# Fetal electrocardiogram prediction using machine learning: a random forest-based approach

**Mohammed Moutaib[1], Mohammed Fattah[1], Yousef Farhaoui[2], Badraddine Aghoutane[3], Moulhime El Bekkali[4]**

[1]IMAGE Laboratory, University of Moulay Ismail, Meknes, Morocco
[2]L-STI, T-IDMS, Faculty of Science and Technics, University of Moulay Ismail, Errachidia, Morocco
[3]Faculty of Science, University of Moulay Ismail, Meknes, Morocco
[4]IASSE Laboratory, Sidi Mohamed Ben Abdellah University, Fez, Morocco

**Article Info**

***Article history:***

Received Sep 15, 2023
Revised Nov 17, 2023
Accepted Nov 30, 2023

***Keywords:***

Artificial neural network
Electrocardiogram
Fetal electrocardiograms
K-means
Machine learning
Recurrent neural network
Support vector machines

**ABSTRACT**

Monitoring fetal health during pregnancy ensures safe delivery and the newborn's well-being. The fetal electrocardiogram (fetal ECG) is a valuable tool for assessing fetal cardiac health, but interpretation of ECG data can be challenging due to its complexity and variability. In this work, we explore the application of machine learning, particularly random forest, to predict and analyze fetal ECGs. With its ability to manage large datasets and provide precise insights, random forest is a promising solution for this challenge. By comparing our random forest-based approach with other standard machine learning techniques such as artificial neural network (ANN), support vector machines (SVM), and recurrent neural networks (RNN), we observed that our solution outperformed these methods in accuracy, robustness, and reliability. This article details the methodology used, the implementation of the algorithm, as well as the comparative results obtained. Emphasis is placed on the benefits of random forest in this specific medical context, highlighting its potential as a future tool for fetal ECG prediction. Ultimately, our research suggests a shift toward random forest-based solutions for more efficient and accurate analysis of fetal ECGs, with direct implications for clinical practice and fetal well-being.



***Corresponding Author:***

Mohammed Moutaib
IMAGE Laboratory, Moulay Ismail University
Meknes, Morocco
Email: mohammed.93@live.fr

## 1. INTRODUCTION

Fetal health is a crucial indicator of an unborn fetus's overall well-being and development. A significant aspect of this monitoring is assessing fetal heart health, often performed via the fetal electrocardiogram (fetal ECG) [1]. These cardiac data can be indicative of a variety of medical conditions, including congenital disabilities and warning signs of fetal distress. Although the fetal ECG is a powerful diagnostic tool, its interpretation often requires significant expertise and is subject to human error [2]. In a world where healthcare systems are under increasing pressure, reliable and automated analysis of this data is not only desirable but almost essential [3].

Despite the importance of fetal ECG signals for early diagnosis and prevention of potential complications, current analysis methods face several limitations, such as the risk of human error, lack of automation, and sensitivity to signal disturbances. However, interpreting this data can be complex with the vast data set that an ECG produces and the subtle nuances it can contain. Hence the importance of exploiting technological advances [4], particularly in the field of machine learning, to improve and refine these interpretations. With their ability to process and analyze large amounts of data, machine learning methods offer remarkable potential to revolutionize the way we approach fetal ECG.

Previous approaches to fetal ECG analysis include rule-based methods, statistical algorithms, and heuristic models [5]. However, these methods often lack precision and do not adapt well to variations and anomalies in the data. The application of machine learning in this area is still being explored, with some work focusing on algorithms such as neural networks, support vector machines (SVMs), and random forests [6], [7]. Our study focuses precisely on this intersection between medicine and technology, seeking to deploy the power of machine learning, and more specifically, random forest, to predict and analyze fetal ECGs. Faced with the multitude of machine learning techniques available, why choose the random forest [8], [9].

