Integrating Symbolic Regression and Photoplethysmography for Monitoring Blood Pressure Estimation

Farangis Johari[1], **Ronaldo C. Prati**[1], **Fabrício O. de França**[1]

[1]Center for Mathematics–Computing and Cognition (CMCC)
Universidade Federal do ABC (UFABC)
Santo André– SP – Brazil

farangis.johari@ufabc.edu.br, ronaldo.prati@ufabc.edu.br

fabricio.olivetti@ufabc.edu.br

Abstract. *This paper advances non-invasive blood pressure (BP) monitoring by leveraging photoplethysmography signals, enhanced through the integration of symbolic regression (SR) and traditional machine learning techniques. Our novel methodology combines traditional SR-based and feature extraction methods, utilizing recursive feature elimination with cross-validation (RFECV) for optimal feature selection. Comparative analysis across extensive datasets shows that integrating SR with RFECV enhances model transparency and predictive accuracy, providing clinically interpretable mathematical expressions that improve our understanding of BP estimation dynamics, which is crucial for healthcare diagnostics.*
Keywords*: Non-invasive blood pressure monitoring, Machine learning techniques, Photoplethysmography (PPG), Symbolic regression (SR).*

1. Introduction

The strong association between hypertension and significant health risks such as stroke and renal dysfunction underscores the essential need for a precise and continuous monitoring of blood pressure (BP). Often called the "silent killer," hypertension is an major contributor to mortality because it generally lacks symptoms, leading to underdiagnosed and untreated conditions, as highlighted in analyses by [Kearney et al. 2005] and [Bittner 2020]. This critical context accentuates the urgency for advancements in BP monitoring technologies.

Traditional BP monitoring methods, particularly invasive techniques, have been critically reviewed due to their complexity and risk of complications. Non-invasive methods, such as sphygmomanometers, although widely used, do not support continuous monitoring due to the impracticality of constant manual operation [Liang et al. 2018]. Recent advances in photoplethysmography (PPG) for BP monitoring have focused on improving accuracy and reliability through machine learning (ML) and deep learning (DL) techniques. In particular, the U-Net architecture predicts arterial BP waveforms from fingertip PPG signals with high precision [Athaya and Choi 2021]. By analyzing continuous waveform data from both PPG and arterial blood pressure (ABP) measurements, these approaches enhance the precision of BP estimates without frequent recalibration [Kachuee et al. 2015].

Recent advances in signal processing and algorithmic interpretation have enhanced the analysis of PPG data, leading to better clinical outcomes in the management of hypertension. Studies support the integration of non-invasive continuous monitoring technologies like PPG into routine clinical practice, suggesting that these methods could improve the early detection and management of hypertension, thus reducing associated diseases [Liang et al. 2018]. ML models leverage rich PPG data to predict BP, extracting features such as pulse arrival time, waveform morphology, and frequency domain characteristics. Techniques such as support vector machines (SVM), random forests (RF) and neural networks have been utilized [Wong et al. 2023].

ML approaches for BP estimation from a single PPG waveform have gained significant attention. Despite numerous ML-based techniques, an optimal methodology remains unclear. A benchmark was established using four open datasets, standardized preprocessing, robust validation, and consistent evaluation metrics. The refined mean absolute scaled error (MASE) improved interpretability, facilitating a comparative analysis of 11 ML-based BP estimation methods [González et al. 2023].

In this context, symbolic regression (SR) emerges as a key methodology. SR is an evolving subfield within ML that focuses on the derivation of symbolic mathematical expressions from the data [Schmidt and Lipson 2009], emphasizing interpretability alongside accuracy [Schmidt and Lipson 2009, Rudin 2019]. Rooted in genetic programming, SR has gained renewed interest in DL advancements, making it powerful in various disciplines [Makke and Chawla 2024].

SR simplifies expressions using the rational transformation-interaction approach, balancing simplicity and approximation capabilities, especially in small datasets [de França 2023]. Benchmarks evaluating different explanations have shown that SR is a viable alternative to traditional ML models [Seidyo Imai Aldeia and Olivetti de Franca 2024].

Addressing the lack of standardized benchmarking, an open-source platform was introduced to evaluate 14 SR and 7 ML methods across 252 regression problems, demonstrating the effectiveness of genetic algorithms combined with parameter estimation or semantic search drivers in real-world scenarios [Cava et al. 2021]. The Operon framework, utilizing local search optimizations, balances accuracy, and simplicity, achieving high performance in synthetic track experiments [Burlacu et al. 2020, Burlacu 2023].

