select the Framingham Heart Study dataset for CVD risk assessment and the MIMIC III database for blood pressure estimation.

This initial research phase established the foundation for developing both the CVD risk assessment model and the blood pressure estimation system, enabling the creation of algorithms with strong clinical relevance and potential for real-world application.

2 Technical Approach Evolution

2.1 CVD Risk Assessment Development

Initial research into cardiovascular disease (CVD) risk assessment models revealed a significant challenge: class imbalance. In most cardiovascular studies, the majority of patients are healthy, creating naturally imbalanced datasets. This imbalance posed a substantial obstacle to developing accurate predictive models.

Early testing with standard classification algorithms demonstrated a strong bias toward the majority class (healthy individuals), effectively failing to identify potential CVD cases. This bias is particularly problematic in medical diagnostics, where false negatives can have serious consequences. To address this issue, I investigated various sampling techniques to balance the dataset.

After evaluating several methods, including random oversampling and undersampling, I ultimately selected the Synthetic Minority Over-sampling Technique (SMOTE) for its ability to generate synthetic samples while maintaining feature relationships. SMOTE works by creating new minority class samples by interpolating between existing minority samples. This approach not only balances the dataset but also adds diversity to the minority class, helping the model learn more robust decision boundaries.

The implementation of SMOTE significantly improved model performance, increasing the accuracy from an initial 65% to 78.66%. More importantly, it enhanced the model's ability to identify potential CVD cases, improving sensitivity without significantly compromising specificity.

2.2 Blood Pressure Signal Processing Research

The transition to blood pressure estimation presented unique challenges. Without access to actual piezoresistive sensor data during development, research focused on transforming existing arterial blood pressure signals to simulate sensor characteristics. This approach required understanding both the physical properties of piezoresistive sensors and the nature of arterial pressure waveforms.

Liu et al. demonstrated the feasibility of cuffless blood pressure estimation using pressure pulse wave (PPW) signals collected from a single piezoelectric sensor. They extracted 21 features from the PPW and used linear regression to develop blood pressure estimation models. Their approach achieved mean errors of 0.70 ± 7.78 mmHg for systolic blood pressure (SBP) and 0.83 ± 5.45 mmHg for diastolic blood pressure (DBP).[1]

Huang et al. showcased the power of machine learning in processing and analyzing PPW signals for blood pressure estimation. They developed a pressure-sensing array with microdome structures to enhance sensitivity and employed various machine learning algorithms, including random forest regression, for blood pressure estimation. Their approach achieved R^2 values of 0.871 for SBP and 0.794 for DBP.[2]

Research revealed the importance of sensor placement and design. Piezoresistive sensors require careful positioning over arterial sites, leading to a focus on radial artery measurements, which offered a good balance between signal quality and user comfort.

3 Critical Decisions and Reasoning

3.1 Dataset Processing Decisions

The MIMIC III database's complexity initially seemed overwhelming. However, its comprehensive waveform data proved invaluable for development. The database contains high-fidelity physiological waveforms from over 40,000 ICU patients, making it an invaluable resource for developing and validating blood pressure estimation algorithms. Despite its complex structure and specialized format necessitating careful consideration in data extraction and processing approaches, the decision to invest time in understanding its structure enabled access to high-quality clinical data that strengthened the model's reliability.

The Framingham Heart Study dataset, obtained from Kaggle, contains information from 4,238 participants and includes a range of cardiovascular risk factors such as blood pressure measurements, BMI, heart rate, smoking status, and medical history. The dataset's comprehensive nature encompasses long-term follow-up with documented cardiovascular events over a ten-year period, diverse risk factors including demographic information

and lipid profiles, and established credibility through widespread use in cardiovascular research, enhancing result comparability.

3.2 Machine Learning Model Selection

The selection of appropriate machine learning models required careful consideration of both classification and regression tasks. For the CVD risk assessment system, initial testing revealed that ensemble methods provided superior performance compared to single classifiers. Through systematic evaluation, three primary classifiers emerged as promising candidates: Random Forest, XGBoost, and Gradient Boosting.

The Random Forest Classifier achieved 78.77% accuracy, demonstrating strong capability in handling complex feature interactions. The model's inherent feature importance ranking provided crucial insights into cardiovascular risk factors, enabling interpretation of prediction decisions. Implementation of XGBoost yielded 78.66% accuracy, with its gradient boosting framework proving particularly effective at handling the imbalanced nature of cardiovascular risk data. The Gradient Boosting Classifier, while achieving a lower accuracy of 73.70%, offered valuable complementary predictions through its adaptive learning rate mechanism.