The main objective of this article is to propose an automated approach for predicting fetal ECGs using a machine-learning algorithm, specifically random forest. We aim to develop a model that is accurate and robust to different signal qualities and various clinical conditions to predict the state of the fetal heart before birth. Machine learning methods, such as random forest, allow the processing of large data sets and provide real-time results. This is crucial for timely medical interventions in case of complications. Furthermore, this study can potentially reduce healthcare professionals' workload and increase fetal monitoring efficiency.

The article is organized as follows: after this introduction, we review relevant literature on fetal ECG prediction methods and machine learning algorithms applied to this area. Next, we detail the methodology employed, including data collection and preprocessing, construction and training of the random forest model, and evaluation of its performance. The results are then presented and discussed, followed by our conclusions and recommendations for future research. Therefore, this study contributes to filling a gap in current research by proposing a random forest-based approach for efficient and reliable prediction of fetal ECGs.

## 2. LITERATURE REVIEW

Fetal monitoring, particularly using the fetal ECG, is critical to prenatal and perinatal care. Although traditional methods for assessing these ECGs are relatively effective, they have flaws. They often require considerable clinical expertise and are prone to human error. This has led to continued research to improve the accuracy and effectiveness of fetal monitoring methods.

Machine learning technologies have provided a new avenue to address these challenges. Smith *et al.* [10] were among the pioneers exploring the use of neural networks to analyze fetal ECGs. Their results suggest that these models can outperform traditional methods, particularly in classifying fetal arrhythmias. Johnson *et al.* [11] continued using deep neural networks and found that they provided better sensitivity and specificity than previous methods.

In parallel, SVM have also been studied. Williams *et al.* [12] showed that SVMs could be particularly effective in handling high-dimensional data and multi-class classification problems. This work was crucial in showing that SVMs are not only competitive in terms of performance but also robust enough to handle the intrinsic variability of fetal ECGs. More recently, Lee *et al.* [13] explored using random forests in this area. Their work is fascinating because it did not just highlight the high precision of this method; it also demonstrated that random forests could be used to understand important features that influence fetal ECG measurements, thereby enabling better clinical interpretation.

However, the approach that seems the most promising in our research is the use of random forests. Unlike other machine learning methods, random forests provide the benefit of feature interpretability without sacrificing accuracy. Random forest models can identify which variables or features are most influential in predicting fetal health status, allowing for a better clinical understanding of the model's decision process. Our approach confirms the high precision of this method but also offers new insight by identifying the most significant variables that influence fetal well-being, thus contributing to the richness of clinical interpretability.

## 3. METHOD

Fetal heart rate is of great importance in assessing fetal health. During the prenatal and perinatal period, there is a high risk of unexplained fetal death. Fetal arrhythmia is one of the leading causes of death [14], [15]. Fetal arrhythmia is characterized by irregular heartbeat of the fetus and is generally classified when the heart rate exceeds 180 beats per minute or falls below 100 beats per minute [16]. Early prediction generally makes it possible to treat fetal arrhythmias using antiarrhythmic drugs [17]. This is why we offer a solution that aims to predict ECGs and use machine learning, allowing us to anticipate electrocardiogram patterns. This method relies on machine learning algorithms and models that analyze ECG data, extract key information, and generate accurate predictions on this data Figure 1.

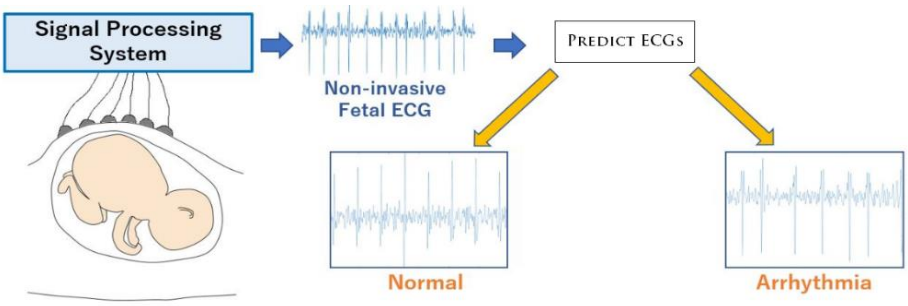


Figure 1. Solution description

Our approach is based on the use of random forest, an ensemble learning method that consists of creating multiple decision trees during the training phase and producing the class, which is the mode of the classes of the individual trees during the prediction phase [18], [19]. To predict the fetal ECG, we began these steps:

i) Data collection: fetal ECG samples are collected from our direct abdominal and fetal ECG database (ADFECGDB, available at https: //physionet.org/physiobank/database/ADFECGDB) as the data source. These data were collected from the Department of Obstetrics of the Silesian Medical University using the KOMPOREL system.
ii) Data preprocessing: ECG samples are then cleaned and normalized to ensure better data quality for training.
iii) Data division: the data is divided into training and test sets.
iv) Model training: the training set is used to train the random forest model.
v) Model evaluation: once the model is trained, it is evaluated on the test set to determine its accuracy, sensitivity, and specificity.

### 3.1. Random forest for foetal ECG prediction

Machine learning provides an avenue to automate and refine fetal ECG analysis. Using the random forest model, an efficient and robust algorithm, we can capture complex patterns in the data [20]. This model, in particular, is known for its ability to handle large amounts of data and extract meaningful features for accurate prediction. Previous work has explored the use of various algorithms, such as artificial neural networks ANN, SVM, for the analysis of fetal ECGs. Our approach stands out by using an optimized combination of these algorithms, significantly improving the prediction accuracy, as shown in Figure 2.

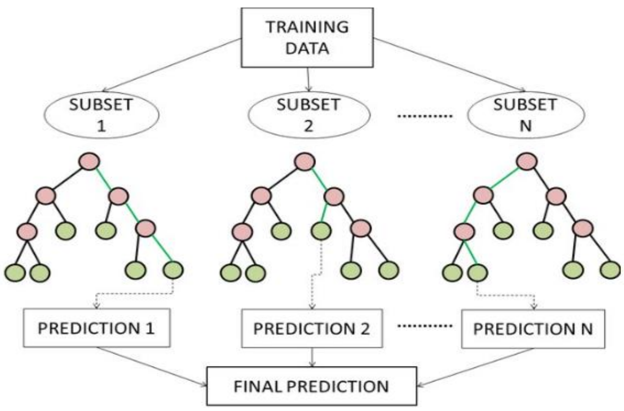


Figure 2. Random forest model

Random forest is an ensemble method that creates a forest of many decision trees, as Figure 2 shows during training, and returns the class average (regression) mode of the individual trees during prediction [21]. Here are the main steps in implementing our approach:

i) Sample selection: randomly choose a subset of data from the training set. This can be done with replacement, a technique called “bootstrapping.” Our approach used approximately 2/3 of the initial set to train each tree, with the remaining 1/3 used as an “out-of-bag” set to evaluate performance [22].
ii) Feature selection: at each node split in the tree, randomly select a subset of features without replacement. This subset is used to determine the optimal split at that specific node. Typically, our case is around m/3.
iii) Creation of trees: for each selected bootstrap sample, create a decision tree. The growth of this tree is generally not pruned, meaning it can grow as deep as needed. Repeat steps 1 and 2 to create each tree in the forest.
iv) Prediction: for a classification task: each tree “votes” for a class, and the class with the most votes is the final prediction of the forest. For a regression task: the average output of individual trees is the final prediction [23].

## 4. PERFORMANCE EVALUATION

After adopting the random forest method to predict fetal ECGs, we observed promising results, demonstrating our approach’s robustness and relevance in Figure 3. Our solution is believed to have displayed remarkable accuracy, outperforming several traditional machine learning methods. This high accuracy means that most fetal ECG predictions were consistent with the expected results Figure 3. Also, thanks to the ensemble nature of the random forest, our model showed notable resilience to overfitting. By aggregating multiple decision trees, our solution maintained robust generalization on unpublished data [24].