In healthcare, the feature engineering automation tool (FEAT) uses SR to create precise and interpretable clinical prediction models from electronic health records (EHR), enhancing clinical decision support and trust in ML applications [La Cava et al. 2023].

1.1. Contributions

This paper significantly advances non-invasive BP monitoring by improving clinical interpretability and predictive accuracy through a novel hybrid approach that combines SR with traditional ML techniques. The specific contributions of our research are detailed as follows:

- Novel methodology: The integration of SR and recursive feature elimination with cross-validation (RFECV) optimizes feature selection, improving BP predictions from PPG signals.

- Improvement of clinical interpretability: By generating explicit mathematical expressions, SR bridges the gap between model accuracy and clinical interpretability, improving model transparency and aiding informed clinical decision-making.
- Comparative analysis: We evaluated several machine learning models, including support vector regression (SVR), adaptive boosting (AdaBoost), and SR-PyOperon. This evaluation helps identify the most effective algorithms for integration with advanced feature extraction techniques, guiding the selection of optimal models for specific clinical applications, and balancing accuracy with computational efficiency.
- Advancement of non-invasive monitoring technologies: The work highlights the potential of PPG-based BP monitoring enhanced by ML and SR techniques, offering a non-invasive solution that meets stringent medical standards, thus improving patient comfort and safety.
- Healthcare diagnostics: The methodology provides insights into BP estimation dynamics, which are crucial for reliable healthcare diagnostics, aiding in early detection and management of hypertension.
- Benchmarking and standardization: The paper addresses the need for standardized benchmarking by evaluating the proposed method against established datasets with robust validation strategies to ensure reliable results.

2. Methodology

This work evaluates various ML and DL models for non-invasive BP monitoring using PPG signals. The evaluation is based on a benchmark developed by Gonzalez et al. [González et al. 2023], using their data sets. These data sets include diverse data from the subject, BP distributions, and recording characteristics. Standardized preprocessing and a validation strategy were applied to maintain data integrity across training, validation, and test sets. Specifically, we used the sensor dataset, which is divided into five folds for a comprehensive analysis.

2.1. Feature-to-Label Methodology

Feature-to-Label (Feat2Lab) is a method proposed by Gonzalez et al. [González et al. 2023] to extract features from PPG signal. It identifies the most effective and widely utilized PPG features, classified into three main groups:

1. **Points-of-Interest and Time-Based Features**: Focus on specific points in the PPG cardiac cycle and its derivatives (VPG and APG), such as the systolic peak and various points from the first and second derivatives. Features include amplitudes, elapsed times, and areas under the PPG curve.
2. **Frequency-Based Features**: Derived from fast Fourier transform (FFT) analysis of the PPG waveform, these features include the dominant frequency, its magnitude, and the average magnitude of the surrounding frequencies.
3. **Operational and Statistical Features**: Provide a comprehensive characterization of the PPG cardiac cycle, including histogram features, slope deviation curve (SDC) features, skewness and kurtosis, and indices such as the aging index.

2.2. PyOperon: Leveraging Symbolic Regression for Blood Pressure Estimation

PyOperon is a cutting-edge symbolic regression library that employs evolutionary algorithms to generate interpretable mathematical models from complex datasets. It is based on the C++ Operon framework for symbolic regression that uses genetic programming to explore a hypothesis space of possible mathematical expressions. Unlike traditional regression techniques, symbolic regression does not assume a predefined model structure. Instead, it explores the space of mathematical expressions to find equations that best fit the data. This approach is particularly powerful in medical applications like BP estimation, where understanding the underlying relationships between physiological signals and outcomes is crucial.

2.2.1. Functionality and Implementation

In our work, PyOperon was used to derive symbolic equations that model the relationship between PPG signal features and blood pressure (SBP and DBP). The process begins with a population of randomly generated mathematical expressions, which are iteratively refined through operations such as crossover, mutation, and selection. These operations mimic the principles of natural evolution, allowing the algorithm to explore a wide range of potential solutions. To optimize feature selection for predicting SBP and DBP, we applied RFECV [Guyon and Elisseeff 2003]. This is a technique that iteratively removes less significant features and builds a model using the remaining ones. This process was crucial in identifying the key PPG features that contribute to accurate BP predictions.

