POLITECNICO DI MILANO
DEPARTMENT OF ELECTRONICS, INFORMATION AND BIOENGENEERING
DOCTORAL PROGRAM IN BIOENGINEERING

PHYSIOLOGY BASED MACHINE LEARNING AND DATA ANALYTICS FOR PERINATAL MONITORING

A novel framework for a comprehensive maternal, fetal, and neonatal profiling

Doctoral Dissertation of:
Nicolò Pini

Advisor:
Prof. **Maria G. Signorini**

Co-Advisor:
Prof. **William P. Fifer**

Tutor:
Prof. **Chiara Guazzoni**

The Chair of the Doctoral Program:
Prof. **Andrea Aliverti**

2017-2020 – XXXIII Cycle

Abstract

The perinatal period plays a crucial role in shaping the early stages of everyone's life. States of poor maternal, fetal, and neonatal wellbeing have profound repercussions with the potential for a lifelong impact. Throughout pregnancy, delivery, and postpartum mothers-to-be and their children are strongly interconnected one another resembling a dynamic system in rapid evolution. Nevertheless, a large portion of the investigational approaches in the perinatal field conceptualized pregnant women, fetuses, and babies as independent entities.

The purpose of this Ph.D. thesis is to inform on a novel characterization of the perinatal period. The objective is to provide a pool of interpretable data-driven models capable of highlighting the potential emergence of states of risk or adverse outcomes. The original methodological contribution centers on the utilization of tools encompassing machine learning, artificial intelligence, and advanced signal processing techniques towards data imputation, non-parametric clustering, prediction, and network physiology analysis. Additionally, the novelty of this Ph.D. thesis lies in the analytic framework which consists of integration of heterogenous data sources. Quantitative data such as fetal, neonatal, and maternal physiological signals are complemented with qualitative data such us maternal lifestyle reports. We utilized Random Forests and Support Vectors Machines for the in-utero detection of fetal growth restriction, K-Nearest Neighbor and finite mixture model were employed to impute and cluster maternal alcohol and tobacco consumption during pregnancy, network physiology for the characterization of neonatal cardiorespiratory regulation, and recursive partitioning to predict neurodevelopmental outcome in toddlers combining longitudinal data recorded throughout the perinatal period. As a result, we provided evidence for the close interrelationship among maternal, fetal, and neonatal wellbeing by investigating i) the effects of maternal lifestyle on fetal and neonatal physiology, ii) the repercussion of fetal growth restriction on neonatal cardiorespiratory regulation, and iii) the influence of perinatal health in shaping subsequent neurodevelopmental trajectories.

In conclusion, the quantitative framework proposed in this Ph.D. thesis is expected to substantially contribute toward promoting healthy pregnancy, safe childbirth, and reduce adverse outcome by building a sustainable network for perinatal health monitoring trained on heterogenous data fusion. The ultimate goal is to inform on monitoring solutions for risk assessment trained on dynamical and longitudinal indicators of perinatal health.

Summary

The definition of *perinatal* period covers pregnancy, delivery, and postpartum [1]. This lifespan interval plays a crucial role in shaping early stages of life and has profound repercussions that attain the potential to extend beyond the early stages of life. In recent years, research has extensively documented several associations between perinatal health and chronic diseases of adulthood [2], [3]. Diverse epidemiological and experimental studies have shown that the perinatal period represents a window of peculiar *sensibility* and *vulnerability* which is ultimately associated with lifelong trajectories which encompass both wellbeing and potentially pathological states. Early prevention, optimal perinatal nutrition, and specific follow-up measures are key factors in the early preservation of long-term health.

On the contrary, *fetal* and *maternal exposures* to substances of abuse, *poor nutrition*, *stress* and *anxiety* have both direct and indirect downstream effects on modifiable and non-modifiable gene expression patterns. As an example, chronic diseases of adulthood such as hypertension, diabetes, and obesity have early developmental origins in the perinatal period [3]. More specifically, perinatal disease, either maternal such as pre-eclampsia, gestational diabetes, or inflammatory disease, fetal and neonatal diseases such as *intrauterine growth restriction* and *preterm birth* are major contributors to altered programming, translated into an increased risk for chronic disease [3].

The illustrated framework is summarized in the so-called *developmental-origins hypothesis* [4]. This theory proposes that a potential increase risk for chronical states of disease in the fetus is trigged by maternal physical state. Fetal or perinatal responses may include modifications in metabolism, hormone production with the potential of persistent alterations in physiologic and metabolic homeostatic set points [4].

Striking examples of such are the perinatal conditions investigated in this Ph.D. thesis. The association between reduced fetal growth rate, small body size at birth, and a later risk of disease may reflect the long-term consequences of abnormal fetal adaptive responses. Specifically, replicated evidence suggests that that growth restriction and consequent low birthweight are associated with increased rates of coronary heart disease, stroke, type 2 diabetes mellitus, adiposity, the metabolic syndrome, and osteoporosis in adult life [4]. Abnormal growth in utero may be associated with increased allocation of nutrients during the critical period of development

and may then result in accelerated weight gain during childhood, which may contribute to a relatively greater risk of the above-mentioned pathological states. This translates to a continuous relation between birth weight and future risk not limited to the condition of extreme low weight.

Prematurity is an often-coexisting factor closely tight to *fetal growth restriction* and birthweight. Prematurity itself, almost independent of the degree of severity, has been associated with insulin resistance and glucose intolerance in adolescents, and it may ultimately result in elevated blood pressure into young adulthood. Moreover, preterm infants are more likely to experience morbidities and mortality compared to term infants [5], [6]. They are also more likely to develop neurological and developmental disorders independently associated with long-term consequences [3]. Taken together, these factors support evidence for the crucial role of perinatal period in shaping individuals' ability to face external stressor and possibly overcome the risk associated to pathological states.

The main purpose of this Ph.D. thesis was to propose an *innovative methodological framework* towards a comprehensive, longitudinal, and rigorous characterization of the diverse phases of the perinatal period as well as their mutual interrelationship. To achieve this goal, we proposed the utilization of an *ensemble of computational approaches* aimed to provide interpretable models capable of highlighting the emergence of alterations in diverse crucial pathways responsible for potential persistent alterations in physiologic and metabolic homeostatic set points throughout the perinatal period.

The employed tools encompassed *machine learning, artificial intelligence, and advanced signal processing techniques* towards prediction of pathological states in the fetal period, data imputation, non-parametric clustering of maternal substance exposure, and network physiology analysis and risk assessment in newborns. Additionally, the novelty of this Ph.D. thesis lies in the analytic framework which consists of the *integration of heterogenous data sources*. Specifically, quantitative data such as fetal, neonatal, and maternal physiological signals and extracted physiology based parameters are contextualized with qualitative data such us maternal lifestyle reports, precise and timely quantification of exposure. The described approach is in contrast with traditional and established univariate methodologies focused on the investigation of pregnancy, delivery, and postpartum by themselves. To overcome this present limitation, we proposed a novel and *rigorous contextualization* of the perinatal period aimed to describe this crucial developmental period as a whole by means of multisource information fusion and advanced monitoring. The

envisioned application of the designed framework is to enhance the current state of the art in the field of perinatal care and be ultimately incorporated as a tool to support and integrate clinical practice.

The presented research was conducted under the primary guidance of Professor Maria G. Signorini at Politecnico di Milano and in collaboration with Professor Giovanni Magenes at Università degli studi di Pavia, and William P. Fifer at Columbia University Irving Medical Center. Datasets for the analyses were used in collaboration with Dr. Marta Campanile at Università degli Studi di Napoli Federico II, and the PASS Research Network, Columbia University Irving Medical Center.

In Chapter 1, we introduced the first methodological approach of this Ph.D. thesis by proposing a novel multivariate technique toward the detection of a pathological fetal condition, namely *intrauterine growth restriction* (IUGR). This condition accounts for a considerable portion of the perinatal causes of increase morbidity and mortality, thus substantially impacting the perinatal period under several aspects. In contrast with the current available approaches, we propose to develop a set of models for the *in-utero detection of the pathology* instead of a at birth assessment. Moreover, we aimed to go beyond the traditional linear and nonlinear univariate approaches [7]–[9]. The innovative approach proposed in this Ph.D. thesis encompasses the design of a more comprehensive framework for the promptly identification of potential risk for these pathological states making use of data integration and multivariate machine learning approaches. In recent years, the field of fetal monitoring has seen a growing interest towards novel nonlinear techniques aimed to provide an enhanced understanding of the different regulatory mechanisms impinging the fetal heart rate variability. Nonetheless, the limitation of employing a single parameter for the identification of a complex and diverse pathology as IUGR has resulted in unsatisfactory performance. Starting from the preexisting knowledge coming from established *physiology based heart rate features*, we investigated different machine learning models for the prediction of both early and late IUGR pathology. Despite utilizing a similar ensemble of features in both scenarios, it appeared as clear than the same framework cannot successfully be employed. Specifically, we found *Random Forests* and *Support Vector Machines* to best perform in the context of early and late IUGR identification respectively. The need of machine learning model adaptation as a function of *gestational age* of the developing fetus reflects the rapidly evolving process of pregnancy itself. As for every dynamical system, boundary conditions as well as

relationship among variables are expected to change across time and thus models must evolve and/or adapt to account for such modifications. This concept of contextualizing is deeply investigated in Chapter 1. A clear example is the approach utilized to *remove the influence of gestational age on the physiology based heart rate features* which are subsequently fed into the design models. The use of robust linear regression to derive residuals employed in machine learning frameworks allows to contextualize the differences between healthy fetuses and IUGR independently of age of assessment, thus reflecting a true difference in autonomic nervous system regulatory profiles. An additional strategy of contextualization is to provide the model with the original set of physiology based heart rate features in conjunction with gestational age at assessment. If the first methodological approach was reported as mandatory in the context of early IUGR identification, for late detection both were found to be of successful implementation. A plausible explanation sees an association between the average difference between time of assessment for the two groups. Specifically, the larger the discrepancy in terms of gestational age weeks between the two classes, the more likely is the model to distinguish on the basis on the time of assessment rather than fetal condition. This becomes crucial as in clinical practice gestational age at examination may vary substantially depending on several non-modifiable factors. An additional contextualization expected to fill a considerable gap in the investigation of fetal physiology is represented by the *introduction of additional signals* into the scenario of investigation. The most favorable candidates are umbilical and middle cerebral artery Doppler fluxes. One major limitation for their integration with CTG recordings relates once more to the different assessment times associated with current clinical practice. At the same time, machine learning and artificial intelligence models fully trained and inspired to features extracted from CTG (and eventually fetal ECG) are aligned with the latest technological advancements on *wearable solutions for fetal monitoring*. In recent years, a wide variety of portal fetal monitors has been introduced into the market and ultimately designed and made available for the broader public. A common shared trait is the collection of almost exclusively fetal electrical activity fetal movement and/or maternal contractions in a limited subset of the devices. Thus, the approach proposed in this Ph.D. does not appear as unreasonable despite begin solely inspired to predictive potentiality of fetal heart rate variability parameters. Additionally, the advantage of a continuous monitoring in diverse scenarios of utilization may pave the way for a deeper understanding of *pregnancy evolution* in contrast to the current standard of care which encompass a rather intermittent approach. Lastly, the goal of a combined approach encompassing (pseudo)continuous

monitoring and artificial intelligence is the timely prediction of pathological states and ultimately anticipation of potential state of adverse outcome. From a methodological perspective, *waveform artificial intelligence* represents the next frontiers for automatic detection of fetal risk. If the use of time series as input to differently designed neural networks is progressively becoming the new standards in fields such as sleep staging [10], [11], and cardiac arrythmia detection [12], [13], the same does not apply in a straightforward manner to fetal monitoring. The most investigated fetal timeframe from the perspective of artificial intelligence techniques is the intrapartum. Nevertheless, a recent review on the topic [14] highlighted that the fact that artificial intelligence and computer analysis for the interpretation of fetal traces during labor does not improve neonatal outcomes. Authors speculate that one major aspect responsible for the reported unsatisfactory performance is the minimal agreement and reliability between experts and computer systems. Because of the known high inter-observer and intra-observer variability in the interpretation of FHR tracings, any artificial intelligence framework is ultimately expected to exhibit lack of optimal performance if trained with poorly labelled data.

The limitation of *poorly labelled data* represents a crucial limitation for the precise and reliable quantification of a diffused mediator responsible for many of the perinatal states of increased risk, such as maternal exposure throughout pregnancy [15]–[18]. Among the modifiable maternal conditions contributing to the insurgence of fetal pathological conditions, exposure to substance of abuse plays a substantial role. Specifically, the two foremost indirect effects induced by in-utero exposure to substance of abuse are the insurgence of IUGR condition and late prematurity (LPT) (to be discussed the in Chapter 3). A common etiology for the two pathologies relates to abnormalities in placenta development and nutrient supply, both trigged by substance consumption.

From a methodological perspective, the integration of poorly estimated exposure features in any given model is expected to *underestimate and undercount their contribution to the quantification of potential states of risk in the perinatal period*. To effectively contextualize fetal and neonatal wellbeing with information on maternal habits in the perinatal period, we proposed the utilization of *advanced data-driven computational methodologies*. The rationale behind the solutions discussed in Chapter 2 is consequent of the lack of rigorous approaches for the quantification of longitudinal exposure to alcohol and tobacco. The paucity of approaches has resulted in *inconclusive public health guidelines*. Consequently, the common shared trait among the different public health indications advises against consuming substance at any magnitude and

frequency, thus a safe level of alcohol or tobacco use during pregnancy has not yet been determined. The main reason behind such is the absence of validated and reliable methodologies for the quantification of missing information on substance intake throughout pregnancy, thus in most occasion quantification of exposure reduces to a binary variable, any or none. The *imputation of longitudinal data*, and more in general missing information is a complex and *multiparametric problem*. The most common strategy assumes that each missing value in a given dataset is equally likely to be missing [19]. The former assumption is often unmet in the context of exposure collection throughout pregnancy as pregnant women might only be queried at specific time points which are likely to overlap with clinical appointments. As a consequence, data are usually missing by design and thus leading to biased estimates of substance consumptions which may ultimately confound methodologies aimed to investigate the downstream effects [20]–[23].

The novel adaption of the KNN algorithm proposed in Chapter 2 exhibited satisfactory performance for the imputation of exposure data not missing at random bur rather by design. KNN is a non-parametric machine learning algorithm. The nature of the derived imputation originates from neighbor-based methods where the imputed value is either measured for the neighbor or the average for multiple neighbors. In a nutshell, samples with similar features are expected to result in similar output values. This assumption applies to the dataset analyzed in this report and it is consistent with previous work describing longitudinal trajectories of exposure during pregnancy [24], [25]. Moreover, the absence of constrain to fit the available data within a given parametric distribution allowed to decrease variance and potential bias on the derived estimates. The remarkable design obtained with the presented analysis allowed to perform imputation on two different populations by means of the same framework. The obtained results support the previously described homogeneity of the patterns of exposure across sites. Despite the satisfactory performance obtained, further work is needed in terms of validation of the KNN framework. Specifically, validation and replication are two fundamental aspects to be taken into consideration. The former must deal with the limitation of testing the framework performance on participants with complete data, thus on an implicit selected subset of individuals whose patterns of consumption are expected to be similar to a certain degree. Additionally, this selection bias might underrepresent the voracity and variety of information contained the dataset. One potential solution is to employ simulated data to generate an independent test set. Lastly, replication of this approach considering different population profiles and type of exposure is imperative. Techniques of transfer learning are expected to bridge the gap between parameters learnt in this investigation and

application to preexisting databases [26]. A family of methodologies expected to benefit from a more precise characterization of maternal exposure to substances of abuse is *clustering*. The need to group participants on the basis of their alcohol and/or tobacco consumption aims to derive interpretable profiles of exposure. This would also inform methodologies investigating the downstream effects of exposure as the ones to be summarized in the following. Furthermore, from a methodological perspective, clustering can be interpreted as methodology for data dimensionality reduction, thus reducing computational burden of subsequent machine learning and artificial intelligence approaches. In this investigation we provided evidence for the successful *application of finite mixture models* [27], nonetheless methodological advancements are required towards novel techniques to *longitudinal and heterogeneous data clustering and contextualization* [28], [29]. Longitudinal data consist of a given variable measured repeatedly over time. The traditional concept of partitioning implies that, individuals within the same cluster are expected to be similar one to each other, regardless of the adopted definition of group similarity. In the majority of the currently available approaches, two individuals are considered part of a given cluster when they present close trajectories at each time point [28], [29]. These approaches only take into account local similarities but ignore the general shapes of the trajectories. As an example, two identical but shifted in time series of feature are considered different and could potentially be assigned to distinct clusters. As a consequence, the mean of a derived group does not necessarily inform on the shape of trajectories, especially in scenarios where the progress of a phenomenon may be the most meaningful information to capture [28], [29]. These same considerations apply to the context of pregnancy. Lastly, given that longitudinal studies may deal with multiple different features, the issue of presenting attributes at different scales needs to be taken into account. The widely adopted solution is to *normalize the data*. Nevertheless, despite decreasing the potential bias of partitioning, normalization also reduces cluster interpretability and does not contribute to feature contextualization.