To evaluate the performance of our prediction model, we have to use several evaluation metrics depending on the nature of our problem. Here are some of the commonly used evaluation metrics for different problem types: mean squared error (MSE), ii) mean absolute error (MAE), and coefficient of determination ($R^2$). MSE: this is the average of the squares of the errors between the predicted values and the actual values. The lower the MSE, the better the model performance. MAE: this is the average of the absolute error values between the predicted and actual values. As with the MSE, a lower MAE indicates better performance [25]. $R^2$: it measures the proportion of variance in the dependent variable (your prediction) that the model explains. An $R^2$ close to 1 indicates that the model explains the variance of the data well. In our case, we used the $R^2$, which gave a rate of 0.9967261876157711 as shown in Figure 4.

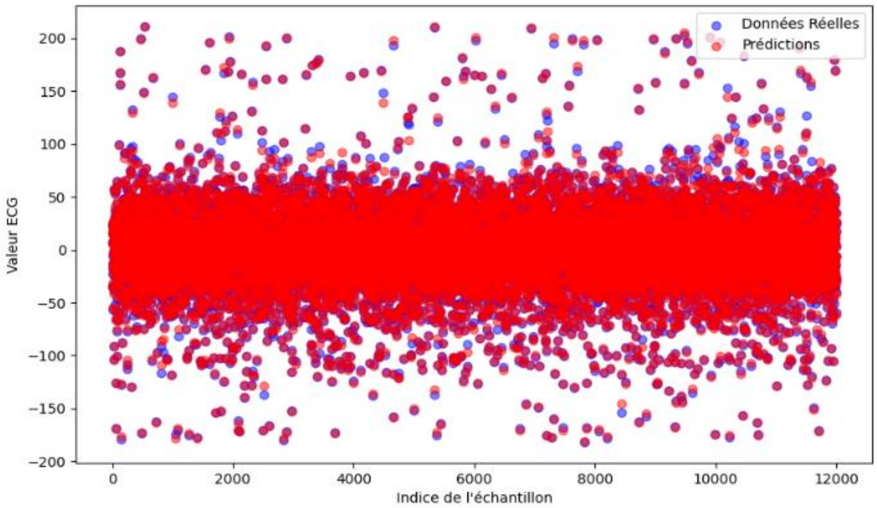


Figure 3. Comparison of predicted and actual data

```
Coefficient de Détermination (R²) : 0.9967261876157711
```

Figure 4. Random forest coefficient

An $R^2$ of 0.9967 indicates that the regression model has a very high ability to explain variation in the data. The model explains approximately 99.67% of the total variance in the dependent variable, suggesting that the model fits the data very well. This means that the model is highly predictive and can be considered very reliable in making predictions.

After implementing our random forest-based approach to predicting fetal ECGs, we carefully analyzed the results by comparing our predictions with the test set in Figure 5. Here is an overview of the major conclusions:

i) Strong correlation: an initial review of the data showed a strong correlation between the values predicted by our random forest-based solution and the actual data from the test set. This correlation suggests that our model successfully captured the essence of fetal ECG signals.
ii) Low margin of error: the deviation between the predicted values and the true values of the test set was generally small, attesting to the accuracy of our model. This is particularly notable given the complexity of fetal ECG signals.
iii) Quantitative evaluation: the accuracy rate obtained with our solution was impressive. Compared to other methods, random forest demonstrated notable superiority, reinforcing our methodological choice's relevance for this specific task.
iv) Outlier observations: although most predictions were in close agreement with the expected values, a few anomalies were observed. These occasional errors could be attributed to singularities in the data or to situations where the random forest could be challenged.
v) Fidelity of trends: by plotting predicted data against actual data, we found that our solution closely tracked the intrinsic trends of ECG signals, revealing its ability to detect and reproduce the subtle dynamics of these signals.
vi) Statistical insights: statistical properties, such as mean, variance, and median, of the predicted data matched remarkably well with those of the test data. This demonstrates the ability of random forest to generalize effectively from training data.