SR predictions were generated using the PyOperon library, with feature significance determined by configurable hyperparameters. These hyperparameters—such as mutation probability, population size, and the number of generations—were meticulously optimized using the Optuna library to enhance model performance. The optimization process, managed through Optuna, was performed using internal cross-validation on the training set, ensuring robustness and generalizability. Key hyperparameters include:

- **Mutation Probability**: Controls the likelihood of random alterations in the mathematical expressions, enabling the discovery of novel and potentially better-performing models.
- **Population Size**: Determines the number of candidate solutions in each generation, balancing the exploration of the solution space with computational efficiency.
- **Generations**: Specifies the number of evolutionary iterations, allowing the model to progressively improve its fit to the data.

2.2.2. Application in BP Estimation

The application of PyOperon in this study involved several critical steps:

- **Feature Selection**: Initially, features were selected from the PPG signal data using RFECV, ensuring that only the most relevant features were used in the symbolic regression process. This step enhanced the interpretability and accuracy of the final model. Feature selection was conducted on five data folds from the sensor dataset, with average performance metrics calculated for both traditional models and the SR model from PyOperon. This process is referred to as *SR_fold* in the analysis.

- **Symbolic Regression**: Using the selected features, PyOperon generated symbolic equations that map the PPG features to BP values. The flexibility of symbolic regression allowed the model to capture complex, nonlinear relationships that might be overlooked by more conventional methods.
- **Model Optimization**: The hyperparameters of PyOperon were optimized using Optuna, a robust framework that systematically tunes the model's settings to achieve higher accuracy and lower error. This optimization, performed with cross-validation, ensured that the model generalized well to unseen data.
- **Model Evaluation**: The symbolic regression models were evaluated using various metrics like MAE, MSE, and the coefficient of determination (R^2). These metrics provided a comprehensive assessment of the model's predictive accuracy and its potential utility in clinical settings.
- **Interpretability**: A significant advantage of PyOperon is the interpretability of the models it generates. Unlike black-box machine learning models, the symbolic equations produced by PyOperon can be directly examined and understood by clinicians, making them more likely to be adopted in practice.

2.3. Evaluation metrics

Following best practices in BP research, several metrics were used to evaluate model performance:

- **Mean Absolute Error (MAE)**: Quantifies the average magnitude of the errors in predictions, disregarding their direction: $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
- **Mean Squared Error (MSE)**: Used to assess the performance of regression models by averaging the squares of prediction errors: $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- **Naive Mean Absolute Error (Naive MAE)**: A baseline metric used for comparison, calculated by taking the mean absolute difference between the actual BP values and a naive prediction model, which typically uses a simple heuristic or historical mean as the predictor.
- **Mean Absolute Scaled Error (MASE)**: Facilitates comparison between different datasets by normalizing the MAE against the Naive MAE, derived from the mean output of the data set:
$$MASE = \frac{MAE}{\text{Naive MAE}}$$
This metric is particularly useful for evaluating and comparing algorithms under consistent standards, especially in time-series forecasting of BP measurements.
- **Coefficient of determination (R^2)**: Measures the proportion of variance in the dependent variable that is predictable from the independent variables:
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$
where $\bar{y}$ is the mean of the actual values.
- **Modified Mean Squared Error (MMSE)**: Facilitates comparison between different datasets by normalizing the MSE against the Naive MAE, derived from the mean output of the data set:
$$MMSE = \frac{MSE}{\text{Naive MAE}}$$

3. Evaluating Machine Learning Techniques in Blood Pressure Estimation: Results and Discussion

In this work, we systematically evaluate various ML models, categorizing them into specific groups to assess their effectiveness in estimating BP using PPG signals. Our comparative analysis, based on the data presented in Table 2, provides a comprehensive evaluation of the performance of different algorithms across various metrics. The proficiency of each model is highlighted in bold, indicating superior performance among the different ML models. These models were rigorously tested and trained using a five-fold cross-validation process. The set of hyperparameter values employed during the grid search in PyOperon is detailed in Table 1.