Feature contextualization plays a central role in the investigation presented in this Ph.D. thesis. As previously discussed, pregnancy, delivery, and early childhood may only be interpreted in the each other's context. Moreover, the boundaries between fetal physiological and pathological states throughout the perinatal period are labile, to the extent of defining many fetal diseases as dynamical. Dynamical disease is a concept introduced by Glass to indicate a disease that arise in a control system when its parameters assumes a specific configuration which is associated with an

abnormal dynamic [30]. These abnormal dynamics would correspond to *bifurcations in the relevant equations describing the physiological system*. Physiological regulations have been explained using the concept of homeostasis, which can be summarized as the ability for a given vital variable to maintain stability by constantly adjusting its deviation from a set point. The process of achieving this stability is called allostasis and it accounts for the ability to adapt successfully to the external demands. In this allostatic state, the spatiotemporal complexity of the control systems may give rise to multiscale complexity in the state variables. While this dynamical adaptation is essential to make physiological regulations, hyper or hypo activation of this state may bring a system to the critical transition point where the emergence of deviations could be early signs of disease onset and/or exacerbation. In the context of complex diseases identification, the discussed notion of transition point between a normal and a disease state is often nontrivial since the transition state may show little apparent change and the path to condition deterioration may vary greatly among individuals [30].

The discussed theory does apply and have been previously used to describe *the condition of prematurity* discussed in Chapter 3. The direct effect of exposure independent or in conjunction with abnormalities in fetal growth and development (indirect effects) may play the role of factors able to elicit a modification of equilibrium set points and thus homeostasis. Even minimal perturbation of physiological equilibria has profound repercussion on a cascade of subsequent developmental aspects. A striking example is the peculiar condition of *late prematurity* investigated in Chapter 3. Late prematurity is often classified of subclinical relevance given the gestational age at birth of babies in this group is at most two weeks smaller than the clinical cutoff of 37 weeks. Nonetheless, their risk of exposure to mortality and morbidities is significantly higher compared to terms babies. This marginal anticipation of physiological time of birth is potentially responsible for profound downstream effects as shown in the research presented in Chapter 3. Interestingly, such modification appears to impinge the mechanisms devoted to the cardiorespiratory control rather than the two system independently, namely cardiac and respiratory. The differences among late, early, and full term were unveiled by the utilization of methodologically advanced signal processing techniques capable of modelling the interaction among subsystems. Specifically, *Transfer Entropy* revealed a pronounced immaturity of cardiorespiratory interaction in both directionalities of interaction in the late preterm group. This intrinsic vulnerability has to the potential to increase the risk of adverse outcome in this population. Specifically, borrowing the previously described notion of dynamical disease, findings support the

hypothesis of different set points in the cardiorespiratory system of late preterm for which a given combination of state variables may elicit unstable mode of functioning and signaling. On the opposite, the same state variables are expected to be associated with an equilibrium state in healthy term infants. The reported results align with the recently proposed concept of *network physiology* [31]. This framework states that the behavior of one physiological network may affect the dynamics of all the others interconnect to the former. The cardiorespiratory regulation is an example of a hierarchical structure associated with diverse dynamics as a function of different physiological states.

The contribution of perinatal contextualization is extensively described in the second part of Chapter 3. Specifically, we provide evidence for the use of the previously derived exposure clusters in relationship with neonatal physiology for the prediction of potential adverse outcome in early childhood. The combination of *heterogenous data sources weighted and updated* based on longitudinal fetal and neonatal physiology, maternal conditions, and diverse environmental conditions is the envisioned approach for a multimodal monitoring framework able to promptly anticipate adverse conditions in the perinatal period. A first attempt towards a comprehensive and longitudinal risk profiling was achieve by the *recursive partition* model proposed in Chapter 3. A combination of features collected during pregnancy and at birth such as neonatal EEG, fetal, neonatal, and maternal chart abstractions, home environment, and in-utero substance use was able to successfully predict neurodevelopment outcome data collected at 24-37 years of age. This finding strongly aligns with the developmental-origins hypothesis, supporting the evidence that perinatal experiences and exposure have a profound repercussion on the early stages of life. At the same time, it must be acknowledged that the potential influence of postnatal modifiable factors has not been taken into considerations. Future studies will expand the applicability and extensively test the validity of the proposed model in terms of sample size, employed features, and cross validation techniques.

Taken the illustrated results together, *the framework presented in this Ph.D. thesis represents a first attempt toward a more comprehensive and cross-sectional conceptualization of the diverse aspects characterizing the perinatal period*. Starting from the developmental-origins hypothesis we proposed a solid set of computational tools inspired to *artificial intelligence, machine learning, and advanced signal processing* for a quantitative and rigorous profiling of potential states of increased risk or adverse outcome. The recent advancements in machine learning

and artificial intelligence methodologies offer the potential to start designing novel approaches to patient care which are expected to be equal or ultimately better than the current standard adopted in clinical practice. Currently, the main utilization of artificial intelligence in healthcare is to build innovative framework for *data collection and storage* [32], [33]. Longitudinal clinical data, such as hospital records for all patients with a given condition, are similarly heterogenous in data type, time scale, sampling rate, and reason for collection. Almost exclusively, electronic health record data are documented without consideration of the development of algorithms and/or with minimal vision regarding the subsequent potential set of analyses. As a consequence, the data collection and retrieve processes need to face several challenges. Firstly, incomplete or missing data need to be handled with caution as meaningful results can only be obtained when training algorithm on high quality and high-fidelity data. *Robust inference and imputation* can only be performed on large and representative datasets. A possible short-term alternative strategy is to employ synthetic data. However, models trained on synthetic data might not be as accurate as those trained on clinical data. Secondly, if outcomes are predicted on the basis of quantitative features, problems can arise when the measurements change considerably over time as it happens in the context of pregnancy. Other examples are the potential of translating findings obtain a given clinical setting to other environments, e.g., a model trained on data from an urban hospital might not be able to predict the same outcome in a rural setting. Finally, data labelling plays a crucial role toward improving model performance. Categorical outcome values are assigned by experts, but the variability among coders may impact in a non-negligible way the model prediction capabilities. Uncertainty can potentially be addressed by using generative models and unsupervised clustering to separate populations into underlying subtypes.

In conclusion, the quantitative framework proposed in this Ph.D. thesis paves the way for a methodological and rigorous sustainable network for perinatal health monitoring trained on heterogenous data fusion. The envisioned application is to promote healthy pregnancy, safe childbirth, and reduce adverse outcome by informing monitoring solutions for risk assessment with novel dynamical indicators of perinatal health.

Table of Contents

List of Abbreviations

ABP	Arterial Blood Pressure
AC	Acceleration Capacity
ANS	Autonomic Nervous System
ApEn	Approximate Entropy
APRS	Acceleration Phase Rectified Slope
AS	Active Sleep
BIC	Bayesian Information Criterion
BPM	Beats per minute
CNS	Central Nervous System
CTG	Cardiotocography
CV	Cross Validation
DC	Deceleration capacity
DPRS	Deceleration Phase Rectified Slope
DV	Ductus Venosus
ECG	Electrocardiography
EEG	Electroencephalogram
EM	Expectation Maximization
ET	Early Term Infant
FAS	Fetal Alcohol Syndrome
FASD	Fetal Alcohol Spectrum Disorders
FHR	Fetal Heart Rate
FIML	Full Information Maximum Likelihood
FT	Full Term Infant
GA	Gestational Age
GA_CTG	GA at CTG examination
HF	High Frequency Band
HoL	Hours of life at time of testing
HRV	Heart Rate Variability
IBI	Inter Breath Interval
II	Interval Index
IUGR	Intrauterine Growth Restriction
KNN	K-Nearest Neighbor
LF	Low Frequency band
LMP	Last Menstrual Period
LPT	Late Preterm Infant
LTI	Long Term Irregularity
LZC	Lempel Ziv Complexity
MCA	Middle Cerebral Artery
MF	Movement Frequency Band
MI	Multiple Imputation
MoD	Mode of Delivery (MoD)
MRI	Magnetic Resonance Imaging
NICU	Neonatal Intensive Care Unit

NP	Northern Plains, USA
NP	Network Physiology
NST	Nonstress Test
PAE	prenatal alcohol exposure
PASS	Prenatal Alcohol and SIDS and Stillbirth Network
PCA	Postconceptional Age
PRSA	Phase Rectified Signal Averaging
PSD	Power Spectral Density
PTE	prenatal tobacco exposure
QS	Quiet Sleep
QSE	Quadratic Sample Entropy
RBF-SVM	Radial Basis Function SVM
RBF-SVM-RFE	RBF-SVM Recursive Feature Elimination
RESP	Respiration Signal
RF	Random Forests
RLR	Robust Linear Regression
RMSSD	Root Mean Square of Successive Differences between heartbeats
RR	Series of the R-to-R intervals
RSA	Respiratory Sinus Arrhythmia
SA	Cape Town, South Africa.
SDNN	Standard Deviation of Successive Differences between heartbeats
SGA	Small for Gestational Age
SIDS	Sudden Infant Death Syndrome
STV	Short Term Variability
SVM	Support Vector Machines
TE	Transfer Entropy measure
TLFB	Timeline Followback Method
UA	Umbilical Artery

Introduction

The perinatal period plays a crucial role in shaping early stages of life and has profound repercussions that attain the potential to extend beyond early childhood. In this timeframe, mothers-to-be and their children are strongly connected one another resembling a dynamic system in rapid evolution. This intuitive and common sense evidence is supported by recent research which extensively documented several associations between perinatal health and chronic diseases of adulthood [2], [3]. The illustrated framework is summarized in the so-called developmental-origins hypothesis [4]. This theory proposes that a potential increase risk for chronical states of disease in the fetus and newborn is trigged by maternal physical state. Fetal and neonatal responses may include modifications and alterations in physiologic and metabolic homeostatic set points [4]. Despite the indisputable tight relationship among pregnancy, delivery, and postpartum periods, the current practice tends to analyze them as individual entities. The same applies to pregnant women, fetuses, and newborns which as often approached with lack of contextualizing of one to another.

Starting from the above described limitations, in this Ph.D. thesis we designed an innovative methodological framework towards a comprehensive, longitudinal, and rigorous characterization of the diverse phases of the perinatal period as well as we modeled the interrelationship among the diverse involved subjects. Moreover, we proposed a novel and rigorous contextualization of the perinatal period aimed to describe this crucial developmental period as a whole by means of multisource information fusion and advanced monitoring. To reach this goal, we utilized an ensemble of computational approaches aimed to provide interpretable models to highlighting the emergence of potential modification in diverse crucial pathways responsible for potential persistent alterations in physiologic and metabolic homeostatic set points throughout the perinatal period. The employed tools encompassed machine learning, artificial intelligence, and advanced signal processing techniques towards prediction of pathological states in the fetal period, data imputation, non-parametric clustering of maternal substance exposure, network physiology analysis and risk assessment in newborns. Furthermore, the novelty of this Ph.D. thesis lies in its analytic framework which consists of the integration of heterogenous data sources. Specifically, quantitative data such as fetal, neonatal, and maternal physiological signals and extracted physiology based parameters are contextualized with qualitative data such us

maternal lifestyle reports, and precise and timely quantification of exposure. The envisioned application of the designed framework is to enhance the current state of the art in the field of perinatal care and be ultimately incorporated as a tool to support and integrate clinical practice. A roadmap of the machine learning and artificial intelligence model ensemble is shown in the figure included in this introduction.

We started the investigation presented in this Ph.D. thesis discussing the pregnancy period. Firstly, we focused on the fetus and specifically on a pathological fetal condition accounting for a considerable portion of the perinatal causes of increase morbidity and mortality, namely intrauterine growth restriction (IUGR). A prompt and timely diagnosis of IUGR condition is key for pregnancy management. From the current clinical practice standpoint, this condition can only be suspected in the fetal period and ultimately diagnosed at birth. In contrast with the current available approaches, we propose to develop a set of models for the in-utero detection of the pathology in contrast with an at birth assessment. In details, the innovative approach encompassed the design of a more comprehensive framework for the promptly identification of potential risk of fetal growth restriction utilizing data integration and multivariate machine learning approaches. We designed a model inspired to the existing literature on advanced signal processing methodologies to inform on fetal wellbeing in the pregnancy period. The former was achieved by complementing the preexisting knowledge of physiology based heart rate features with maternal information, whereas the latter by designing different models as a function of pregnancy progression. This approach comes from the evidence that gestation can be modelled as a dynamical system whose boundary conditions as well as relationship among variables are expected to change across time, thus any framework would be required to evolve and/or adapt to such modifications. The methodological aspect of contextualization described in this Ph.D. thesis allowed to create models able to account changes in fetal autonomic regulation throughout pregnancy yet producing results independent of the time of fetal assessment. This technical solution promoted the generalization of the framework and its insensitivity to potential differences in time of assessment between healthy and pathological fetuses.

The subsequent investigated aspect of pregnancy encompassed an unsupervised and data-driven framework for the imputation and clustering of maternal exposure data. The limitation of

imprecise and missing data on maternal substance use/abuse in pregnancy represents a crucial limitation for the quantification of their effects on perinatal states of increased risk [15]–[18]. As an example, among the modifiable maternal conditions contributing to the insurgence of fetal pathological conditions, exposure to alcohol and tobacco plays a substantial role. Specifically, the two foremost perinatal pathological conditions induced by in-utero exposure are the insurgence of IUGR condition and late prematurity. From a methodological perspective, the integration of poorly estimated maternal exposure features in any given model is expected to underestimate and undercount their contribution to the quantification of potential states of risk in the perinatal period. To effectively contextualize fetal and neonatal wellbeing with information on maternal habits in the perinatal period, we proposed the utilization of advanced data-driven computational methodologies. Specifically, we addressed the design of a technique for the imputation of missing data on alcohol and tobacco consumption. The imputation of longitudinal data, and more in general missing information is a complex and multiparametric problem. The most common strategy assumes that each missing value in a given dataset is equally likely to be missing [19]. The innovative approach proposed in this Ph.D. thesis consists of the utilization of a KNN-based imputation. The absence of constrain to fit the available data within a given parametric distribution allowed to decrease variance and potential bias on the derived estimates. A family of methodologies expected to benefit from a more precise characterization of maternal exposure to substances of abuse is clustering. Furthermore, from a methodological perspective, clustering can be interpreted as methodology for data dimensionality reduction, thus reducing computational burden of subsequent machine learning and artificial intelligence approaches.