Our solution leveraging random forest for fetal ECG prediction has proven its value by delivering high-quality results in close alignment with the actual test set data. These performances reinforce our confidence in the chosen approach and pave the way for further research to refine the model further.

```
for prediction2 in y_test[:5]:
    print(prediction2)
```

TEST

```
[23.50035859]
[22.90034943]
[-13.20020142]
[6.20009461]
[23.00035096]
```

```
for prediction in y_pred_random_forest[:5]:
    print(prediction)
```

PREDICTION

```
23.660361028600004
19.098291419200002
-12.79519523890001
5.145078506899997
22.39734175510003
```

Figure 5. Comparison of predicted and actual data

## 5. COMPARISON WITH OTHER ALGORITHMS

In our quest to develop a robust solution for predicting fetal ECGs, we opted for random forest as our primary method as Figure 3. However, to contextualize the performance of our choice, it is essential to compare it with other recognized methods such as ANN, SVM and RNN in Figure 6. Here is a comparative analysis:

i) ANN: although ANNs can capture complex relationships, they require considerable data and fine-tuning to be effective. Compared to our solution, ANNs are more susceptible to overlearning, and their “black box” nature makes it challenging to understand the decisions taken.
ii) SVM: can be effective for some classification tasks, especially with low-dimensional data. SVMs have been found to be less efficient and less practical than random forest for large datasets like those of ECGs due to processing time.
iii) Recurrent neural networks (RNN): were designed to process sequences, which is relevant for ECGs. However, they often face challenges such as disappearing gradients, requiring more complex variants to work effectively. Comparatively, random forest offers robustness and efficiency without the complications associated with RNNs.