Table 1. Hyperparameter values used during Symbolic Regression

Parameter	**Value**
Offspring Generator	brood
Initialization Method	btc
Comparison Factor	0.1
crossover probability	0.9
Epsilon	1×10^{-5}
Female Selector	tournament
Objectives	r2
Mutation Probability	$[0.3, 0.8, 0.9]$
Reinserter	keep-best
Max Evaluations	1×10^{6}
Tournament Size	3
Pool Size	500
Population Size	[500, 1000]
Generations	[400, 800, 1000]
Time Limit	90
crossover internal probability	0.3
Max Depth	5
Max Length	[5, 10,20, 50, 100, 150]
initialization max length	[100, 150]

3.1. Analysis and Discussion of SBP and DBP for Sensor-dataset

Table 2 shows the performance of various machine learning algorithms in predicting SBP and DBP using PPG signals. The models assessed include AdaBoost, Naive, RF, SVR, and SR. The RF model consistently outperforms other models across all evaluated metrics (MSE, MAE, Score, MMSE, and MASE) for both SBP and DBP predictions. This demonstrates its effectiveness in leveraging PPG signals for accurate blood pressure estimation. The AdaBoost model performs moderately well but does not surpass the RF model in any metric. The Naive model, serving as a baseline, shows the highest errors, highlighting the complexity of accurately predicting blood pressure using PPG signals without advanced machine learning techniques. Symbolic Regression models are examined for their potential to derive interpretable mathematical models from sensor data, although they may encounter challenges with large or complex datasets compared to more robust algorithms. Among the metrics considered, MASE is particularly valuable due to its scale-independent nature and its ability to provide a normalized measure of error relative to simple naive predictions. The low MASE in the RF model signifies not only its performance against basic benchmarks but also its ability to manage the variability of the SBP and DBP data.This assessment aids in selecting appropriate machine learning tools to enhance SBP and DBP monitoring and management in healthcare applications, ensuring alignment with specific clinical needs and dataset characteristics.

Table 2. Performance of the ML algorithms grouped on Sensor dataset for SBP and DBP

Models	Targets	Metrics				
		MSE	*MAE*	*Score*	*MMSE*	*MASE%*
Naive	SBP	468.1	17.52	-0.0023	1.0023	100.30
	DBP	110.25	8.23	0.0015	1.0015	100.21
AdaBoost	SBP	332.75	14.49	0.2827	0.7173	83.25
	DBP	87.72	7.66	0.1949	0.8051	93.60
RF	SBP	**320.27**	**13.78**	**0.3059**	**0.6941**	**79.34**
	DBP	**62.63**	**5.93**	**0.4220**	**0.5780**	**72.62**
SVR	SBP	467.87	17.44	-0.0024	1.0024	99.89
	DBP	90.14	7.07	0.1684	0.8316	86.43
SR	SBP	403.43	16.09	0.1437	0.8562	91.92
	DBP	98.13	7.70	0.1084	0.8915	93.64

Figure 1 displays MASE percentages for SBP and DBP across various ML models. The models are ordered from the lowest to highest MASE values, which allows for easy comparison of their performance.

SBP MASE Analysis: RF has the lowest MASE percentage (79.34%), indicating the highest accuracy among the models for predicting SBP. AdaBoost follows with a slightly higher MASE percentage (83.25%). SR_ fold, SVR, and Naive models have progressively higher MASE percentages, with Naive having the highest MASE (100.3%).

DBP MASE Analysis: RF again shows the best performance with the lowest MASE percentage (72.62%). SVR and AdaBoost have similar performance, with MASE percentages around 86.43 and 93.6%, respectively. SR_fold and Naive have the highest MASE percentages, with Naive reaching 100.21%.

The RF model consistently outperforms the other models in both SBP and DBP predictions, demonstrating the lowest MASE percentages. This suggests that RF is the most reliable model for blood pressure estimation in this dataset. The Naive model, on the other hand, shows the highest MASE percentages, indicating poorer performance compared to the other models.

RF consistently demonstrates the best performance for both SBP and DBP predictions, with the lowest MMSE values in both cases. AdaBoost is also a strong performer, showing competitive results close to RF. Symbolic Regression and SVR show moderate performance, with higher MMSE values than RF and AdaBoost. Naive model exhibits the highest MMSE values for both SBP and DBP, indicating that it is the least effective model among those tested.