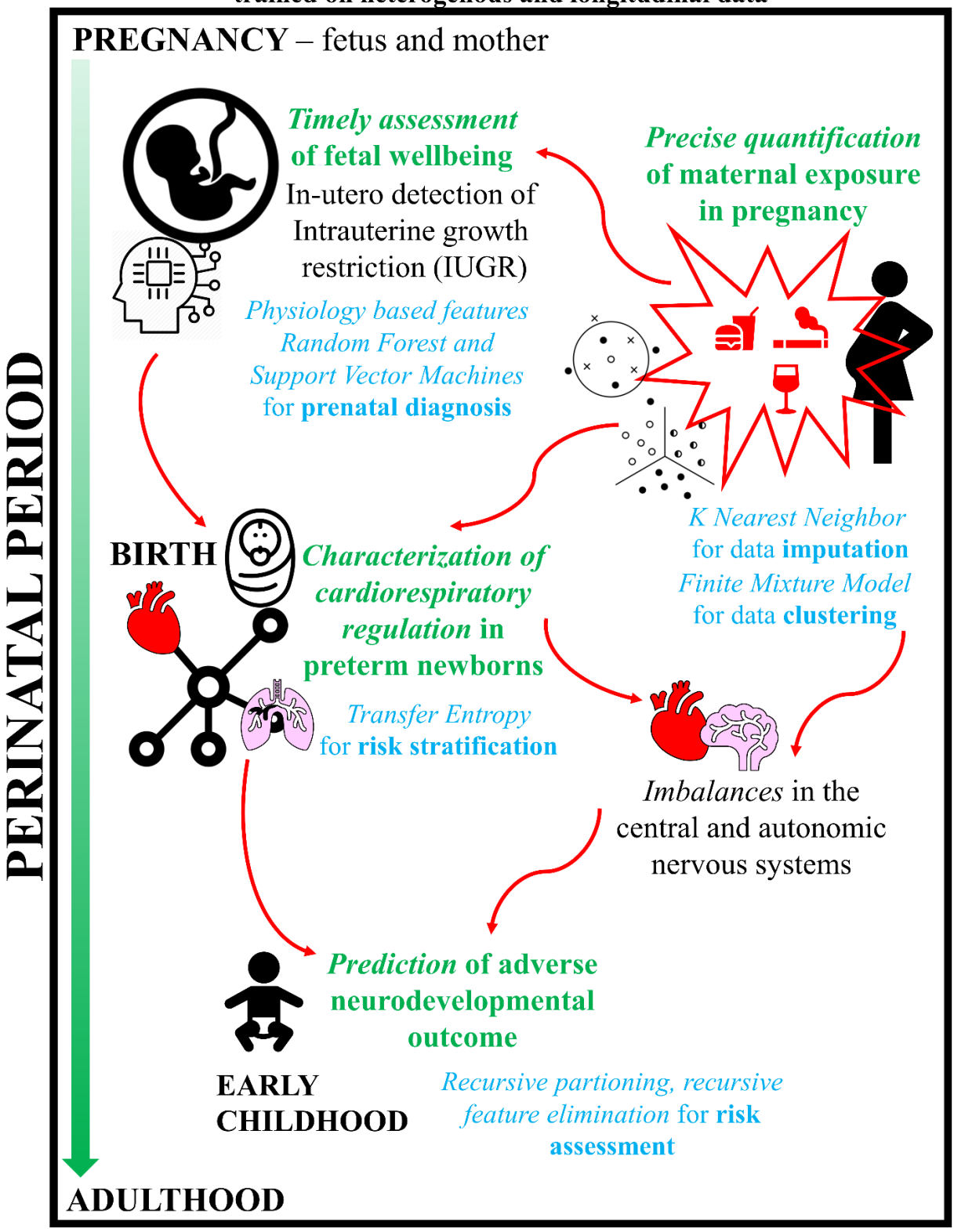
Feature contextualization plays a central role in the investigation presented in this Ph.D. thesis. As previously discussed, pregnancy, delivery, and early childhood may only be interpreted in the each other's context. Moreover, the boundaries between fetal physiological and pathological states throughout the perinatal period are labile, to the extent of defining many fetal diseases as dynamical. This translate into a continuum between the fetal, neonatal, and early childhood period. A contextualized example of such is the combination of the effect of exposure independent or in conjunction with abnormalities in fetal growth and development (indirect effects) to play the role of factors able to elicit a modification of equilibrium set points and thus homeostasis. This attains the potential to elicit the condition of late prematurity. A graphical representation of this concept

is illustrated in the roadmap at the end of this chapter. Late prematurity is often classified of subclinical relevance given the gestational age at birth of babies in this group is at most two weeks smaller than the clinical cutoff of 37 weeks. Nonetheless, their risk of exposure to mortality and morbidities is significantly higher compared to terms babies. We proposed to build a signal processing framework able to perform risk stratification making use of advanced signal processing methodologies. The differences among late, early, and full term were unveiled by the utilization of techniques capable of modelling the interaction among subsystems. Specifically, Transfer Entropy revealed a pronounced immaturity of cardiorespiratory interaction in both directionalities of interaction in the late preterm group. This intrinsic vulnerability has the potential to increase the risk of adverse outcome in this population.

Lastly, to provide evidence for the sustainability of the proposed ensemble of machine learning and artificial intelligence models we proposed an approach for the prediction of neurodevelopmental outcome in early childhood. The combination of heterogenous data sources weighted and updated based on longitudinal fetal and neonatal physiology, maternal conditions, and diverse environmental conditions is the envisioned approach for a multimodal monitoring framework able to promptly anticipate adverse conditions in the perinatal period. The combination of features collected during pregnancy and at birth such as neonatal EEG, fetal, neonatal, and maternal chart abstractions, home environment, and in-utero substance use was able to successfully predict neurodevelopment outcome data collected at 24-37 years of age. This finding strongly aligns with the developmental-origins hypothesis, supporting the evidence that perinatal experiences and exposure have a profound repercussion on the early stages of life.

In conclusion, the quantitative framework proposed in this Ph.D. thesis paves the way for a methodological and rigorous sustainable network for perinatal health monitoring trained on heterogenous longitudinal datafusion. The envisioned application is to promote healthy pregnancy, safe childbirth, and reduce adverse outcome by informing monitoring solutions for risk assessment with novel dynamical indicators of perinatal health.

An ensemble of machine learning and artificial intelligence models
toward a sustainable network for perinatal health monitoring
trained on heterogenous and longitudinal data



Chapter 1

One of the most common pathological condition resulting from abnormal placental development is Intrauterine Growth Restriction (IUGR) [34]. IUGR is a fetal condition defined as the abnormal rate of fetal growth. The pathology is a documented cause of fetal and neonatal morbidity and mortality. This pathological fetal condition is characterized by different timing of insurgence as well as diverse associated profiles of alteration of doppler parameters [35]–[37]. From a clinical prospective, diagnosis is confirmed at birth and may only be suspected during in utero. The crucial aspect to achieve an optimal or suboptimal pregnancy management is the identification of the fetuses at greatest risk for adverse outcome. Pregnancies complicated by the IUGR condition commonly evolve from a preclinical phase to clinically apparent growth delay and may progress to fetal deterioration. Although placental pathology changes over the continuum of gestational age, the classification of early-onset and the late-onset is widely accepted [37]

The early-onset (<32 weeks GA) represents 20–30% of all IUGRs and it is associated with decreased umbilical artery (UA) end-diastolic velocity and severe placental insufficiency. Additionally, the pathophysiology of early-onset comprises of a substantial reduction of the vascular area of the chorionic villi, resulting in severe placental insufficiency and chronic fetal hypoxia [38], [39]. Conversely, the late-onset IUGR (≥ 32 weeks GA) represent the 70–80% of IUGR cases. It is frequently associated with increased incidence of perfusion abnormalities and UA index elevation, or a decrease in the cerebroplacental Doppler ratio [37], [40]. Nonetheless, a large portion of the IUGR cases (regardless of the onset) do not exhibit marked abnormalities in the Doppler flow parameters, thus the diagnose process is characterized by a considerable portion of false negative cases [39]. For these reasons, it is imperative to design innovative and accurate integrated fetal testing toward a prompt and precise detection of significant variations in the clinical evolution of IUGR.

In recent years, the advancement in cardiotocography (CTG) analysis allowed to a progressive shift from pure visual observation of the traces to its computerized version [41]. This led to dissemination and integration of novel methodologies for the generation and interpretation of quantitative parameters potentially associated with diverse fetal conditions [42], [43]. As an example, fetal heart rate variability (HRV) extracted from a CTG trace acquired during a reactive nonstress test (NST) has been shown to virtually excludes hypoxemia. Nevertheless, in the context

of a nonreactive NST, HRV parameters have not been univocally associated with a restricted range of pH variability. As a consequence, current nonreactive NSTs do not provide conclusive understanding of fetal hypoxic condition. This consideration becomes extremely relevant for the diagnosis of both early and late IUGR conditions in the context of physiological Doppler velocities and derived parameters.

Studies investigating the predictive power of CTG-derived parameters have initially focused on traditional time and frequency domain features [43], [44]. Nonetheless, the traditional time domain techniques have been reported to be associated with a limited predictive power, and low sensitivity and specificity. This limitation has to be attributed to their inability to describe the complex linear and nonlinear dynamics of fetal HRV rather than the CTG technique by itself, although intrinsically less precise than abdominal ECG [19]. Further advancements in the characterization of fetal HRV dynamics saw the introduction of nonlinear parameters as potentially applicable [43], [44]. The IUGR condition has been reported to be associated with high estimates of Lempel-Ziv complexity [45]. At the same time, the main limitation of nonlinear domain parameters is their interpretability. As a consequence, their application in the clinical practice remains limited. The most recent advancement of CTG analysis for the purpose of IUGR diagnosis encompasses the use of Phase Rectified Signal Averaging (PRSA) technique. Results obtained by the analysis of features extracted from the PRSA curve indicate that IUGR fetuses have a lower cardiac acceleratory and deceleratory competence compared to those with an appropriate growth for gestational age [46]–[48]. Nonetheless, the reported studies focus primarily on the early phenotype with a limited investigation of the late onset IUGR cases.

Based on the above described findings, the added value of advanced signal processing tools appears as crucial to support clinical diagnostic power. At the same time, a very limited portion of the literature has evolved towards efficient and robust methodologies able to combine features from different domains. Nowadays, it appears clear that a single index cannot be descriptive of all pathophysiological processes taking place in the pregnancy period, thus supporting the need for a multivariate and more comprehensive analysis of FHR. To achieve this goal, we provided validation of an accurate artificially intelligence framework for the diagnosis of IUGR condition in the antepartum period considering both the early and the late phenotypes. The employed physiology based heart rate features constitute an interpretable link between the machine learning

results and the quantitative estimators of fetal wellbeing. Moreover, a purely CTG-based framework is expected to benefit from the large availability of novel wearable devices able to record FHR by means of noninvasive and passive solutions. The ultimate goal is to design a remote cloud platform to monitoring fetal wellbeing and thus prevent the insurgence of pathological states.

Chapter 1 is based primarily on the work illustrated in the [9], [49], [50]. Additional reports supporting the methodologies utilized in [9], [49], [50] can be found in [51]–[53].

Keywords – **Early and Late Intrauterine Growth Restriction, Machine Learning and Statistical Models, Fetal Heart Rate monitoring, Predictive analytics, Physiology-based features**

Integrating Machine Learning Techniques and Physiology Based Heart Rate Features for Antepartum Fetal Monitoring

Nowadays, antepartum fetal monitoring is a routine methodology adopted in clinical practice to assess fetal wellbeing throughout pregnancy, namely in the context of pathological fetal state identification [54], [55]. The most used technique consists in recording the Fetal Heart Rate (FHR) by means of the CardioTocography (CTG) [56]. The rationale for its utilization relies on the fact that it has been extensively shown how FHR changes can anticipate and/or even predict fetal distress as well as adverse conditions before the insurgence of any other symptom [57].
CTG analysis has been progressively shifting from pure visual observation of the traces to its computerized version [41], which consists of extracting various quantitative parameters associated with fetal conditions [42], [43]. Morphological [58], frequency [43], [59], and nonlinear/complexity indices [45], [60]–[62] are usually thought to summarize the various pathophysiological aspects of FHR. Despite the large availability of FHR quantitative indicators, a very limited portion of fetal-related literature addresses the investigation of fetal surveillance by means of multivariate approaches. If this latter consideration was to be attributed to scarce data availability in the past, recent years have seen the endless growth of data generated during patients' care path [32]. Additionally, the technological advancements in parallel with novel parameters contributed to an increase in the amount of available data related to fetal monitoring [63].

As a result, if adding more measurements could hopefully contribute to better insights into pathophysiological systems, inevitability it increases the complexity of data analysis as well as the interpretation of the extracted results. Machine learning methodologies appear as a possible solution to this issue, as they can face large and complex datasets [63], [64]. However, it is also to be underlined that when a subset of features is automatically extracted from a large amount of data, the interpretation of the results is usually difficult to be linked to the a priori knowledge of the underlying physiological mechanisms.

In the presented study, we designed a two-step methodology for the early identification of a pathological fetal state, namely: Intrauterine Growth Restriction (IUGR). The implementation was achieved by deriving features from a single antepartum CTG trace by means of advanced

signal analytics. Subsequently, various machine learning techniques were trained with the extracted FHR features. The rationale for employing such physiology based heart rate features aimed to realize a tool capable of deriving an interpretable link between the machine learning results and the physiological mechanisms of fetal regulation. Moreover, the requirement of early identification is achieved by removing the influence of gestational age (GA) at which the available traces were acquired, thus providing a reliable and effective set of tools for the antenatal IUGR discrimination.

As a proof of concept of an impactful and clinically relevant application of artificial intelligence in the field of fetal monitoring, in this paper we compared the validity and performances of several machine learning techniques for the classification of healthy fetuses versus fetuses affected by IUGR. The former pathology along with small for gestational age (SGA) represent the second cause of perinatal mortalities, contributing to 52% of stillbirths [38]. Moreover, the IUGR condition has been extensively reported as affecting perinatal and postnatal development under several different aspects [34]. As reported in [39], the key point in IUGR management is the early identification of the pathology to the aim of improving both the time setting and the management of delivery. Unfortunately, methodologies towards a reliable and timely detection of IUGR condition are still pending, to the point where the assessment can only be performed at birth [39]. Throughout the years, the lack of consensus in IUGR definition has resulted in uncertainty in the diagnosis of the pathology. Additionally, antepartum CTG is not consensually adopted as the routine approach in this context. As a consequence, the overall outcome of IUGR babies has not changed much over time [39]. The crucial challenge which is yet to be addressed is aimed to develop reliable tools which ideally would be able to provide antenatal identification of IUGR condition, starting from the available and clinically recorded data.