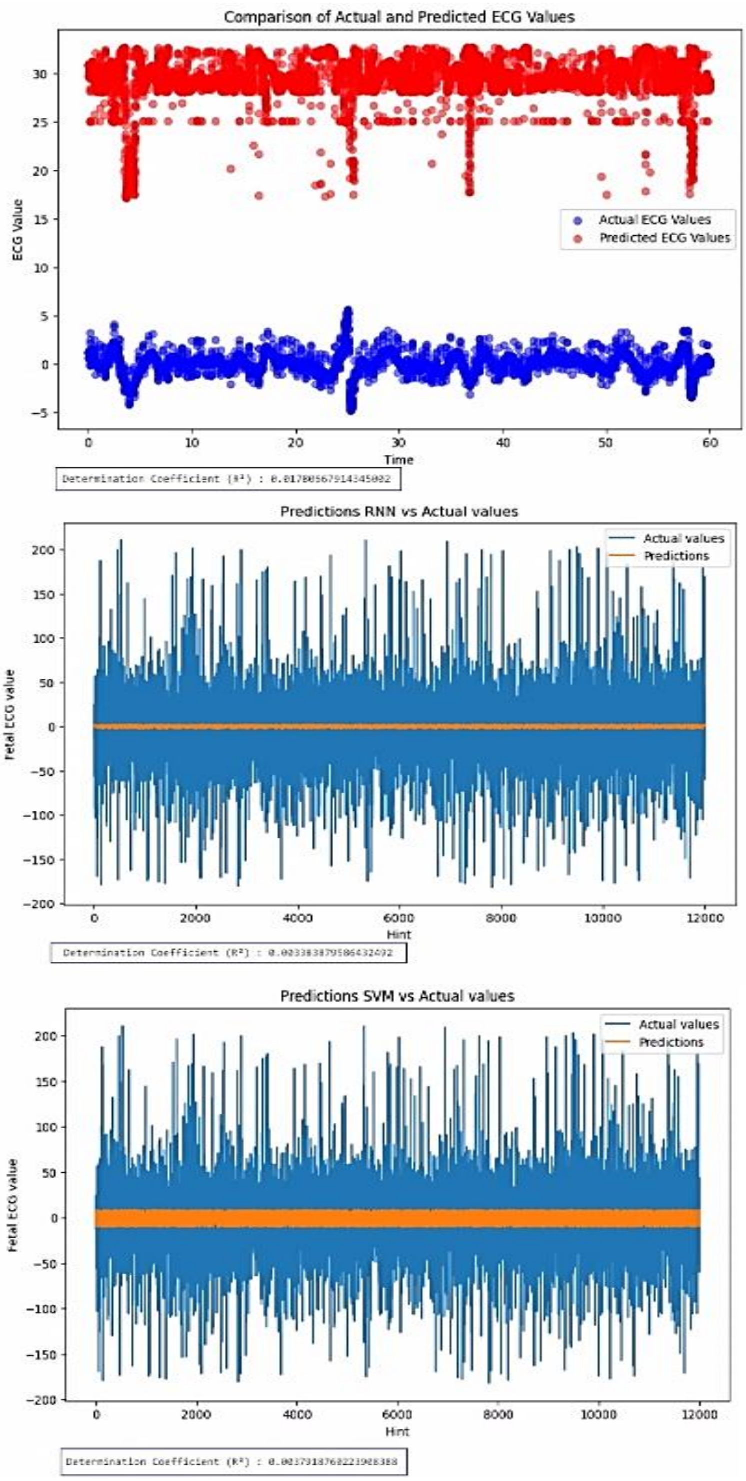


Figure 6. Comparison with other algorithms

Random forest demonstrated an exceptional ability to be robust to noisy data while accurately assessing feature importance. Combining multiple trees ensured excellent generalization, reducing the risk of overfitting, a common problem with ANNs. Furthermore, unlike SVM and RNN, random forest combines efficiency and interpretability, making the results more transparent. Although each method has its merits, our random forest-based solution was more effective and robust in predicting fetal ECGs than other methods. The ability of random forest to provide accurate predictions while being less susceptible to common problems associated with ANN, SVM, and RNN confirms the relevance of our choice.

## 6. CONCLUSION

Through this study, we explored the intricacies of fetal ECG prediction, an area of medicine that holds crucial importance for fetal health monitoring. Faced with a myriad of machine learning techniques, choosing an optimal solution is often complex. However, after our extensive research and multiple experiments, our random forest-based method stood out as a preeminent tool. It should be noted that although random forest was this study's flagship method, exploring other techniques was equally informative. It allowed us to understand each approach's intrinsic strengths and weaknesses, thus enriching our overall understanding of the field of fetal ECG prediction. In short, this study demonstrates the immense potential of machine learning in medicine. By combining medical expertise and data engineering, we demonstrated a powerful solution for fetal ECG prediction, contributing to ensuring fetuses' health and well-being. As the medical world continues to evolve, with the increasing integration of artificial intelligence technologies, we are confident that our work will serve as a cornerstone for other future innovations in this field.

## REFERENCES


[1] M. Moutaib, T. Ahajjam, M. Fattah, Y. Farhaoui, B. Aghoutane, and M. E. Bekkali, "Reduce the energy consumption of IoTs in the medical field," in *International Conference on Digital Technologies and Applications*, 2022, pp. 259–268, doi: 10.1007/978-3-031-02447-4_27.

[2] M. Moutaib, M. Fattah, and Y. Farhaoui, "Internet of things: energy consumption and data storage," *Procedia Computer Science*, vol. 175, pp. 609–614, 2020, doi: 10.1016/j.procs.2020.07.088.

[3] M. Moutaib, T. Ahajjam, M. Fattah, Y. Farhaoui, B. Aghoutane, and M. E. Bekkali, "Optimization of the energy consumption of connected objects," *International Journal of Interactive Mobile Technologies*, vol. 15, 2021, doi: 10.3991/IJIM.V15I24.26985.

[4] M. Moutaib, T. Ahajjam, M. Fattah, Y. Farhaoui, and B. Aghoutane, "Reduce the energy consumption of connected objects," in *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning*, 2021, pp. 86–91, doi: 10.5220/0010728900003101.