3.2. Symbolic Regression

This section delves into the derivation and implications of an equation formulated to estimate SBP and DBP based on various features extracted from the sensor dataset. The following equation represents the best expressions obtained through combining folds for the DBP target obtained in

$$Y_{DBP} = 52.730 + 5.877 \cdot \left(\frac{0.776}{-0.357 \cdot \text{DiaRise}} \right) \cdot \left(\tanh \left(\sqrt[3]{\sin \left(\sqrt[3]{-0.029 \cdot \text{ppg\_fft\_peaks\_neighbor\_avgs\_0}} \right)} \right) \right)$$

We analyze the components of the equation:

1. **Constant Term (52.730):** This is the baseline value in the equation, indicating the initial DBP without the influence of other variables.
2. **Coefficient (5.877):** This multiplier affects the overall contribution of the non-linear transformation to the final DBP value.
3. **Hyperbolic Tangent Function ($\tanh$):** This function maps the input values to a range between -1 and 1, introducing non-linearity and controlling extreme values.
4. **Cube Root of Sine Function ($\sqrt[3]{\sin(\sqrt[3]{\cdot})}$):** The nested sine and cube root functions apply a non-linear transformation to the ‘ppg_fft_peaks_neighbor_avgs_0‘ feature, making the equation sensitive to this feature’s variations.
5. **Fractional Component ($\frac{0.776}{-0.357 \cdot \textbf{DiaRise}}$):** This part introduces another layer of complexity, incorporating the ‘DiaRise‘ feature inversely. The negative coefficient suggests an inverse relationship between ‘DiaRise‘ and the contribution to DBP.

The feature *ppg_fft_peaks_neighbor_avgs_0* likely represents a frequency-domain characteristic of the PPG signal, indicating the average peak values of neighboring points in the FFT spectrum. Its transformation through non-linear functions suggests that it plays a critical role in capturing subtle variations in the signal that correlate with DBP. The feature *DiaRise* appears to be a characteristic in the time domain, probably related to the rate

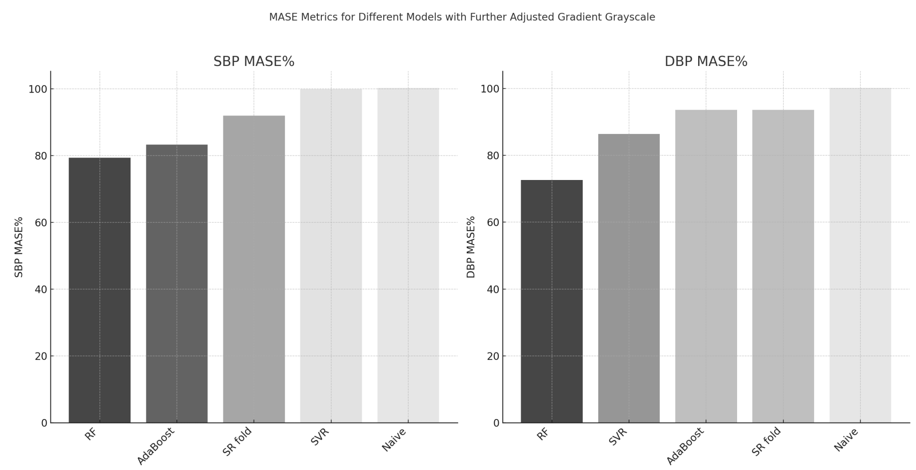

Figure 1. Performance Comparison of MASE% for SBP and DBP

of change or slope of the diastolic rise phase in the sensor signal. The inverse relationship implies that higher values of *DiaRise* are associated with lower DBP predictions. The equation incorporates several non-linear transformations essential for modeling complex physiological processes that linear relationships cannot adequately capture. The interaction between *ppg_fft_peaks_neighbor_avgs_0* and '*DiaRise* through the non-linear functions and fraction indicates that the model considers combined effects rather than treating each feature independently. The equation derived for DBP reflects an advanced approach to physiological signal processing, utilizing sophisticated mathematical transformations to improve prediction accuracy. It underscores the importance of non-linear modeling in capturing the intricate dynamics of cardiovascular signals.

The best equation for SBP takes into account various physiological parameters, such as pulse pressure (*ppg_min_0*), normalized cardiac time (*T_c_norm*), and ultrasound-derived diameter change (*usdc_3*). The equation provided:

$$Y_{SBP} = 81.645 + \left(-0.598 \times \frac{(-1.952 \times (\exp(1.187 \times \text{ppg\_min\_0}))^2)}{(1.419 \times \text{T\_c\_norm}) - (0.368 \times \text{usdc\_3})} \right)$$

aims to predict SBP based on specific physiological inputs.