Material and methods

Data collection and participant selection

In a collaboration framework among the Ob-Gyn Clinics at the Azienda Ospedaliera Universitaria Federico II, Napoli, Italy, Biomedical Engineering Labs of Politecnico di Milano, Italy, and Università di Pavia, Italy, FHR traces were collected in a large population of pregnant women. Among the available CTG recordings, we asked clinicians to select 120 CTG recordings: 60 Healthy and 60 early IUGR fetuses. The left-hand side of Figure 1 displays a 30-minutes

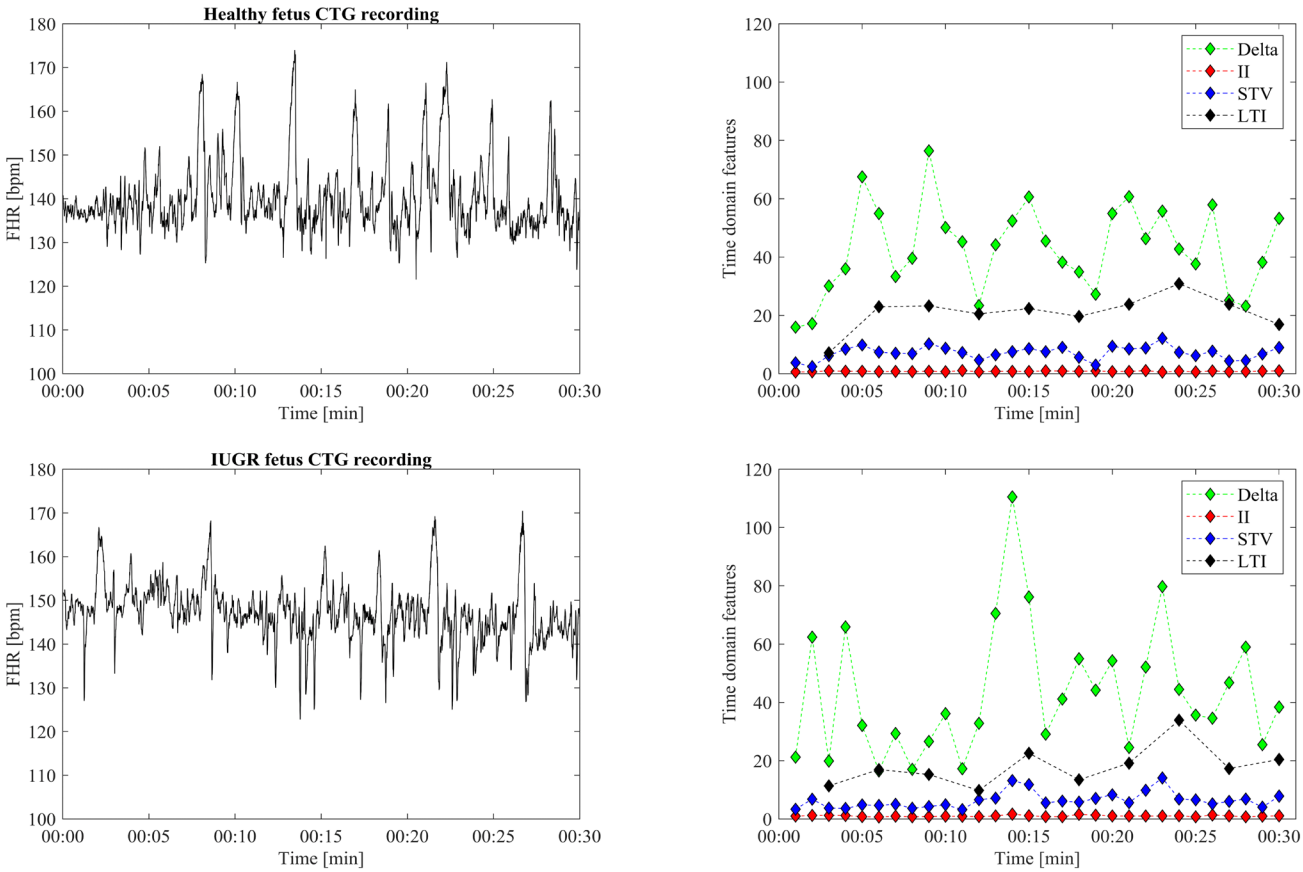


Figure 1. Left panel display two 30-minutes CTG recordings of a healthy (top) and an IUGR (bottom) participants respectively. On the right-hand side, the corresponding time series of time domain features are shown. Delta, II, and STV are computed by considering 1-minute window, thus resulting in 30 estimates throughout the reported recording. On the other hand, a 3-minutes window is employed in STV computations, thus 10 values are obtained.

segment for a healthy (top) and an IUGR (bottom) CTG traces. The prenatal fetal condition for each participant was verified after delivery to confirm group membership previously suspected at the CTG timepoint. Exclusion criteria encompass history of pre-existing maternal diseases or drug abuse during gestation, fetal chromosomal and/or major congenital anomalies and inadequate umbilical cord samples at birth. Healthy fetuses at birth presented the following characteristics: weight and abdominal circumference ±10% with respect to the normative ranges and 5-minutes Apgar score = 10. The in-utero diagnosis of IUGR pathology was assessed by ultrasound biometry, Doppler velocimetry of Umbilical Artery (UA), Middle Cerebral Artery (MCA), Ductus Venosus (DV). PI of UA and DV were considered abnormal when >95th percentile for gestational age, in case of absent or reverse A-wave or end diastolic flow in DV and UA, and PI of MCA <5th percentile. At birth, IUGR fetuses presented estimated weight below the 10th percentile, abdominal circumference below the 10th percentile for the corresponding GA and 5-minutes Apgar score ≤ 8.

The average duration of the FHR tracing was > 30 minutes for both healthy fetuses and IUGRs to contain both activity and quiet periods of the fetus. FHR recordings were collected in a controlled clinical environment, with the pregnant woman lying on a bed during the standard protocol of non-stress test. The average GA at CTG measurement for healthy fetuses was 34.78±0.53 weeks (Inter Quartile Range (IQR) = 34-35) whereas for IUGR fetuses was 32.27±2.79 weeks (IQR = 30-34). The reason for a nonoverlap in terms of GA between the two groups relies on the fact that in

Table 1. Demographics of healthy and IUGR fetal groups (mean±standard deviation).

Population	Healthy	IUGR
Subject number	60	60
Maternal Age (years)	32.34 ± 5.64	29.68 ± 6.21
GA at CTG (weeks)	34.78 ± 0.53	32.27 ± 2.79
GA at birth (weeks)	39.74 ± 1.15	34.15 ± 2.99
Newborn weight at birth (g)	3275 ± 518	1479 ± 608
Fetal gender (male/female)	25 / 35	34 / 26
1-minute Apgar score	8.46 ± 0.53	10.00 ± 0.00
5-minutes Apgar score	6.61 ± 5.64	7.44 ± 0.56
Delivery mode	58% Spontaneous 42% Caesarean	15% Spontaneous 85% Caesarean

clinical practice normal pregnancies are usually monitored only after the 33rd week of gestation whereas earlier assessments are usually available when considering suspected IUGR cases. Additional information regarding the clinical characteristics of the population analyzed in this work is reported in Table 1.

FHR signal acquisition and preprocessing

The FHR signals were measured through Philips fetal monitor Avalon FM30, connected to a PC. The fetal monitor series HP-135x employs an autocorrelation technique to compare the demodulated Doppler signal of a heartbeat and the following one. Such Doppler signal is sampled at 200 Hz. AutoCorrelation Function (ACF) is computed based on time windows of length 1.2 s, corresponding to a lower bound for FHR equivalent to 50 beats per minute (bpm). A peak detection software determines the heart period (equivalent to the RR interval) based on ACF. By using the interpolation of the peaks temporal position, the system achieves an effective resolution below 2 ms [65]. The resulting heart period is then converted into a heart frequency in beats/minute (bpm).

Due to historical reasons, almost all commercially available fetal CTG monitors employed in clinical practice display the fetal heart rate expressed in bpm. These monitors produce an FHR value in bpm every 250 ms and the series is progressively stored in a buffer. In this work, FHR

time series were acquired and analyzed by means of the 2CTG2 software [66], realized by a collaboration among our Academic research groups (University of Pavia and Politecnico di Milano), a software company (S.E.A. s.r.l. – Pavia, Italy), and Hewlett Packard Italy. 2CTG2 retrieves FHR values from the buffer at 2 Hz. Thus, Nyquist frequency for FHR series is 1 Hz. The rationale behind reading the FHR values every 0.5 sec represents a reasonable compromise to achieve enough bandwidth and acceptable accuracy. In fact, as reported in [67] significant correlations between beat-to-beat FHR values and the 2 Hz derived series were found. Moreover, for some nonlinear indices, as the ones to be discussed in the following section, 2 Hz sampling frequency has been reported as the best solution as it contributes to simpler and faster FHR signal processing and archiving [67].

As known by practitioners in CTG tests, many factors may worsen the accuracy of FHR detection and/or cause a signal loss. In the employed monitors, a quality index provides a direct measure of FHR signal quality by means of three different levels: optimal (green), acceptable (yellow), and insufficient/unavailable signal (red). Such index is based on the output of the autocorrelation procedure upon which FHR signal extraction is built. Each FHR recording has been divided in subintervals of either 120 (60 s) and 360 points (180 s), after having red-quality points being corrected. The use of 1-minute or 3-minutes subintervals is related to differences in the extracted parameters and it will be explained in the following sections. Isolated red-quality points were substituted, through a moving average procedure, with the average of the nearest five FHR points. On the other hand, subintervals including either more than five consecutive red-quality points or more than 5% of red-quality values (6 FHR values out of 120 points per subinterval or 18 FHR values out of 360 points per subinterval) were discarded in further analysis.

Selection and statistical preprocessing of features

In previous works, we approached the identification of IUGR fetuses by means of various FHR-based encompassing time domain, frequency, and nonlinear domains.

Time domain indices were computed as suggested by Arduini et al. [58]: Delta, Interval Index (II), Short Term Variability (STV), and Long Term Irregularity (LTI). The former three parameters were computed dividing the signal in windows of length equal to 60 s, LTI in windows of length equal to 180 s.

The frequency content of the FHR signal was analyzed by means of Power Spectral Density (PSD) [68]. This technique provides the power associated with specific frequency components of

FHR, as described in detail in [43]. In this study, the power associated to Low Frequency band (LF_pow) is computed in the frequency range (0.03-0.15 Hz), Movement Frequency power (MF_pow) in the frequency range (0.15-0.5 Hz), High Frequency power (HF_pow) in the frequency range (0.5-1 Hz). Powers in the different bands have also been combined to extract the ratio LF/(MF+HF).

Regarding nonlinear features of the FHR signal, we estimated Approximate Entropy (ApEn) [60], Lempel Ziv Complexity (LZC) with binary alphabet [44], and Phase Rectified Signal Average (PRSA) features [69]. Nonlinear FHR measures were computed considering nonoverlapping windows of length equal to 180 s, with the exception of PRSA features, namely APRS and DPRS which consist of a single estimate since they are computed based on the entire recorded CTG trace. Additional information on the extracted parameters is reported in the Appendix.

The complete procedure of parameter extraction produced N=12 indices, 10 of which are extracted by averaging the corresponding time series (extracted by subdividing the FHR recording in windows), and 2 of them are global parameters computed considering the entire recording. The right-hand side of Figure 1 displays an example of time domain parameters. The length of each series is equal to the number of available windows (number of acceptable intervals) in the original CTG trace after performing the quality assessment. As reported in [43], the majority of FHR parameters can noticeably vary depending on the fetal state (quiet or activity). In order to reduce such intrasubject source of variability and considering that fetal state annotation cannot be performed routinely in clinical practice, the average of the parameters of each time series was calculated. This approach is justified by the fact that our database contains recordings with both activity and quiet periods.

The reason behind the selection of this restricted subset of features relies on their individual peculiar ability in discriminating IUGRs and healthy fetuses as described in the following. To summarize: the a priori knowledge parameters to be employed as the starting set of features for further analysis are: Delta, II, STV, LTI (time domain) [58]; LF_pow, MF_pow, HF_pow, LF/(MF+HF) (frequency domain) [43]; ApEn(1, 0.1) [60], LZC(2, 0) , Acceleration Phase Rectified Slope (APRS) and Deceleration Phase Rectified Slope (DPRS) (nonlinear domain) [70]. Parameters employed in the computation of ApEn were m=1 and r=0.1 thus resulting in the feature ApEn(1, 0.1). LZC was computed within a binary approach, having the factor value (p) set to zero,

the computed quantity is reported as LZC(2, 0). Additional information regarding LZC applied to FHR analysis may be found in [44]. The last nonlinear technique employed to investigate FHR was the so-called PRSA method, introduced by Bauer [69]. In this context, Acceleration Phase Rectified Slope (APRS) and Deceleration Phase Rectified Slope (DPRS) were computed as reported in [46].

All preprocessing operations on the extracted features were performed by R, a free software environment for statistical computing [71]. A very limited portion of the total number of subjects (8 out of 120) presented some missing features (the percentage of missing features is equal to 1.5% of the total number of features). In order to account for features missingness, we employed the R package missForest [72]. It is suitable to be used in the case of mixed-type data. The imputation procedure is based on the training of a random forest which is capable of predicting the missingness based on the observed and available data [72]. The majority of extracted features showed evidence of intermediate correlation accordingly to the definition by Cohen [73] ($0.30 < |\text{Spearman's Rank Correlation coefficient } (\rho)| < 0.50$) with the GA at which trace was acquired (GA_CTG). This assumption stands considering both the whole cohort but even when limiting the analysis to the considered IUGR population as reported in Table 1.

Table 2. Spearman's rank correlation coefficient between extracted parameters and GA_CTG — unadjusted (U) and adjusted (A) distributions (* indicates statistically significant correlation $p < 0.05$).

	Correlation Coefficient ρ					
	Unadjusted (U)			Adjusted (A)		
Parameter	**Overall**	**Healthy**	**IUGR**	**Overall**	**Healthy**	**IUGR**
Delta	0.3295 *	-0.1000	0.2128	0.0117	-0.0184	-0.0339
II	0.0024	0.0286	-0.1455	0.0174	0.0061	-0.0412
STV	0.4170 *	-0.0388	0.3505 *	0.0178	-0.0388	-0.0399
LTI	0.3030 *	-0.1266	0.0937	0.0103	-0.0429	-0.0101
LF_pow	0.1684	-0.0265	-0.1254	0.0393	0.0143	0.0052
MF_pow	-0.0286	-0.0122	0.1120	-0.0017	0.0347	-0.0124
HF_pow	-0.2407 *	-0.0041	-0.0732	-0.0544	-0.0408	-0.0049
LF/(MF+HF)	0.1684	-0.0265	-0.1254	0.0498	-0.0653	-0.0092
ApEn(1, 0.1)	0.2313 *	0.0551	0.1527	0.0083	0.0245	-0.0288
LZC(2, 0)	0.2310 *	-0.2799 *	0.0204	0.0839	-0.1556	0.0429
APRS	0.4112 *	0.0408	0.2502	0.0147	0.0184	-0.0283
DPRS	-0.4896 *	-0.0408	-0.3478	-0.0152	0.0163	0.0087

The reason behind feature adjustment for GA_CTG relies on previous studies showing the evolution of linear and nonlinear indices throughout pregnancy [40], [74]. Therefore, to address the dependence of measures with respect to time of the assessment, all variables were adjusted using a Robust Linear Regression (RLR). If compared to classical linear regression, the peculiar advantage of RLR methodology relies on deriving a model less sensitive to outliers present in the observed data. Additionally, the robust fitting method is less affected by consistent changes in small portions of the data, with respect to ordinary least squares techniques. RLR consists of an iterative process. At each step, the weight associated to each data point is updated using a process called iteratively reweighted least squares. In the first iteration, each point is assigned equal weight and model coefficients are estimated using ordinary least squares. At subsequent iterations, weights are recomputed so that points farther from model predictions in the previous iteration are given lower weight. RLR model coefficients are then recomputed using weighted least squares. The process continues until the values of the coefficient estimates converge within a predefined specified tolerance. Once the final RLR model is derived, it is possible to compute the so-called residuals which represent the difference between the observed data and the RLR fit. The derived residuals were employed in the further analyses to provide machine learning classifier with features independent from the GA at which the CTG traces were recorded. Results of correlation between parameters (Unadjusted and Adjusted) and GA_CTG are reported in Table 2. As an example, distribution of Unadjusted (_U) and Adjusted (_A) of four of the further employed features are shown in Figure 2 along with the corresponding regression lines. Furthermore, to the aim of validating the proposed methodology we tested the influence of GA_CTG by means of linear regression models. Specifically, we considered alternatively each feature or the corresponding residuals as a dependent variable and GA_CTG only, GA_CTG + Group (healthy versus IUGR), and GA + Group + GA*Group as independent variables. Results showed a consistent absence of significant effect of GA_CTG (regardless the regression model) on the residuals derived from RLR correction.