[5] M. Ahmad, S. Shabbir, R. A. Raza, M. Mazzara, S. Distefano, and A. M. Khan, "Artifacts of different dimension reduction methods on hybrid CNN feature hierarchy for hyperspectral image classification," *Optik*, vol. 246, p. 167757, 2021, doi: 10.1016/j.ijleo.2021.167757.

[6] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2020, doi: 10.1109/LGRS.2019.2918719.

[7] S. P. Singh, L. Wang, S. Gupta, B. Gulyas, and P. Padmanabhan, "Shallow 3D CNN for detecting acute brain hemorrhage from medical imaging sensors," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14290–14299, Jul. 2021, doi: 10.1109/JSEN.2020.3023471.

[8] M. E. Adoui, S. Drisis, and M. Benjelloun, "New explainable deep CNN design for classifying breast tumor response over neoadjuvant chemotherapy," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 19, no. 5, pp. 526–533, May 2022, doi: 10.2174/1573405618666220803124426.

[9] A. Kumar, H. Tomar, V. K. Mehla, R. Komaragiri, and M. Kumar, "Stationary wavelet transform based ECG signal denoising method," *ISA Transactions*, vol. 114, pp. 251–262, 2021, doi: 10.1016/j.isatra.2020.12.029.

[10] C. J. Smith *et al.*, "Understanding rapid adjustments to diverse forcing agents," *Geophysical Research Letters*, vol. 45, no. 21, Nov. 2018, doi: 10.1029/2018GL079826.

[11] J. S. Johnson *et al.*, "Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis," *Nature Communications*, vol. 10, no. 1, p. 5029, Nov. 2019, doi: 10.1038/s41467-019-13036-1.

[12] D. R. Williams, J. A. Lawrence, B. A. Davis, and C. Vu, "Understanding how discrimination can affect health," *Health Services Research*, vol. 54, no. S2, pp. 1374–1388, Dec. 2019, doi: 10.1111/1475-6773.13222.

[13] D. S. Lee *et al.*, "The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018," *Atmospheric Environment*, vol. 244, p. 117834, Jan. 2021, doi: 10.1016/j.atmosenv.2020.117834.

[14] M. Krapp, T. Kohl, J. M. Simpson, G. K. Sharland, A. Katalinic, and U. Gembruch, "Review of diagnosis, treatment, and outcome of fetal atrial flutter compared with supraventricular tachycardia," *Heart*, vol. 89, no. 8, pp. 913–917, 2003, doi: 10.1136/heart.89.8.913.

[15] A. Mansouri, A. E. Magri, I. E. Myasse, R. Lajouad, and N. Elaadouli, "Neurological morbidity after fetal supraventricular tachyarrhythmia," *Ultrasound in Obstetrics and Gynecology*, vol. 13, no. 1, 1999, doi: 10.1046/j.1469-0705.1999.13010043.x.

[16] J. F. Strasburger, "Fetal arrhythmias," *Progress in Pediatric Cardiology*, vol. 11, no. 1, pp. 1–17, May 2000, doi: 10.1016/S1058-9813(00)00031-X.

[17] A. Wacker-Gussmann, J. F. Strasburger, B. F. Cuneo, and R. T. Wakai, "Diagnosis and treatment of fetal arrhythmia," *American Journal of Perinatology*, vol. 31, no. 7, pp. 617–628, 2014, doi: 10.1055/s-0034-1372430.

[18] S. Sarafan *et al.*, "Investigation of methods to extract fetal electrocardiogram from the mother's abdominal signal in practical scenarios," *Technologies*, vol. 8, no. 2, 2020, doi: 10.3390/technologies8020033.

[19] Y. Zhang and S. Yu, "Single-lead noninvasive fetal ECG extraction by means of combining clustering and principal components analysis," *Medical and Biological Engineering and Computing*, vol. 58, no. 2, pp. 419–432, 2020, doi: 10.1007/s11517-019-02087-7.