To analyze this equation, we need to examine each component and its influence on the overall prediction of SBP:

1. **Baseline Constant**: The equation begins with a baseline constant of 81.645, serving as a reference point for systolic blood pressure (SBP). This ensures that the model starts with a realistic SBP value.
2. **Modifier Coefficient (-0.598)**: The negative coefficient (-0.598) indicates an inverse relationship between the ratio of the parameters and SBP.
3. **Numerator**: The numerator features an exponential function influenced by *ppg_min_0* (minimum pulse pressure). The term $(-1.952 \times (\exp(1.187 \times \text{ppg\_min\_0}))^2)$ highlights the significance of pulse pressure in predicting SBP. This exponential component amplifies the effect of *ppg_min_0*, meaning that small changes in pulse pressure lead to substantial variations in SBP due to the squared exponential function.
4. **Denominator**: The denominator incorporates normalized cardiac time (*T_c_norm*) and ultrasound-derived diameter change (*usdc_3*). The term $(1.419 \times \text{T\_c\_norm}) - (0.368 \times \text{usdc\_3})$ shows how these two parameters balance each other in influencing SBP. The positive coefficient (1.419) suggests that longer normalized cardiac times increase SBP, while the smaller coefficient (0.368) indicates that greater diameter changes, which are subtracted, lead to lower SBP predictions.

This equation provides a robust framework for predicting SBP by integrating key physiological metrics. The exponential amplification of the pulse pressure, combined with the balanced contributions of cardiac time and diameter change, ensures a nuanced approach to the estimation of SBP. Future work should focus on validating this model with empirical data to enhance its predictive accuracy and reliability.

3.2.1. Comparative Analysis of Symbolic Regression

An analysis of the aspects of SR highlighting its distinct advantages and challenges underscores the unique position of SR within the broader spectrum of machine learning techniques. This analysis emphasizes the potential for SR for applications where transparency and simplicity are as important as predictive capability.

- **Interpretability:** SR's primary advantage lies in its ability to produce understandable mathematical models from complex data sets. This feature is particularly valuable in domains where stakeholders require clarity on how decisions are derived from the model.
- **Accuracy Challenges:** Although SR excels in interpretability, it may not always achieve the highest accuracy, especially when compared to more complex machine learning models. This aspect often presents a trade-off between understanding the model's outputs and achieving the lowest possible prediction errors.
- **Verification:** The clarity and transparency of SR models foster trust and allow for easy verification of results, which is essential in sensitive fields such as healthcare. Medical professionals can understand and justify automated decisions made based on SR, which is crucial for clinical acceptance.
- **Integration into Clinical Practice:** The ability of SR to integrate into clinical practice enhances its utility. Decisions based on SR models can be transparent and justifiable, aligning with the regulatory and ethical standards required in healthcare environments.
- **Model Complexity and Implementation:** One of the benefits of SR is its relative simplicity in terms of model structure. SR can often distill complex phenomena into simpler, more comprehensible mathematical expressions. Despite its conceptual simplicity, implementing SR effectively can be challenging, especially in handling noisy or incomplete data. This may require sophisticated preprocessing steps or enhancements to the basic SR algorithm.
- **Scalability and Performance:** SR models are generally less resource-intensive, making them more scalable to larger datasets compared to deep learning models. This scalability is advantageous in settings where computational resources are limited. SR's performance can vary significantly depending on the nature of the data and the specific configurations of the SR algorithm. Fine-tuning SR to maintain robustness across different scenarios remains a critical area for ongoing research.

4. Conclusions and Future Work

This work presents a thorough evaluation of various ML algorithms for non-invasive BP estimation using PPG signals. Among the models evaluated, the RF model consistently demonstrated superior accuracy and reliability in multiple metrics, including MSE, MAE, Score, MMSE, and MASE. The robust performance of the RF model highlights its effectiveness in leveraging PPG signals for precise BP estimation, establishing it as a valuable tool for clinical applications. SR models, while not achieving the highest accuracy, offered significant advantages in interpretability. This feature is particularly crucial in healthcare settings, where understanding and verifying the decision-making process is essential. The transparency and simplicity of SR models foster trust and facilitate integration into clinical practice, aligning with regulatory and ethical standards. Despite the moderate performance of other models, such as AdaBoost and SVR, the evaluation underscores the importance of selecting models based on the specific requirements of the application, balancing trade-offs between accuracy, interpretability, and computational resources.