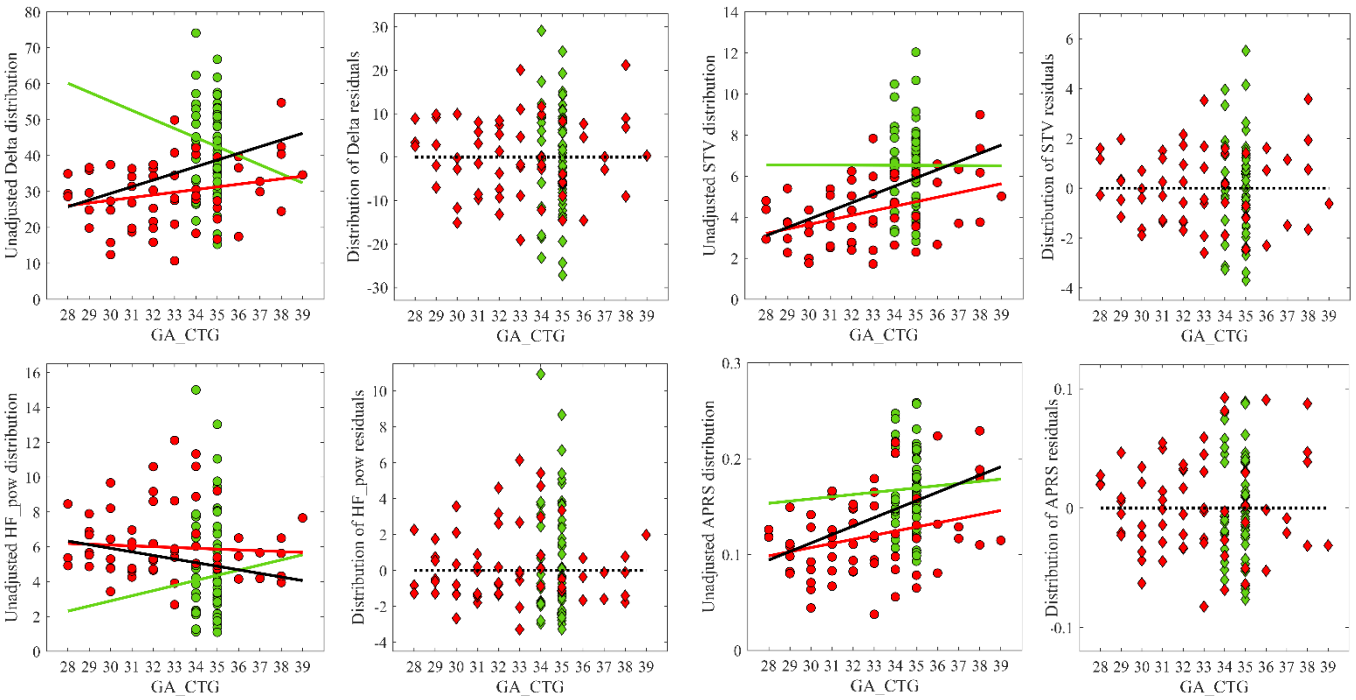


Figure 2. Scatterplots showing Delta, STV, HF_pow, APRS distributions as a function of GA_CTG. Green items and red items represent the Healthy and IUGR populations respectively. Circles and rhombuses represent Unadjusted and Adjusted features respectively. Solid lines represent the RLR fit in the healthy (green), IUGR (red), and combining the two groups (black). The three RLR fits coincide in a single fit line (black dotted) in the graphs displaying residuals given the absence of any trend after performing RLR.

Multivariate analysis

In this investigation, we deepened the preliminary results obtained by analyzing the same database analyzed in [75]. While our previous work was mainly focused on the comparison between the performances of univariate versus multivariate classifiers, in this paper we investigate within a more detailed approach the possible influence of GA over the performances of machine learning methodologies and their feature robustness and insensitivity to GA_CTG. Moreover, the more precisely conducted analysis on feature space will provide validation for the utilization of physiology based heart rate features for the early identification of IUGR pathology. As a general consideration, multivariate analysis was designed to search for an optimal decision rule in the multidimensional space of the parameters to predict the class of interest, namely healthy versus IUGR. A complete roadmap from CTG signal to binary classification is depicted in Figure 3.

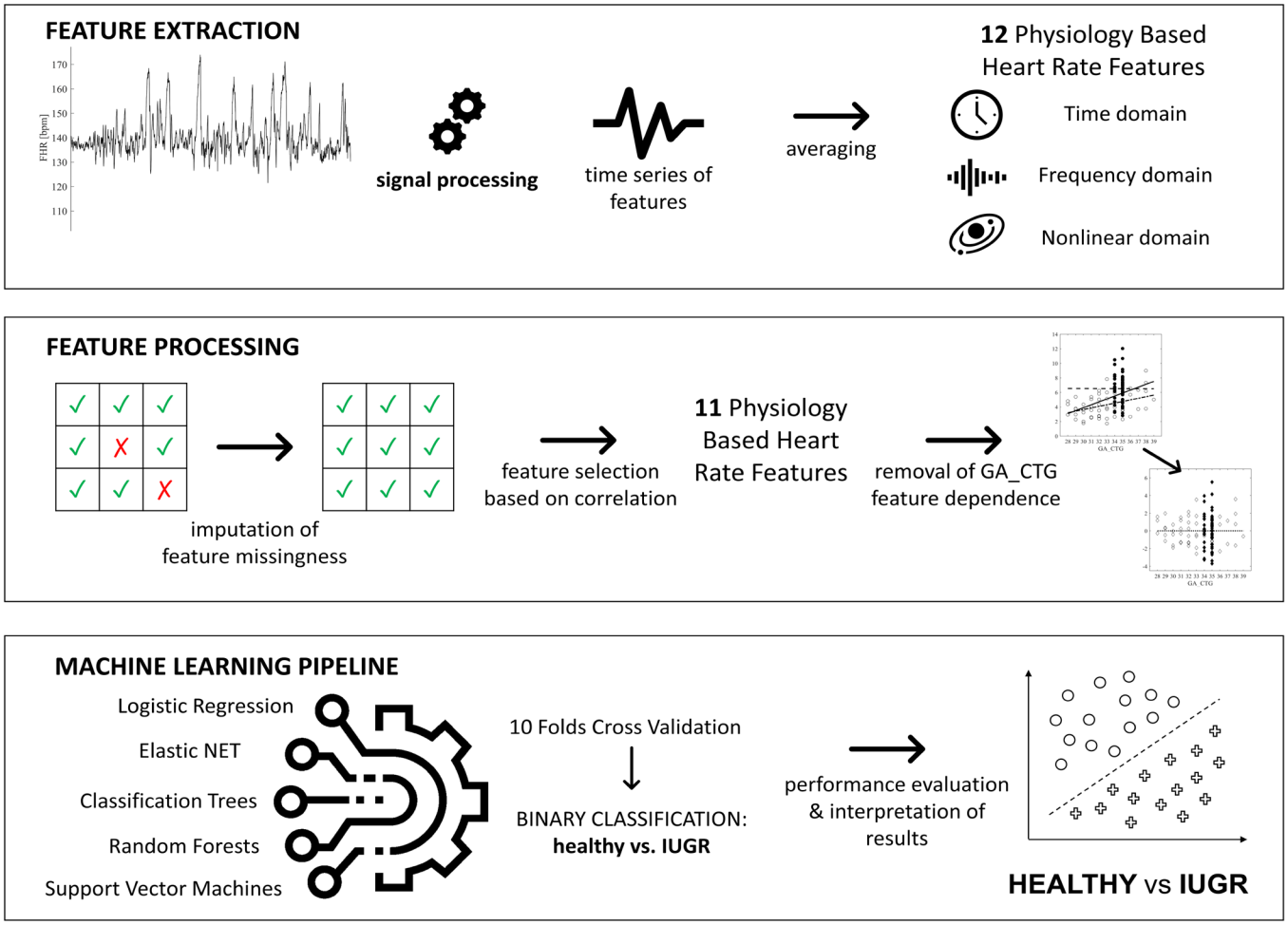


Figure 3. Machine learning approach to the classification of antepartum fetal heart rate signal. Top panel: from signal to feature extraction; middle panel: feature processing, imputation of missing parameters, feature space reduction based on physiological knowledge; bottom panel: machine learning techniques and validation of performances.

Several multivariate models were employed towards to aim of identifying the most reliable technique for predicting IUGR condition. The employed machine learning techniques and the corresponding employed R packages are reported in Table 3.

The following algorithms were applied:

Logistic Regression (LR): is a regression model where the probability of a class of interest is obtained as the results of a logistic function provided with a linear combination of the features. The general formulation for Logistic Regression is expressed in Eq. 1:

$$P(y|\boldsymbol{x}) = \frac{e^{\alpha+\sum_{i=1}^{N}\beta_i x_i}}{1+e^{\alpha+\sum_{i=1}^{N}\beta_i x_i}} \quad (1)$$

Table 3. Tested machine learning techniques, their corresponding acronyms, and employed R packages.

Machine Learning Technique	Acronym	R package
Logistic Regression including all covariates	LR	stats
Logistic Regression, stepwise feature selection and pairwise interactions between features	LR-SW-INT	stats
Logistic Regression, stepwise feature selection and without pairwise interactions between features	LR-SW	stats
RIDGE regression	RIDGE	glmnet
Elastic NET, alpha = 0.25	ENET 0.25	glmnet
Elastic NET, alpha = 0.50	ENET 0.5	glmnet
Elastic NET, alpha = 0.75	ENET 0.75	glmnet
Least Absolute Selection and Shrinkage Operator	LASSO	glmnet
Naïve Bayes	NB	e1071
Classification Trees	CT	rpart
Random Forests	RF	randomForest
Support Vector Machines, linear kernel	SVM-LIN	e1071
Support Vector Machines, polynomial kernel	SVM-POLY	e1071
Support Vector Machines, radial kernel	SVM-RAD	e1071
Support Vector Machines, sigmoid kernel	SVM-SIGM	e1071

where y is a target class (Healthy versus IUGR), x_i are the available features, α and βs are the regression coefficients estimated by the algorithm. The method, as formulated in the previous Eq. 1 generates a linear decision boundary, i.e. a hyperplane in the multidimensional space. LR can be utilized including all covariates (LR), namely the whole set of previously extracted parameters or coupled with a features selection algorithm called stepwise selection of informative features (step function) [76], allowing (LR-SW-INT) or not (LR-SW) for pairwise interactions between features.

Within the family of approaches based on regression we also employed *RIDGE regression* for binary outcomes *(RIDGE)* [77]; *elastic net regression* for binary outcomes with different alpha settings (alpha = 0.25 *(ENET 0.25)*, alpha = 0.50 *(ENET 0.5)*, alpha = 0.75 *(ENET 0.75)* [78]; and *Least Absolute Selection and Shrinkage Operator* for binary outcomes *(LASSO)* [79]. For *RIDGE*, *ENET* and *LASSO* regressions, the optimal lambda parameter was computed by considering each

training set separately (as described in the following) and the same seed was imposed for each analysis.

Naïve Bayes (NB): is a classification algorithm based on the Bayes theorem. NB assumes that the attributes x_i are conditionally independent given the class *y*, as formulated in Eq. 2:

$$P(y|\boldsymbol{x}) \propto P(y)\prod_{i=1}^{N} P(x_i|y) \quad (2)$$

Despite these rather simplistic assumptions, NB often outperforms more sophisticated machine learning algorithms. This is due to the fact that although the individual class density estimates may be biased, the assumption of feature independence (given the class variable) is not effectively affecting the posterior probabilities [40] of belonging to a specific class.

Classification Trees (CT): are widely used in the machine learning field and they consist of a set of rules that defines a tree-like structure, in which branches represent different decisional paths and terminal nodes (leaves) corresponds to the assignment to a target class. CT generates a set of nonlinear decision boundaries through piecewise constant functions in the multidimensional space. In this work, the information gain was employed as the splitting criterion for CT [80].

Random Forests (RF): are ensemble classifiers that consist of a variable number of CTs grown based on a set of attributes selected randomly from the complete set of parameters; each CT contributes with its own classification of the analyzed examples. As a result, the final classification is provided by a voting approach, which considers the complete set of CTs. Thanks to their scalability and generalization performance, RFs are increasingly exploited in clinical research [81].

Support Vector Machines (SVM): are a family of classifier capable of mapping the training samples into high-dimensional attributes space, to the aim of defining a hyperplane that maximizes the distance between observations belonging to the different classes. If the training set cannot be separated by a linear boundary, the optimal hyperplane that best discriminates between/among examples of different class labels is identified resorting to a suitable space transformation through kernel functions. In this work, we employed SVMs with linear kernel *(SVM-LIN)*, polynomial kernel *(SVM-POLY)*, radial kernel (SVM-RAD), and sigmoid kernel *(SVM-SIGM)* [82].

From an implementation point of view, models were learned on the training sets using the default settings. The available data were split into training and testing sets according to a 10 Folds Cross Validation (CV). The training sets were employed to the aim of evaluating the performances

of classification algorithms and different feature selection while the corresponding test sets were used to test the relative discriminative performances. The above-described machine learning methods were tested on either the U and A set of features towards to aim of comparing the two approaches and identifying if GA_CTG had a significant effect on IUGR classification.

Multivariate model evaluation

In this work, an IUGR subject correctly classified as such is counted as a true positive (TP), and a healthy subject correctly classified is counted as a true negative (TN). On the contrary, an IUGR subject erroneously classified as healthy is counted as a false negative (FN), and a healthy subject erroneously classified as IUGR is counted as a false positive (FP).

The performances of each model are reported in terms of four different figures of merit, namely Classification Accuracy: $CA = (TP + TN)/(TP + TN + FP + FN)$, sensitivity: $sensitivity = TP/(TP + FN)$, specificity: $specificity = TN/(TN + FP)$, positive predictive value: $PPV = TP/(TP + FP)$, and negative predictive value: $NPV = TN/(TN + FN)$. The Area Under the Receiver Operating Characteristic (AUROC) was estimated by averaging the results obtained by providing the model with different test sets, namely the ones obtained using 10 Folds CV procedure. Since the healthy/IUGR ratio was 1, model ranking was performed based on CA.

Results

Multivariate analysis has been performed considering alternatively the U and A set of features. Prior to multivariate testing, a preliminary analysis of the correlation between covariates has been performed. As a general consideration, features to be provided to any machine learning algorithm should be highly correlated with the classes to be distinguished but not be highly correlated with one another [83]. By way of example, values of correlation for the A set of covariates are reported in Table 4.

Table 4. Tested machine learning techniques, their corresponding acronyms, and employed R packages.

	Delta	II	STV	LTI	LF_pow	MF_pow	HF_pow	LF/(MF+HF)	ApEn(1, 0.1)	LZC(2, 0)	APRS
DPRS	-0.60	-0.03	-0.62	-0.42	-0.37	0.22	0.29	-0.37	-0.03	-0.26	-0.84
Delta		0.03	0.93	0.43	0.4	-0.17	-0.45	0.40	0.06	0.33	0.57
II			-0.10	0.03	-0.11	0.15	0.01	-0.11	0.01	0.13	-0.07
STV				0.38	0.36	-0.13	-0.41	0.36	0.02	0.31	0.57
LTI					0.29	-0.18	-0.21	0.28	0.08	0.17	0.37
LF_pow						-0.72	-0.75	0.99	-0.27	0.24	0.33
MF_pow							0.15	-0.72	0.20	0.07	-0.27
HF_pow								-0.75	0.32	-0.39	-0.18
LF/(MF+HF)									-0.27	0.26	0.33
ApEn(1, 0.1)										0.14	0.02
LZC(2, 0)											0.19

The correlation coefficient values in each domain are on average higher than comparing feature correlation in the same area. The former result is related to the fact that the proposed features have the ability to grasp different characteristics of FHR, thus their information content is different, resulting in a low value of correlation. Regarding high values of correlation among indexes of the same domain, a clear example is the parameter LF/(MF+HF) which is highly correlated to the other frequency extracted indexes (LF_pow, MF_pow, and HF_pow). Based on this criterion, the ratio LF/(MF+HF) was excluded from the set of employed parameters, resulting in a reduced parameter space of 11 features: Delta, II, STV, LTI (time domain); LF_pow, MF_pow, HF_pow (frequency domain); ApEn(1, 0.1), LZC(2, 0), APRS, DPRS (nonlinear domain).