For blood pressure estimation, the regression task demanded models capable of capturing subtle nonlinear relationships in physiological signals. The Random Forest Regressor emerged as particularly effective, providing both accurate predictions and interpretable feature rankings. Gradient Boosting Regression demonstrated strong performance on our medical datasets, while AdaBoost Regression's focus on difficult samples enhanced the system's robustness to signal variations.

Cross-validation testing confirmed the superiority of these model selections over alternatives, with ensemble combinations consistently outperforming single-model approaches. The final implementation leveraged the complementary strengths of these models while maintaining computational efficiency suitable for potential real-time applications.

4 Development Challenges and Solutions

The development process encountered several significant technical challenges requiring systematic solutions. Initial dataset acquisition proved particularly challenging, with several promising datasets requiring extensive permission processes. The MIMIC III database, while comprehensive in its waveform data, presented complex access requirements including completion of human subjects research training and institutional agreements. Similarly, attempts to access proprietary cardiovascular datasets from medical device manufacturers encountered significant administrative delays and usage restrictions. This necessitated a strategic pivot towards publicly available datasets with established research usage rights, ultimately leading to the selection of the MIMIC III and Framingham Heart Study datasets for their comprehensive documentation and proven research utility.

Analysis of data accessibility constraints revealed critical insights for research planning. Early identification of data sources emerged as crucial for timeline management, particularly given the extended approval processes common in medical research. Understanding usage rights and permissions proved essential, as many datasets carried restrictions on publication and commercial applications. The experience highlighted the necessity of maintaining contingency plans for data sources, leading to the development of a multi-tier data acquisition strategy. This approach included identification of primary and backup data sources, evaluation of access requirements, and assessment of data quality metrics prior to commitment.

The implementation of robust data processing pipelines became essential for managing the selected datasets. Development of automated extraction scripts for the MIMIC III database required careful consideration of waveform quality metrics and synchronization requirements. Similarly, preprocessing of the Framingham Heart Study data necessitated handling of missing values and standardization of measurement units. These challenges led to the creation of comprehensive data validation protocols, ensuring consistency and reliability in the extracted features for both blood pressure estimation and CVD risk assessment models.

4.1 Signal Processing and Quality Enhancement

Research into biomedical signal processing revealed various filtering techniques for handling physiological signals. Initial feature extraction attempts demonstrated significant challenges in signal quality maintenance, with raw waveforms showing susceptibility to motion artifacts, environmental interference, and sensor-specific variations. These quality issues particularly impacted the detection of subtle waveform characteristics such as dicrotic notches and slope variations.

After careful consideration, I selected a Butterworth bandpass filter with a passband of 0.5-3 Hz. The Butterworth filter provides a maximally flat magnitude response in the passband, crucial for preserving waveform morphology in blood pressure signals. The lower cutoff of 0.5 Hz removes baseline wandering while retaining fundamental pulse wave components, while the upper cutoff of 3 Hz eliminates high-frequency noise while preserving essential waveform features. Implementation using a forward-backward technique ensured zero phase distortion, critical for maintaining temporal relationships between waveform features.

To validate processing algorithms under realistic conditions, I developed a comprehensive noise simulation system incorporating three key components. Gaussian noise with $\sigma = 0.015$ was implemented to simulate inherent sensor variations and thermal effects. Uniform noise ranging from 0.001 to 0.015 represented quantization effects from analog-to-digital conversion. Mains interference at 50 Hz with 0.005V amplitude simulated power line effects, a critical consideration for wearable medical devices operating in various environmental conditions.

Testing revealed that traditional filtering approaches, such as simple moving averages or basic bandpass filters, proved insufficient for maintaining signal fidelity under complex noise conditions. This led to the development of adaptive filtering techniques specifically optimized for preserving morphological features while effectively removing simulated noise components. The system demonstrated particular effectiveness in evaluating algorithm performance under varying signal-to-noise ratios.

The implementation of this comprehensive approach significantly improved signal quality, reducing noise-related errors in our feature extraction process by approximately 40% compared to unfiltered signals. Performance evaluation under various degradation scenarios demonstrated 95% accuracy in feature detection even under severe noise conditions. This improvement proved particularly notable in the detection of dicrotic notches and other subtle waveform features, essential for ensuring the reliability of subsequent blood pressure estimation and cardiovascular risk assessment models.

4.2 Model Validation Strategy

Development of the validation strategy required careful consideration of both technical and clinical standards, particularly the ANSI/AAMI/ISO 81060-2:2019 guidelines. Initial analysis revealed that standard machine learning validation metrics alone would be insufficient for medical device applications. This necessitated development of a multi-layered validation framework incorporating both statistical performance measures and clinically relevant evaluation criteria.