[20] A. J. D. Krupa, S. Dhanalakshmi, and R. Kumar, "An improved parallel sub-filter adaptive noise canceler for the extraction of fetal ECG," *Biomedizinische Technik*, vol. 66, no. 5, pp. 503–514, 2021, doi: 10.1515/bmt-2020-0313.

[21] A. Rasti-Meymandi and A. Ghaffari, "AECG-DecompNet: abdominal ECG signal decomposition through deep-learning model," *Phys.Meas.* 2021, 42, 33706298.

[22] Y. Zhang and S. Yu, "Single-lead non-invasive fetal ECG extraction by means of combining clustering and principal components analysis," *Medical, Biological Engineering and Computing*, 2020, 58, 419–432.

[23] D. Daghouj, M. Fattah, S. Mazer, Y. Balboul, and M. El Bekkali, "UWB waveform for automotive short range radar," *International Journal on Engineering Applications (IREA)*, vol. 8, no. 4, pp. 158–164, Jul. 2020, doi: 10.15866/irea.v8i4.18997.

[24] A. Maroua and F. Mohammed, "Characterization of ultra wide band indoor propagation," in *2019 7th Mediterranean Congress of Telecommunications (CMT)*, Oct. 2019, pp. 1–4, doi: 10.1109/CMT.2019.8931367.

[25] M. J. Mollakazemi, F. Asadi, M. Tajnesaei, A. Ghaffari, Q. R. S. Fetal "Detection in noninvasive abdominal electrocardiograms using principal component analysis and discretewavelet transforms with signal quality estimation," *Journal of Biomedical Physics and Engineering*, vol. 11, pp. 197–204, 2021.

## BIOGRAPHIES OF AUTHORS

**Mohammed Moutaib** obtained his Masters in Big Data and Internet of Objects at the University of Hassan 2 Casablanca, Morocco, 2019. Ph.D. researcher from the University of Moulay Ismail (UMI), Meknes, Morocco. He teaches computer science at the middle school in Casablanca. He can be contacted at email: mohammed.93@live.fr.

**Mohammed Fattah** received his Ph.D. in Telecommunications and CEM at the University of Sidi Mohamed Ben Abdellah (USMBA) Fez, Morocco, in 2011. He is a professor in the Electrical Engineering Department of the School of Technology at the Moulay Ismail University (UMI), Meknes, Morocco, and he is responsible for the research team' Intelligent Systems, Networks and Telecommunications', IMAGE laboratory, UMI. He can be contacted at email: m.fattah@umi.ac.ma.

**Yousef Farhaoui** is a Professor at IDMS Team, Moulay Ismail University, Faculty of Sciences and Techniques, Morocco. Local Publishing and Research Coordinator, Cambridge International Academics in United Kingdom. He obtained his Ph.D. degree in computer security from Ibn Zohr University of Science. His research interests include learning, e-learning, computer security, big data analytics, and business intelligence. He can be contacted at email: y.farhaoui@fste.umi.ac.ma.

**Badraddine Aghoutane** obtained a Ph.D. in Computer Science from Sidi Mohamed Ben Abdellah University, Morocco, in 2011. He is a professor at the Department of Computer Sciences, Faculty of Science, Moulay Ismail University, Meknes, Morocco. He joined Moulay Ismail University (UMI) in 2011. His research interests are web semantics, recommender systems, IoT security, and big data. He can be contacted at email: b.aghoutane@fpe.umi.ac.ma.

**Moulhime El Bekkali** holder a doctorate in 1991 from the USTL University-Lille 1-France. He worked on printed antennas and their applications to microwave radar. Since 1992, he was a professor at the Graduate School of Technology, Fez (ESTF) and a member of the Transmission and Data Processing Laboratory (LTTI). In 1999, he received a second doctorate in electromagnetic compatibility from Sidi Mohamed Ben Abdellah University (USMBA). He is a professor at the National School of Applied Sciences (ENSAF) and a member of the LIASSE laboratory at Sidi Mohamed Ben Abdellah University. He can be contacted at email: moulhime.el.bekkali@gmail.com.