Table 6. Median (25th, 75th percentiles) of classification accuracy (CA), sensitivity, specificity, positive and negative predictive values (ppv and npv) for adopted machine learning techniques learned on the **ADJUSTED** set of covariates. Machine learning techniques are sorted in descending order of CA.

Model	CA	Sensitivity	Specificity	PPV	NPV
RF	0.855 (0.794, 0.916)	0.838 (0.729, 0.947)	0.871 (0.766, 0.976)	0.889 (0.799, 0.980)	0.862 (0.773, 0.951)
LR-SW	0.833 (0.759, 0.908)	0.867 (0.773, 0.961)	0.800 (0.665, 0.935)	0.835 (0.737, 0.934)	0.870 (0.785, 0.955)
LR	0.825 (0.743, 0.907)	0.850 (0.731, 0.969)	0.800 (0.665, 0.935)	0.830 (0.730, 0.931)	0.862 (0.766, 0.958)
SVM-RAD	0.818 (0.738, 0.897)	0.850 (0.719, 0.981)	0.786 (0.687, 0.885)	0.806 (0.723, 0.888)	0.866 (0.756, 0.977)
LASSO	0.817 (0.716, 0.917)	0.867 (0.744, 0.990)	0.767 (0.627, 0.907)	0.800 (0.687, 0.914)	0.873 (0.761, 0.985)

Table 6. Median (25^{th}, 75^{th} percentiles) of classification accuracy (CA), sensitivity, specificity, positive and negative predictive values (PPV and NPV) for adopted machine learning techniques learned on the **UNADJUSTED** set of covariates. Machine learning techniques are sorted in descending order of CA.

Model	CA	Sensitivity	Specificity	PPV	NPV
RF	0.911 (0.860, 0.961)	0.902 (0.820, 0.985)	0.919 (0.819, 1.019)	0.936 (0.859, 1.013)	0.918 (0.852, 0.984)
CT	0.911 (0.846, 0.975)	0.871 (0.766, 0.976)	0.950 (0.892, 1.008)	0.949 (0.890, 1.009)	0.893 (0.808, 0.978)
LR-SW	0.867 (0.797, 0.937)	0.900 (0.817, 0.983)	0.833 (0.721, 0.946)	0.859 (0.774, 0.944)	0.900 (0.822, 0.978)
SVM-RAD	0.867 (0.781, 0.952)	0.850 (0.762, 0.938)	0.883 (0.770, 0.996)	0.893 (0.790, 0.996)	0.856 (0.775, 0.938)
SVM-POLY	0.850 (0.762, 0.938)	0.933 (0.850, 1.017)	0.767 (0.627, 0.907)	0.817 (0.712, 0.922)	0.927 (0.838, 1.017)

Table 5 and Table 6 report the mean discriminative performances of the top five machine learning techniques in classifying the test sets (10 Folds CV) for _U and _A set of covariates respectively. The average performances in the first case are: CA=0.8812, Sensitivity=0.8912, Specificity=0.8704, PPV=0.8908, NPV=0.8988, whereas on the second one: CA=0.8296, Sensitivity=0.8544, Specificity=0.8048, PPV=0.8320, NPV=0.8666.

The two machine learning techniques which outperformed, showing the best discriminative performances were: RF_U (Random Forests_Unadjusted): mean CA=0.911 and CT_U: CA = 0.911 when considering the model learned on U covariates. In the case of adjusted covariates by GA_CTG, RF_A(Random Forests_Adjusted): mean CA=0.855, and LR-SW_A: mean CA=0.833, showed the best CA among the proposed machine learning models.

Focusing the attention on the comparison of AUROC for RF_U, CT_U, RF_A, and LR-SW_A, no statistically significant difference was observed after performing post-hoc tests between models' AUROC. The values of AUROC (averaged over the 10 test sets) for RF_U: AUROC=0.974, CT_U: AUROC=0.892, RF_A: AUROC=0.935, and LR-SW_A: AUROC=0.933 are reported in Figure 4.

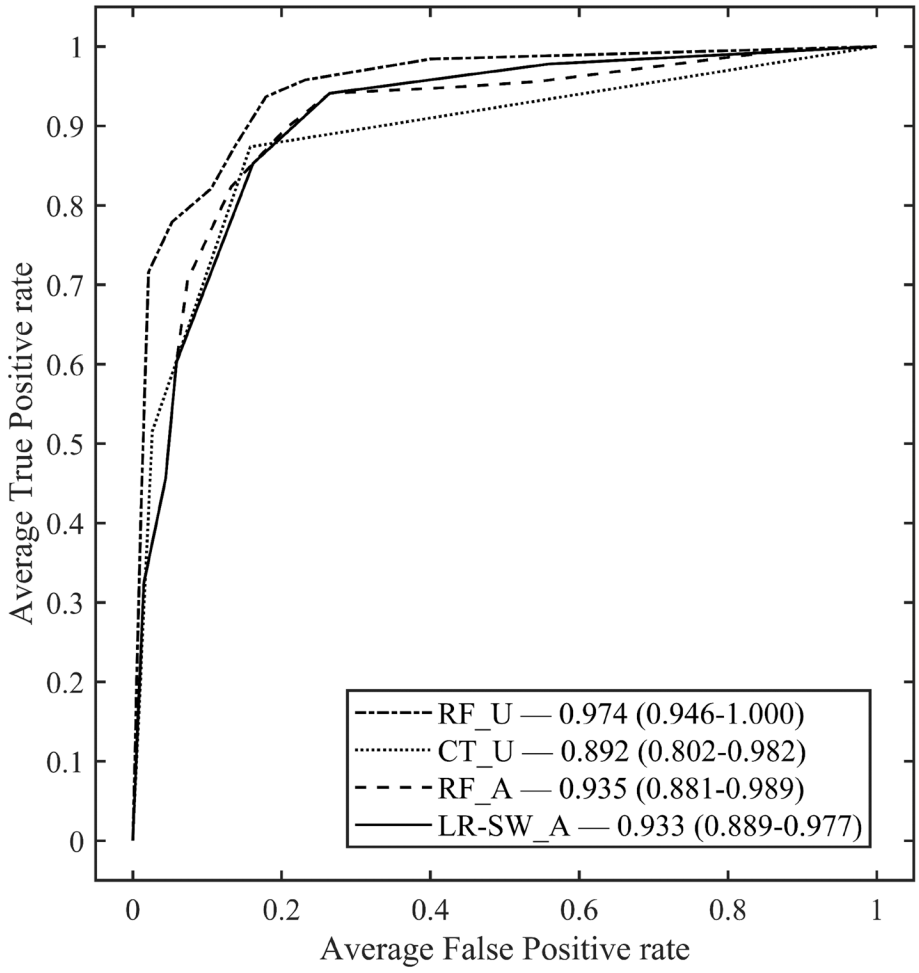


Figure 4. Mean AUROC (95% CI) for the multivariate analysis: RF_U, CT_U, RF_A, and LR-SW_A. The label _U indicates UNADJUSTED features. The label _A indicates ADJUSTED features. False positive rate is defined as FP/(FP+TN) and true positive rate corresponds to sensitivity.

In order to define the final models, RF and LR-SW were learned on the whole set of computed features. Nevertheless, features selection procedures for both RF and LR-SW were performed. It is crucial to pinpoint that investigating the performances of the former machine learning techniques on a reduced feature space may be helpful in reducing the amount of FHR extracted parameters within achieving the same level of prediction accuracy.

Regarding RF_U and RF_A, the relative importance of each feature is shown in Figure 5. Results indicate that for the U set of covariates LZC(2, 0), ApEn(1, 0.1), HF_pow, LTI and DPRS caused the greatest decrease in terms of CA if removed from the model. Similarly, when considering RF_A, LZC(2, 0), HF_pow, ApEn(1, 0.1), LTI and LF_pow were identified as most explanatory variables for the model. It is crucial to highlight that the most explanatory parameters

are encompassing all the investigated domain: time, frequency and nonlinear. The latter assumption is valid for both RF_U and RF_A enforcing the idea that combining FHR features belonging to different domains provides a more comprehensive and extensive snapshot of the interacting mechanism leading to the IUGR condition.

Coming to the second-best performing machine learning techniques, CT_A identifies the same covariates previously found for RF_U as most explanatory. In the case of LR-SW_U, the covariates producing the highest decrease in accuracy if excluded are: LTI, LZC(2, 0), STV, LF_pow, HF_pow, ApEn(1, 0.1), MF_pow. Consistently with RF, despite the different ranking of importance, CT_A and LR-SW_U select a reduced set of variables encompassing the three different domains. Figure 5 displays variables ranked in descending order of mean decrease in accuracy. It is crucial to pinpoint that RF methodology appears stable and quite insensitive to covariance dependence upon GA_CTG. LZC(2, 0) is identified as producing the most impactful decrease in accuracy by both RF_U and RF_A. The remaining covariates are on average ranked

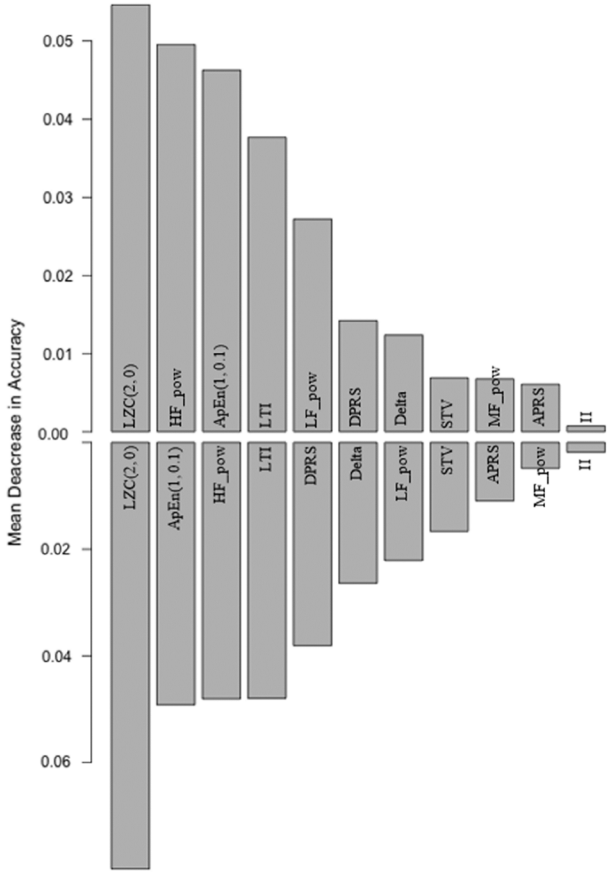


Figure 5. Variables importance according to the RF classifier on the whole cohort and associated decrease in model CA when progressively excluding features. Top bar graph is relative to RF_A and bottom one to RF_U.

in a similar fashion by the two models, strengthening the limited impact of GA_CTG on classification accuracy. On the opposite, it is peculiar to observe the dramatic drop of performances when comparing CT_U and CT_A. In particular, CT_U is ranked as the second-best performing machine learning technique while CT_A is the least performing one (CA=0.771, Sensitivity=0.757, Specificity=0.786, PPV=0.805, NPV=0.802) followed by SVM-SIGM only (CA=0.702, Sensitivity=0.705, Specificity=0.700, PPV=0.709, NPV=0.710). In the latter case, GA_CTG plays a fundamental role and it becomes evident that the correction for such dependence is mandatory to provide accurate discrimination between healthy and IUGR fetuses. Regarding the remaining machine learning techniques, namely SVM and ENET, yet not giving the best performances, they appear less dependent upon GA_CTG providing comparable results in terms of CA when employing either U or A set of covariates.

Discussion

The presented investigation provides evidence of the successful application of machine learning techniques for the early identification of IUGR condition in the antepartum period. Such design appears as radically different with respect to the up-to-date clinical practice where IUGR condition is assessed at birth and only suspected in the antepartum period. The rationale for the utilization of the presented physiology based heart rate features relies on the fact that these features as standalone parameters had shown enhanced individual discrimination power in classifying healthy versus IUGR fetuses [44], [46], [84].

However, throughout the years it has become clear that a single index cannot be descriptive of all pathophysiological processes taking place in the pregnancy period thus the need for multivariate analysis of FHR emerged as evident. Additionally, the pronounced variability of each parameter as pregnancy progression poses a substantial barrier for their immediate translation into a clinically applicable tools. These are the main reasons contributing to the use of physiology based heart rate features and gestational age adjustment in this investigation. Moreover, our approach demonstrated the independence of different machine learning methodologies to the time at which CTG recordings were acquired. These findings are reflected in the discriminative performance of the employed machine learning classifiers as reported in Figure 4. Although characterized by different diagnostic results, no statistically significant differences were found in the classification

of healthy versus IUGR fetuses comparing the top-performing techniques trained with the unadjusted and adjusted set of features.

Coming to the discussion of the results section, it appears evident that both RF_U and RF_A achieved adequate performances, thus proposing as a possible candidate as a tool for early discrimination in the context of the presented investigation. The assumption of RF insensitivity to GA_CTG found in this investigation is based on the results reported in Figure 4 and in Figure 5. Specifically, as reported in Figure 4 and statically verified by a post-hoc analysis on AUROC, no difference in the classification performance between RF_U and RF_A was found. Moreover, Figure 5 shows agreement in feature ranking comparing the results obtained providing RF classifier with either the unadjusted and the adjusted set of features.

Random Forest is becoming a popular machine learning technique and it has been claimed as particularly accurate and interpretable by several authors [85], [86]. A clear example of interpretability of the results is the feature ranking results reported in Figure 5. Consistently with previous findings [84], LZC(2, 0) is associated with the most considerable mean decrease in accuracy. On average, IUGR fetuses have been reported as characterized by lower values of LZC with respect to healthy ones [46], as this is also verified in this analysis. The reported difference is to be attributed to lower complexity of FHR for pathological subjects, thus supporting the hypothesis of an unbalance in the autonomic nervous system mechanisms in IUGR condition. Similarly, values of ApEn(1, 0.1) are greater in healthy versus IUGR. Nevertheless, this entropy index resulted in a lower mean decrease in accuracy accordingly to the reported lower discriminative power with respect to LZC measures [84]. Moreover, the corresponding time domain index (LTI) which quantifies FHR variability considering windows of analogous time duration was found among the top informative features. As for both LZC(2, 0) and ApEn(1, 0.1), LTI values in healthy are greater to the ones for IUGR subjects as previously found in [46]. This latter finding contributes to the hypothesis of an impaired ANS regulation in the pathological conditions. Lastly, PRSA-derived index DPRS was found significantly greater in IUGR versus healthy. Despite not providing an analogous definition of acceleration and deceleration as ones found in the clinical context, the PRSA slope is dependent upon both the amplitude and duration of the ANS-related events modulating the FHR [59].

To summarize, the reported results reinforce the idea that several controlling mechanisms affect HRV, acting linearly and nonlinearly. This specifically happens when a pathological condition arises, and the analytic frameworks need to merge and combine information coming from different domains to obtain an exhaustive and comprehensive description of FHR dynamics. The latter consideration is reflected in the obtained findings considering feature ranking in RF, reporting the first five features encompassing the three domains of investigation, namely time, frequency, and nonlinear. Findings reported in this investigation confirm the importance of a multivariate approach to investigate the variety of implications resulting from a pathological condition such as IUGR. The advantages resulted by the application of several machine learning techniques rely on: i) easy-to-use model capable of providing an early and interpretable antenatal diagnosis of IUGR condition; ii) parameters extracted from routinely CTG examination can be fed into the model regardless the considered GA_CTG. The latter novelty is of primary importance given that, in nowadays clinical practice, IUGR fetuses are usually monitored far in advance with respect to healthy ones so that the proposed model may see its direct translation in the clinical field. At the same time, it is imperative to underly that the GA distributions analyzed in this report do not cover the wider GA ranges which would potentially pertain the early IUGR phenotype. Thus, the validity of the introduced model requires further and extensive validation large and extant available dataset. The cutting-edge frontier for the methods described in this work would be focusing on tracking the evolution from health condition to pathological state in a patient-specific way by integrating heterogeneous data which are dynamically evolving in time.