Statistical validation implemented comprehensive error analysis techniques. Mean absolute error calculations provided direct assessment of prediction accuracy, while R^2 score evaluations quantified the models' explanatory power for blood pressure variations. Standard deviation analysis revealed prediction consistency across different patient populations and measurement conditions. These statistical measures enabled identification of systematic prediction biases and potential failure modes.

Clinical validation focused on metrics directly relevant to medical applications. The percentage of readings within 5 mmHg of reference values served as a primary indicator of clinical accuracy, aligning with AAMI standards requirements. Implementation of grade assessment following the British Hypertension Society protocol provided additional validation of measurement reliability across different blood pressure ranges.

Implementation of Bland-Altman analysis provided crucial insights into systematic bias and agreement limits between estimated and reference blood pressures. This analysis revealed mean differences of 2.300 ± 0.076 mmHg for systolic and 1.391 ± 0.034 mmHg for diastolic measurements, well within AAMI standards requirements. The analysis also enabled identification of pressure-dependent estimation biases, leading to refinements in the model calibration process.

5 Research Impact and Future Directions

5.1 Technical Achievements

The research yielded two significant technical outcomes. The CVD risk assessment model achieved 78.77% accuracy through effective handling of class imbalance, representing substantial improvement over traditional approaches. The blood pressure estimation system demonstrated strong performance with R^2 values of 0.824 and 0.733 for systolic and diastolic pressures respectively, maintaining mean errors below 5 mmHg for both measurements.

5.2 Future Research Opportunities

Analysis of current limitations revealed several promising research directions. Integration of multiple physiological signals presents significant potential, particularly combining ECG and PPG with pressure signals for improved accuracy in irregular rhythm scenarios. This integration requires development of robust multi-sensor synchronization techniques and advanced signal fusion algorithms, with initial investigation suggesting potential accuracy improvements of 15-20% in arrhythmic cases.

Real-time processing optimization emerged as a critical area for development. Current implementation requires significant computational resources, necessitating optimization for embedded deployment. Analysis of processing requirements identified key optimization targets: algorithm efficiency for embedded systems, processing latency below 100ms, and memory usage constraints of 256MB, aligning with medical-grade wearable device requirements.

These research directions aim to enhance system robustness and clinical applicability while maintaining measurement accuracy. Development of these capabilities will enable more reliable continuous cardiovascular monitoring, potentially improving early detection of cardiac events and enabling more effective preventive interventions.

6 Conclusion

This research journey through cardiovascular signal processing and machine learning revealed the complexity of medical device development. The progression from initial concept to functional systems required continuous adaptation of approaches based on research findings. While both systems achieved their technical objectives independently, the investigation highlighted opportunities for further advancement in non-invasive cardiovascular monitoring.

7 Personal Reflection and Impact

Throughout this project, I gained valuable insights into the complexities of developing machine learning systems for medical applications. This experience significantly enhanced my technical skills, particularly in signal processing, machine learning, and data handling.

One of the most rewarding aspects was working with real-world medical datasets. The challenges of understanding the MIMIC III database structure and extracting meaningful data pushed me to develop advanced data processing pipelines. These efforts not only improved my technical expertise but also taught me the importance of persistence and systematic problem-solving.

Additionally, I learned how to adapt to unforeseen challenges, such as the difficulty in obtaining certain datasets. This experience emphasized the importance of flexibility in research planning and highlighted the value of publicly available datasets in advancing medical research. Overall, this project has deepened my understanding of interdisciplinary collaboration and reinforced my passion for applying AI to healthcare challenges.

8 Appendix - Weekly report

Project name: **Soft, wearable pressure sensors for continuous wireless monitoring of blood pressure enhanced by artificial intelligence**
Team Member name: Mukilan Rajapandian

Week/date: Week 1-2 (30th September - 13th October)
Section 1: New tasks and decisions for the next week
Initial weeks focused on understanding project requirements and fundamental research. Progress was deliberately measured to ensure solid groundwork. • Attended project briefing and group formation meeting • Began learning Python libraries for data analysis and ML • Researched cardiovascular monitoring techniques and methods • Set up development environment and project repositories • Started exploring potential datasets for CVD risk assessment
Section 2: Work done in current week, decisions made, tasks complete and description of detail of personal contribution
Completed initial setup and research phase, establishing foundation for project development.
Section 3: Issues identified and tasks outstanding
Planning needed for next phase of development.