The findings of this work open several avenues for future research. Key areas of focus should include:

- **Enhancement of Symbolic Regression Models:** While SR models excel in interpretability, their accuracy can be further improved. Future research should explore advanced preprocessing techniques and algorithmic enhancements to handle noisy or incomplete data effectively. Furthermore, developing hybrid models that combine the strengths of SR with other ML techniques could provide a balanced approach to accuracy and interpretability.
- **Comprehensive Benchmarking:** Extending the benchmarking process to include a broader range of datasets and real-world scenarios will ensure the robustness and generalizability of the evaluated models. This effort should also incorporate standardized protocols to facilitate consistent and reliable comparisons across studies.
- **Clinical Validation and Implementation:** To bridge the gap between research and practical application, extensive clinical validation of the proposed models is necessary. Collaborations with healthcare institutions will be crucial in testing the models in various clinical settings, ensuring their reliability and acceptance by medical professionals.
- **Scalability and Resource Optimization:** Given the resource-intensive nature of some ML models, future work should focus on optimizing the computational efficiency of these algorithms. This includes exploring lightweight models and efficient training methodologies that can scale to large datasets without compromising performance.
- **Integration with Wearable Technology:** The integration of ML-based BP estimation models with wearable devices offers a promising direction for continuous and non-invasive health monitoring. Research should investigate the feasibility, accuracy, and user acceptability of such integrations, with the aim of improving patient comfort and health outcomes.

5. Acknowledgments

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001 at the Federal University of ABC–UFABC. Fabricio Olivetti de Franca received support from CNPq through the Grant 301596/2022-0.

References

Athaya, T. and Choi, S. (2021). An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: A u-net architecture-based approach. *Sensors*, 21(5):1867.

Bittner, V. (2020). The new 2019 aha/acc guideline on the primary prevention of cardiovascular disease. *Circulation*, 142(25):2402–2404.

Burlacu, B. (2023). Gecco'2022 symbolic regression competition: Post-analysis of the operon framework. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, pages 2412–2419.

Burlacu, B., Kronberger, G., and Kommenda, M. (2020). Operon c++: An efficient genetic programming framework for symbolic regression. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, GECCO '20, page 1562–1570, New York, NY, USA. Association for Computing Machinery.

Cava, W. G. L., Orzechowski, P., Burlacu, B., de França, F. O., Virgolin, M., Jin, Y., Kommenda, M., and Moore, J. H. (2021). Contemporary symbolic regression methods and their relative performance. *CoRR*, abs/2107.14351.

de França, F. O. (2023). Alleviating overfitting in transformation-interaction-rational symbolic regression with multi-objective optimization. *Genetic Programming and Evolvable Machines*, 24(2):13.

González, S., Hsieh, W.-T., and Chen, T. P.-C. (2023). A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram. *Scientific Data*, 10(1):149.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182.

Kachuee, M., Kiani, M. M., Mohammadzade, H., and Shabany, M. (2015). Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE int. symp. on circuits and systems (ISCAS)*, pages 1006–1009. IEEE.

Kearney, P. M., Whelton, M., Reynolds, K., Muntner, P., Whelton, P. K., and He, J. (2005). Global burden of hypertension: analysis of worldwide data. *The lancet*, 365(9455):217–223.

La Cava, W. G., Lee, P. C., Ajmal, I., Ding, X., Solanki, P., Cohen, J. B., Moore, J. H., and Herman, D. S. (2023). A flexible symbolic regression method for constructing interpretable clinical prediction models. *NPJ Digital Medicine*, 6(1):107.

Liang, Y., Chen, Z., Ward, R., and Elgendi, M. (2018). Photoplethysmography and deep learning: enhancing hypertension risk stratification. *Biosensors*, 8(4):101.

Makke, N. and Chawla, S. (2024). Interpretable scientific discovery with symbolic regression: a review. *Artificial Intelligence Review*, 57(1):2.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215.

Schmidt, M. and Lipson, H. (2009). Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85.

Seidyo Imai Aldeia, G. and Olivetti de Franca, F. (2024). Interpretability in symbolic regression: a benchmark of explanatory methods using the feynman data set. *arXiv e-prints*, pages arXiv–2404.

Wong, M. K. F., Hei, H., Lim, S. Z., and Ng, E. Y. K. (2023). Applied machine learning for blood pressure estimation using a small, real-world electrocardiogram and photoplethysmogram dataset. *Mathematical Biosciences and Engineering*.