A Machine Learning Approach to Monitor the Emergence of Late Intrauterine Growth Restriction

Introduction

Antenatal fetal heart rate (FHR) is a widely used tool to monitor fetal wellbeing [87]. The assessment of fetal Heart Rate Variability (HRV) has been described as a marker of the functional state of the autonomic nervous system (ANS) thus, an indicator of the fetal development throughout pregnancy. In the context of fetal pathological states, Intrauterine Growth Restriction (IUGR) is one of the most relevant complications of pregnancy and it has been reported to alter HRV [8], [49]. IUGR is associated with a decreased rate of fetal growth, which is the result of an abnormal supply of maternal nutrients and placental transfer to the fetus. IUGR is a pathological fetal state characterized by an increased mortality and/or morbidity [34], [38], [39]. The two phenotypes of IUGR (Early and Late) can be identified based on onset, evolution, Doppler parameters modifications, and postnatal outcome [37].

In this study, we will focus on a population of Late IUGR. This condition is associated with a substantial increased prevalence if compared to Early IUGR. The main cause for the insurgence of Late IUGR is fetal hypoxemia/hypoxia secondary to placental insufficiency. Moreover, it is often associated with multiple placental anomalies that by contrast have less influence on placental resistance. Therefore, the umbilical Doppler indices are often unaffected, thus making the diagnosis of late IUGR more difficult, due to the large variability of fetal parameters on growth charts in the third trimester [88]. Late IUGR is suspected when the fetal growth curve slows down or even become plateaus. Undetected IUGR in the third trimester of pregnancy represents the main cause of unexplained stillbirths in low-risk pregnancies, thus better antenatal diagnosis, treatment and timely delivery could diminish the risks significantly [89].

In order to investigate identification of Late IUGR through FHR analysis, we used the cardiotocography (CTG), which combines the measure of FHR through a Doppler ultrasound probe with the detection of uterine contractions using a pressure sensor. Although CTG analysis is still performed visually in the vast majority of Ob-Gyn clinical settings (following guidelines edited by national and international scientific societies, such as the International Federation of Gynecology and Obstetrics (FIGO) [90]), a progressive transition to computerized approaches has

been reported in recent years. Computerized systems are able to extract FHR parameters from multiple domains (time domain and frequency domain, complexity/nonlinear methodologies [68]) and represent the initial step towards more sophisticated FHR analyses able to benefit from machine learning algorithms. As a matter of fact, the FHR regulation is the result of multiple and diverse neurological feedback loops, hormones, and various external factors, thus resulting into complex temporal dynamics, which are usually missed by the simple visual inspection. Additionally, previous studies have shown the strength of a multivariate framework in detecting fetal acidemia [91] and a previous paper from our group used machine learning approaches to diagnose IUGR, but mainly focusing on the early phenotype [49].

In this study, classical FHR features were complemented with advanced nonlinear features and subsequently employed to train machine learning algorithms for the detection of Late IUGR condition in a database of 262 fetuses. As a first step, we adopted an approach similar to the analyses presented in [49]. Nevertheless, the superior performance of Support Vector Machines (SVM) appeared clear in contrast with the results previously obtained in the context of early IUGR phenotype (data not shown). Thus, we decide the primarily focused on a novel adaption of SVM which allowed to generate meaningful and interpretable results with the potential of their translation into clinical practice. Specifically, we report the enhanced performance of nonlinear features over traditional parameters and the significant improvement in specificity and sensitivity of multiparametric machine learning approaches over univariate analysis. Furthermore, by utilizing an interpretable variant of support vector machines, we were able to identify the features that contributed the most to classification accuracy.

Material: database and preprocessing

Dataset

Antepartum FHR recordings were collected at the Azienda Ospedaliera Universitaria Federico II (Napoli, Italy), during daily routine monitoring. Traces were recorded in a controlled clinical environment, with participants lying supine on a bed undergoing a standard non-stress test protocol. Data were acquired using Philips cardiotocography (CTG) fetal monitor Avalon FM30 connected to a computer. The device employs an autocorrelation technique to compare the demodulated Doppler signal of a given heartbeat and the subsequent one. The resulting resolution for beat detection is below 2 ms. The derived CTG signal consists of a series of FHR values

sampled at 2 Hz and expressed in beats per minute (bpm). Additionally, each FHR sample is accompanied by an indication of signal quality: optimal, acceptable, or insufficient based on the results of autocorrelation technique.

The cohort analyzed in this work comprised 102 Late Intrauterine Growth Restriction (IUGR) fetuses, and 160 healthy fetuses matched for GA at the first CTG examination. Fetuses in both groups underwent a routine ultrasound examination at approximately 34 weeks GA which did not exhibit any alteration in fetal growth, abnormalities in the doppler velocimetry, middle cerebral artery, and ductus venosus. Once subsequently admitted for a CTG recording at 37.54 ± 0.77 (mean ± std) weeks, fetuses in the healthy group did not shown any abnormality in the FHR trace, whereas IUGR fetuses (admitted at 36.94 ± 0.59 weeks) did present irregularities in their CTG recordings. A concurrent ultrasound examination showed alteration in both growth and doppler profiles in this group of fetuses. The clinical definition of Late IUGR adopted in this work reflects the guidelines reported in [92]. Each prenatal fetal condition was verified after delivery to confirm group membership previously suspected at the CTG timepoint. The length of the CTG recordings considered in this study is 40 minutes in both populations.

Clinical data of the analyzed populations are reported in Table 7. Fetuses in the healthy group were characterized by higher birthweight, Apgar score 1 min, and rate of spontaneous

Table 7. Clinical data for the healthy and IUGR fetuses expressed as mean ± standard deviation or number (%). † denotes a significant difference between healthy and IUGR fetuses, $p < 0.05$.

	Healthy *n* = 160	IUGR *n* = 102
GA at CTG [weeks] †	37.54 ± 0.77	36.94 ± 0.59
Maternal age [years]	32.23 ± 5.16	32.36 ± 5.82
Birthweight [g] †	3311.62 ± 373.87	2038.40 ± 348.15
Umbilical cord pH †	7.28 ± 0.08	7.32 ± 0.06
Fetal sex [male]	55.00 %	46.08 %
Apgar score 1 min > 7 †	91.88 %	78.43 %
Apgar score 5 min > 7	100.00 %	98.04 %
Mode of delivery †	59.38 % Vaginal	28.43 % Vaginal
	40.62 % Caesarean	71.57 % Caesarean

vaginal delivery compared to the considered IUGR population. Similar results have been reported in other populations of Early [47] and Late [37] IUGR fetuses.

FHR time series and preprocessing

The first preprocessing step towards the computerized analysis of the acquired traces consisted on subdividing each FHR recording in shorter segments of length 120 points (60 s) or 360 points (180 s). The choice of 1-min or 3-min subintervals is related to the different time scales on which CTG-derived features are computed and whose procedure will be described in the following sections. Subsequently, segments including more than five consecutive red-quality points or more than 5% of red-quality samples (6 FHR values out of 120 points per subinterval or 18 FHR values out of 360 points per subinterval) were discarded in further analysis. Lastly, isolated insufficient-quality points were substituted, through a moving average procedure, with the average of the nearest five FHR points.

Methods: Features and Radial Basis Function Support Vector Machines (RBF-SVM)

Features

The present contribution focuses on building a machine learning-based screener of Late IUGR pathology rather than a feature design-oriented approach. Thus, the employed features were selected on the basis of the a priori knowledge on various quantifiers of fetal ANS dynamics in different domains, complemented by fetal and maternal information. Figure 6 reports a schematic workflow for the framework implemented in this work.

Morphological and Time Domains (MTd)

Morphological and time domain features represent the computerized and automated extraction of FIGO guidelines from FHR recordings, in terms of baseline evolution, accelerations/decelerations, and variability. Starting from the identification of FHR baseline (by means of Mantel's approach [93]), it is possible to derive the automatic counts of: large accelerations (more than 15 beats per minute over the baseline lasting 15 seconds or more), (*#acc_large*), small accelerations (fewer than 15 beats per minute), (*#acc_small*), decelerations (*#dec),* and contractions (*#contr*) accordingly to [90], [94].

FHR variability features are derived from FHR signal excluding events of accelerations and decelerations. Specifically, the overall variability is quantified by the mean and standard deviation of entire FHR signal (*FHR_mean* and *FHR_std*). Short Term Variability (*STV*), Interval

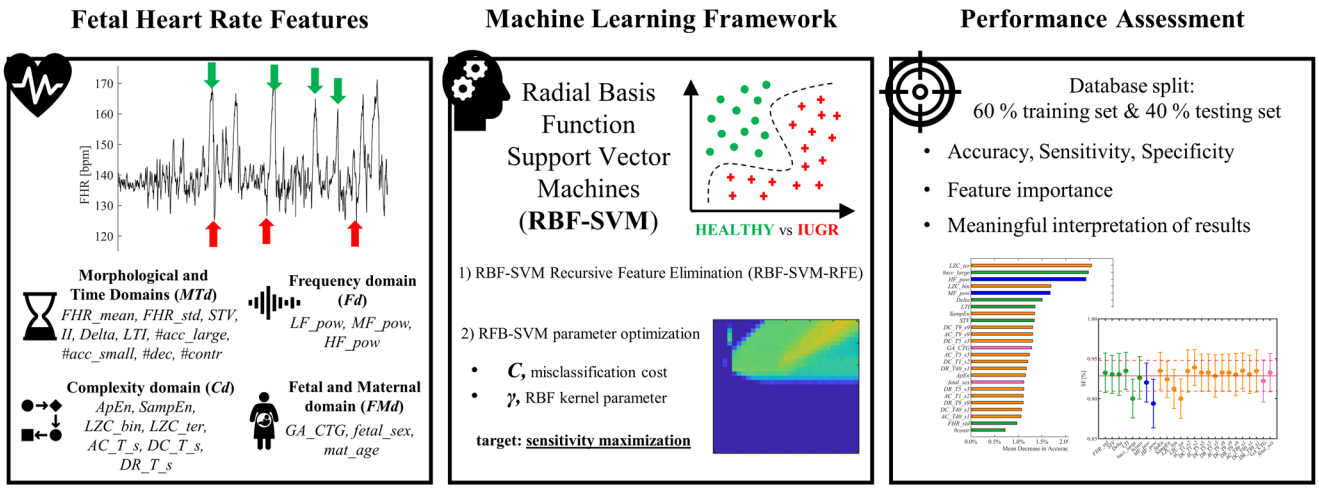


Figure 6. A schematic depiction of the machine learning approach implemented. The starting point is the extraction of FHR-based quantitative features able to describe the dynamic of FHR under different domains. The RBF-SVM approach describes the multivariate relationship among features by means of nonlinear separation boundaries. Furthermore, the proposed method allows to gain insight into the obtained results, as well as the individual contribution of each feature towards classification.

Index (*II*), and *Delta* provide estimates of short-term FHR variability considering 1-min FHR intervals. Long Term Irregularity (*LTI*) quantifies variability on a longer time scale (3-min FHR intervals). A more comprehensive description and characterization of the employed FHR variability features can be found in [43], [58], [66].

Frequency domain (Fd)

Power Spectral Density (PSD) is a largely exploited method for HRV frequency analysis. It decomposes the signal power in oscillatory components which are an indirect measure of ANS modulation over the cardiac system. PSD is computed employing autoregressive (AR) modelling, specifically by Levinson-Durbin algorithm. Based on previous findings, three specific frequency bands of interest can be identified, namely low frequency (LF) [0.03-0.15] Hz; movement frequency MF [0.15-0.5] Hz; high frequency HF [0.5-1 Hz], which quantifies the different branches of ANS modulation [43], [91]. The FHR signal undergoes an automatic decomposition into a sum of sinusoidal contributions identified by their central frequencies and the associated amount of power, thus obtaining the power in the LF band (*LF_pow*), MF band (*MF_pow*), and HF band (*HF_pow*) for each 3-min FHR segments.

Complexity domain (Cd)

The application of nonlinear methodologies to investigate FHR variability has demonstrated its usefulness in predicting fetal wellbeing in several investigations [43], [74], [95]. In the context of Early IUGR detection, Lempel Ziv Complexity (LZC) [45] has been previously

reported to have enhanced discriminative power in both univariate [44] and multivariate [49] approaches, considering binary (*LZC_bin*) and ternary (*LZC_ter*) alphabets. Additional measures of complexity such as Approximate Entropy (*ApEn*) [61] and Sample Entropy (*SampEn*) [96] have been employed for the quantification of ANS profiles in the perinatal period. The above described features are computed for each 3-min FHR segments. The last nonlinear technique is Phase Rectified Signal Averaging (PRSA) [69]. Average Acceleration and Deceleration Capacities (*AC* and *DC*) are among the various parameters which can be derived from the PRSA curve [46]. More recently, Deceleration Reserve (*DR*) [48] was defined as the simple summation of *AC* and *DC* and it has been shown to achieve enhanced performance in detecting fetal hypoxia compared to *AC* and *DC* standalone parameters in the context of intrapartum FHR recordings. Regarding the specific implementation of these methodologies in this work, for the computation of *LZC_bin* and *LZC_ter* the factor value (p) was set to zero, whereas for entropy computation the length of the pattern (m) and tolerance (r) were set equal 1 and 0.1, respectively, accordingly with the prior knowledge on their application for fetal investigations [43], [49], [74]. On the other hand, a technical aspect that complicates the physiological understanding of PRSA-derived features is their dependence on three parameters, namely T, s, and L. Different combinations of the former parameters allow to gain insight about the ANS branches separately. In this work, AC, DC, and DR were computed considering: $T = 1$ and $s = 2$, $T = 5$ and $s = 5$, $T = 9$ and $s = 9$, $T = 40$ and $s = 1$, L was constant and equal 100.

Fetal and Maternal domain (FMd)

The evolution of fetal ANS regulation throughout pregnancy has been extensively investigated in regard to GA, sex, and various aspects [40], [74], [97]. This evidence is consistently reported among MTd, Fd, and Cd features. To address this issue, GA at the recording (GA_CTG), fetal sex (fetal_sex) and maternal age (mat_age) are included in the machine learning analyses.

Feature preprocessing

The time series of each parameter was averaged throughout the recording to derive a single set of features for each subject. At this step, features were preprocessed for outliers (Winsorization in the interval [Q1 – 3IQR, Q3 + 3IQR], where Qi is defined as the ith quartile and IQR = Q3 – Q1). Lastly, features were standardized across the entire population to obtain zero mean and unitary variance distributions.