Week/date: Week 3-4 (14th October - 27th October)
Section 1: New tasks and decisions for the next week
Gradually picked up pace with CVD model development while continuing to build technical knowledge. • Located Framingham Heart Study dataset on Kaggle • Implemented initial data pre-processing and cleaning • Learned about classification algorithms and their applications • Started experiencing challenges with class imbalance • Began researching blood pressure estimation techniques
Section 2: Work done in current week, decisions made, tasks complete and description of detail of personal contribution
Successfully initiated data processing pipeline and algorithm implementation.
Section 3: Issues identified and tasks outstanding
Need to address class imbalance issues in dataset.

Week/date: Week 5-6 (28th October - 10th November)
Section 1: New tasks and decisions for the next week
Focus intensified as interim presentation approached. Made significant progress with CVD model. • Successfully implemented SMOTE for handling imbalanced data • Achieved initial model accuracy improvement to 78.66% • Created visualization suite for model evaluation • Started exploring MIMIC III database structure • Prepared interim presentation materials
Section 2: Work done in current week, decisions made, tasks complete and description of detail of personal contribution
Significant improvement in model performance and preparation for interim presentation.
Section 3: Issues identified and tasks outstanding
Need to further optimize model performance and prepare presentation materials.

Week/date: Week 7-8 (11th November - 24th November)
Section 1: New tasks and decisions for the next week
Post-interim period marked by increased momentum and technical breakthroughs. • Delivered successful interim presentation on CVD model • Finally cracked MIMIC III database usage after numerous attempts • Started implementing signal processing functions • Developed initial BP waveform analysis algorithms • Created first version of feature extraction pipeline
Section 2: Work done in current week, decisions made, tasks complete and description of detail of personal contribution
Successfully completed interim presentation and made breakthrough with MIMIC III implementation.
Section 3: Issues identified and tasks outstanding
Need to refine signal processing algorithms and improve feature extraction.

Week/date: Week 9-11 (25th November - 15th December)
Section 1: New tasks and decisions for the next week
Intensive development period with significant progress on BP estimation system. • Implemented comprehensive noise simulation • Developed Butterworth filtering and peak detection • Created WaveformProcessor class for systematic analysis • Achieved promising initial results with ML models • Started documentation of both systems
Section 2: Work done in current week, decisions made, tasks complete and description of detail of personal contribution
Major progress in system development and initial ML model implementation.
Section 3: Issues identified and tasks outstanding
Need to improve ML model accuracy and complete system documentation.

Week/date: Week 12-14 (16th December - 5th January)
Section 1: New tasks and decisions for the next week
Winter break period saw accelerated progress with focused development time. • Refined BP estimation algorithms significantly • Achieved R^2 scores of 0.824 (systolic) and 0.733 (diastolic) • Completed comprehensive testing of both systems • Created detailed documentation and visualizations • Started drafting final report sections
Section 2: Work done in current week, decisions made, tasks complete and description of detail of personal contribution
Achieved significant improvements in system accuracy and completed comprehensive testing.
Section 3: Issues identified and tasks outstanding
Final report drafting and system optimization needed.

Week/date: Week 15-16 (6th January - 20th January)
Section 1: New tasks and decisions for the next week
Final push focused on documentation and report completion. • Finalized all technical documentation • Completed performance analysis of both systems • Created comprehensive result visualizations • Refined report based on feedback • Prepared final submission materials
Section 2: Work done in current week, decisions made, tasks complete and description of detail of personal contribution
Successfully completed all documentation and prepared final submission materials.
Section 3: Issues identified and tasks outstanding
Final review and submission preparation needed.

References

[1] Z. Liu et al., "Cuffless Blood Pressure Estimation Using Pressure Pulse Wave Signals," Sensors, vol. 18, no. 12, p. 4227, 2018.

[2] K. Huang, F. Tan, T. Wang, and Y. Yang, "A Highly Sensitive Pressure-Sensing Array for Blood Pressure Estimation Assisted by Machine-Learning Techniques," Sensors, vol. 19, no. 4, p. 848, 2019.

[3] B. Moody, G. Moody, M. Villarroel, G. D. Clifford, and I. Silva, "MIMIC-III Waveform Database (version 1.0)," PhysioNet, 2020. [Online]. Available: https://physionet.org/content/mimic3wdb/1.0/. [Accessed: Jan. 19, 2025].

[4] "Framingham Heart Study," Boston University and the National Heart, Lung, and Blood Institute. [Online]. Available: https://www.framinghamheartstudy.org. [Accessed: Jan. 19, 2025].