RFB-SVM

Classical SVM

Support Vector Machines (SVM) are a class of machine learning algorithms highly exploited for the purposes of data classification and regression. As a general consideration, SVMs aim to derive models learnt on the training data able to predict the target values contained in the test data [98]. Given a training set of labelled instance pairs ($\boldsymbol{x}_i$, y_i), $i = 1, \ldots, l$ where $\boldsymbol{x}_i \in \mathbb{R}^n$ and $y \in \{1, -1\}^l$, l is equal to the number of observed pairs, n is the dimensionality of the feature space, and y corresponds to healthy/Late IUGR binary classification assigned to each participant. SVM searches for the optimal hyperplane $\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b$, which maximizes the separation margin between the two classes by solving the optimization problem in Eq. 3:

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i$$

$$\text{subject to} \quad y_i(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i, \ \xi_i \geq 0 \tag{3}$$

where the training vectors $\boldsymbol{x}_i$ are mapped into a higher dimensional space by the function ϕ. In the context of classical SVMs, such function is linear, thus the corresponding term reads $\boldsymbol{w}^T\boldsymbol{x}_i + b$, which translates into a linear separating hyperplane. $C > 0$ is the so-called penalty parameter of the error term. C controls the tradeoff between misclassification and data sparsity. Large values of C constrain the optimization procedure to derived smaller-margin hyperplane if such boundary contributes to have the training points classified correctly. Conversely, a smaller value of C causes the optimizer to search for larger margins, even if the derived hyperplane misclassifies more observations. Classical SVMs promote data sparsity given only few subjects contribute to the margin determination at the expenses of involving all the features, thus being affected by the curse of dimensionality [98]. To address the described issues, we propose to employ a more efficient kernel function: Radial Basis Function SVM (RBF-SVM) [98], and a novel feature elimination algorithm, namely RBF-SVM Recursive Feature Elimination (RBF-SVM-RFE) [99].

Radial Basis Function SVM (RBF-SVM)

The main shortcoming of classical SVMs is the constraint of describing the relationship between the class labels and the features as linear. On the opposite, the kernel of RFB-SVMs maps

observations into a higher dimensional space, thus allowing for a nonlinear relationship between observations and attributes. In this scenario, the function ϕ can be expressed accordingly to Eq. 4:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma \|x_i - x_j\|^2}, \qquad \gamma > 0 \tag{4}$$

where K is called kernel function, and the parameter γ defines the radius of influence of a given training example. Specifically, low values of γ code for far influence and a very broad decision region, whereas high values of γ result in the opposite. Furthermore, it can be shown that RFB kernel is equivalent to the linear one for some combinations of (C, γ) [100]. RBF-SVM are suitable to be employed in the presented study given the well-documented nonlinear relationship between several features and the target binary outcomes: healthy or IUGR fetuses [43], [74], [91].

RBF-SVM Recursive Feature Elimination (RBF-SVM-RFE)

Linear SVM Recursive Feature Elimination (SVM-RFE) is a largely exploited category among the wrapper models [101] which performs feature selection [102]. Wrapper methodologies are computationally demanding but they exhibit enhanced performance compared to filter approaches [103]. If SVM-RFE allows to derive an interpretable feature ranking, the same is not valid when considering nonlinear SVM (as for RBF-SVM). This relates to the fact the mapping function ϕ is unknown, thus the vector $\boldsymbol{w}$ cannot be explicitly computed and consequently cannot be used to rank features as for SVM-RFE. In this work, we employed a recent extension of SVM-RFE which performs feature elimination in the context of nonlinear SVM, namely RBF-SVM Recursive Feature Elimination (RBF-SVM-RFE) [99]. In a nutshell, RBF-SVM-RFE expands RBF kernel into its Maclaurin series. The weight vector $\boldsymbol{w}$ is derived from the series by computing the contribution made to the classification hyperplane of each feature. The algorithm allows to rank features by their relative importance starting by including all features, and progressively eliminating each of them until all attributes are ranked. Moreover, RBF-SVM-RFE allows to derive the most informative subset of feature among all possible permutations of the original set. A comprehensive and rigorous description of the algorithm can be found in [99].

Performance assessment

Performance is quantified in terms of the area under receiver-operation-characteristic (ROC) curve (AUC), Sensitivity (SE), and Specificity (SP). In the context of supervised machine learning approaches as for RBF-SVMs, it is usually required to perform the following: i) make use

of cross-validation (CV) to identify the best pair of parameters C and γ; ii) train the whole training set using the previously identified C^{opt} and γ^{opt} and evaluate the performance; iii) test the validity, replicability, and stability of the model on a new set of observation which have never been used in the training phase. The prediction accuracy obtained from the unknown observations are thought to reflect in a more precise way the classification performance of the trained algorithm.

In the context of this work, the training set was obtained by including 60% of the original dataset and the remaining 40% was used to derive the test set. The ratio between healthy and IUGR (~ 1.5:1) was maintained constant in both sets. C^{opt} and γ^{opt} were derived by performing a grid search on C and γ using cross-validation. Specifically, several pairs of (C, γ) were tested and the one achieving the best cross-validation figure of merit was chosen. Exponentially growing sequences of C and γ were employed in a the grid search framework [91], [98] along with a 10-Fold CV repeated 5 times on the training set to identify C^{opt} and γ^{opt}. In this work, SE was identified as the figure of merit to be maximize to the aim of deriving a screening tool suitable to be employed in clinical practice for the identification of Late IUGR. Despite its straightforward implementation compared to more advanced methodologies, a grid search approach has the advance of avoiding approximations by performing an exhaustive parameter search. Additionally, it can be easily parallelized since each (C, γ) is independent. On the opposite, iterative processes can hardly be run simultaneously [98].

Results

Univariate analysis

For benchmark, the performance in discriminating healthy versus IUGR fetuses for each of the previously described features was computed. Specifically, a set of logistic regression models were trained including each attribute individually. The optimal cut-off (c) for a given feature was derived by the maximization of Youden Index [104] defined as: $J = max_c(SE(c) + SP(c) - 1)$. J allows to compute the optimal c and consequently the corresponding SE(c), SP(c), and AUC(c) values. Table 8 reports the ten best performing features ranked by their AUCs. Cd features yielded the best univariate classification results, followed by MTd ones. Notably, neither Fd nor FMd attributes have a significant individual power. The selected features clearly point to the importance of more sophisticated analyses of FHR, rather than the traditional time and frequency approaches. Despite satisfactory values of AUC, the corresponding SE and SP suggest the need for a

Table 8. Univariate Performance Feature cut-offs associated with the optimal values of Sensitivity and Specificity are derived based on Youden's index maximization.

	AUC	SENSITIVITY	SPECIFICITY
LZC_bin [bits]	0.78	0.78	0.68
LZC_ter [bits]	0.78	0.88	0.57
#acc_large	0.72	0.84	0.48
AC_T9_s9 [bpm]	0.68	0.76	0.50
AC_T5_s5 [bpm]	0.67	0.66	0.58
FHR_std [bpm]	0.66	0.72	0.55
#contr	0.65	0.66	0.58
LTI [ms]	0.63	0.66	0.58
Delta [ms]	0.62	0.83	0.40
STV [ms]	0.61	0.98	0.21

multivariate framework in order to improve and balance the overall performance. Prior to multivariate classification, correlation among the all pairs of features was performed: 1) short and longer term *MTd* features were moderately correlated; 2) short term variability measured in the different domains: *ApEn*, *SampEn*, *HF_pow*, *LZC_bin*, and *LZC_ter* were highly correlated as expected given their definitions; 3) *ApEn*, *SampEn*, *LZC* parameters did not exhibit any relationship with PRSA-derived features; 4) *AC*s and *DC*s at different scales exhibited marked negative correlations; 5) *DR*s were weakly positive correlated with the corresponding *DC*s but not with *AC*s.

Multivariate analysis

Feature selection

The original set of features comprised $n = 32$ attributes, of which $n = 10$ from *MTd*, $n = 3$ from *Fd*, $n = 16$ from *Cd*, and $n = 3$ from *FMd*. The first step prior to multivariate analysis was to reduce the feature space accordingly to RBF-SVM-RFE, as described in the Methods section. The minimum and maximum allowed numbers of features for each subset were $n = 1$ and $n = 32$, respectively. Among the tested subsets, the selected one consisted of *25 retained feature and 7 eliminated*. Specifically, the features with the least squared weights were: *FHR_mean*, *II*, *#acc_small*, *#dec*, *LF_pow*, *DR_T1_s2,* and *mat_age*. This result was in accordance with the findings for the univariate analysis. Consistently, the dropped attributes exhibited poor discriminative performance as standalone parameters. Additionally, the results of the correlation

analysis for *FHR_mean*, *II*, and *mat_age* highlighted their independence from any other variable included in this analysis. *LF_pow*, and *DR_T1_s2* were highly correlated with frequency and PRSA-extracted indices, thus it is likely that their contribution in classification resulted as limited. Lastly, *#acc_small* and *#dec* did not exhibit substantial variations in the two groups.

RFB-SVM parameter optimization

Various pairs of (C, γ) were tested to identify the combination (C^{opt}, γ^{opt}) which corresponded to the maximization of the figure of merit J. Each value of J was obtained by training the model on the whole set of selected features employing a 10-Fold CV scheme, repeated 10 times. Exponentially growing sequences of $C = \gamma = 2^{-15}, 2^{-14}, \ldots, 2^{14}, 2^{15}$ were adopted as practical implementation of the RBF-SVM grid search previously described. The distribution of J as a function of C and γ is shown in Figure 7. About half of the tested pairs (corresponding to positive exponents of γ) resulted in an unsatisfactory performance (J = 0), which is mapped in the lower half of the plane displayed in Figure 7. The remaining portion of the investigated two-dimensional space is associated with more satisfactory values of SE and SP. Specifically, the optimal combination was achieved by setting $C^{opt} = 2^{12}$ and $\gamma^{opt} = 2^{-14}$. In this case, J^{opt} was equal 0.7682 and the corresponding SE and SP were equal to 0.9287 and 0.8395, respectively. Noticeably, SE associated to the reported J^{opt} is the highest achieved for the presented parameter searching. On the other hand, the best SP was equal 0.8881 but the corresponding SE was 0.7467 ($C^{opt} = 2^5$ and $\gamma^{opt} = 2^{-15}$), thus being unsatisfactory from the perspective of building a screening tool.

Performance assessment on training and testing sets

The pair $C^{opt} = 2^{12}$ and $\gamma^{opt} = 2^{-14}$ was employed as optimal set of parameters for the final adopted model. This was learned on the training set by a 10-Fold CV scheme repeated 10 times, including the restricted set of selected features. The resulting AUC was 0.9277 [0.9109, 0.9445], corresponding to SE equal 0.9287 [0.9095, 0.9479] and SP equal 0.8395 [0.8024, 0.8766]. Results are reported as mean and 95% confidence interval. A main drawback of the proposed pipeline is the opportunity for overfitting the model on the training data. The practice of testing the derived model on a validation set aims at evaluating its robustness and insensitivity to overfitting. As previously described, the validation set encompasses 40% of the original dataset with the requirement of a similar ratio of healthy versus IUGR cases. The model tested on the validation set achieved a close agreement with the one obtained on the training data. In details, classification

accuracy was 0.8462 [0.7622, 0.9094] and the associated values of SE and SP were 0.8438 and 0.8500 respectively. The resulting performance did not exhibit a drastic decrease of AUC, SE, or SP, strengthening the validity of the proposed model as a screening tool. This assumption was highlighted by additional figures of merit such Positive Predicted Value (PP V), 0.9000, and Negative Predictive Value (NPV), 0.7727.

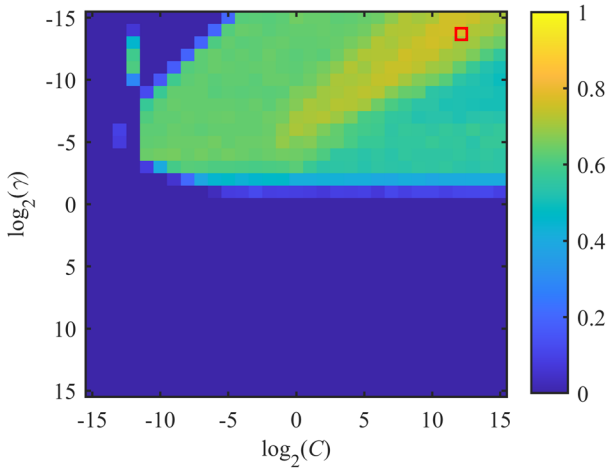


Figure 7. Distribution of J as a function of misclassification cost C, and SVM-RBF kernel parameter γ. X and Y axes are expressed in logarithmic units for better interpretability of the adopted grid search. J^{opt} is achieved by considering the pair (C^{opt}, γ^{opt}), which is indicated by the red box.

Feature importance

The main advantage in employing interpretable fetal heart rate features becomes evident for the purpose of providing meaningful machine learning findings. Specifically, the combination of heart rate attributes and RFB-SVM-RFE allows to investigate the relative influence of each attribute towards classification. The results of described approach are displayed in Figure 8 and Figure 9. The operation of ranking features accordingly to the weight vector $\boldsymbol{w}$ was also found to reflect the mean decrease in accuracy of classification when a given feature was removed from the original set employed in the training phase, as shown in Figure 8. The features that, when removed, generated the biggest decrease in accuracy, were found to belong to different domain namely: *LZC_ter* for *Cd*, *#acc_large* for *MTd*, and *HF_pow* for *Fd*. Additionally, the reported mean decrease in discriminative power appeared limited if compared to the reference accuracy achieved in the training set. At the same time, the associated SE and SP highlighted a more pronounced decrease in performance as displayed in Figure 9. The feature specific decreases in SE shown in

the left-hand panel in Figure 9 were highly correlated with the results reported in Figure 8. In fact, *LZC_ter*, *#acc_large*, and *HF_pow* accounted for the greatest decrease in SE, whereas SE stayed stable once the remaining features were removed from the model. On the opposite, it was possible to observe a dramatic drop in SP when each of the analyzed attribute was removed from the model as shown in the right-hand panel in Figure 8. The described evidences suggest that the employed features contribute differently to SE and SP. In SE, a limited number of attributes impact on the detection of Late IUGR pathology. On the other hand, it is possible to speculate that the optimal SP can be achieved only when the entire set of features is employed, thus SP appears as a less robust figure of merit in the context of the proposed model.

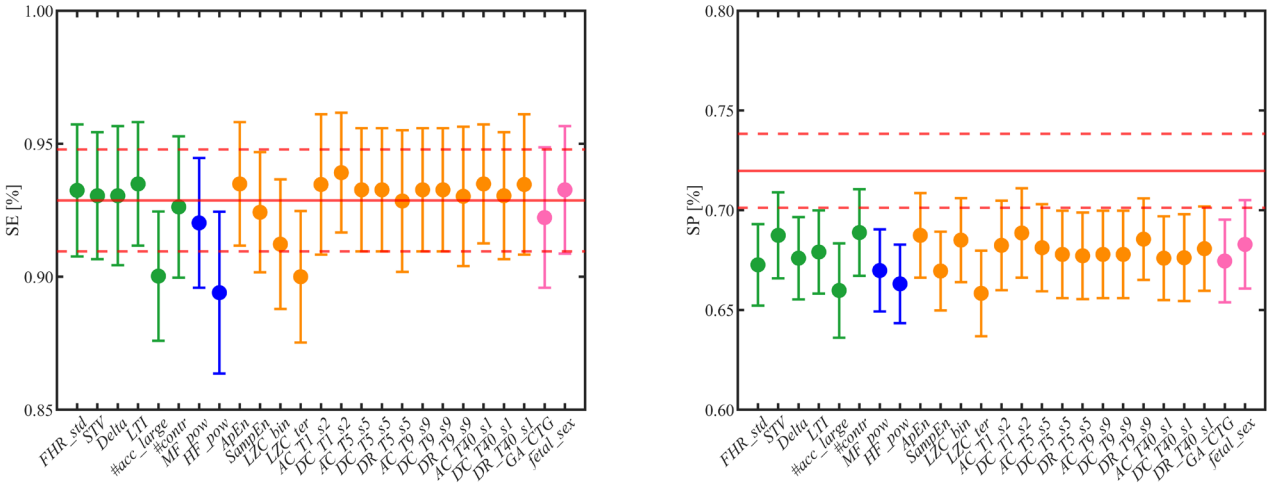


Figure 8. Left and right panels show the resulting SE and SP (mean and CI) when each feature is alternatively excluded from the trained model. The reference value of SE and SP are reported in red, solid and dashed lines corresponds to mean and CI, respectively. The displayed colors codes for the different feature domains: green for MTd, blue for Fd, orange for Cd, and pink for FMd.

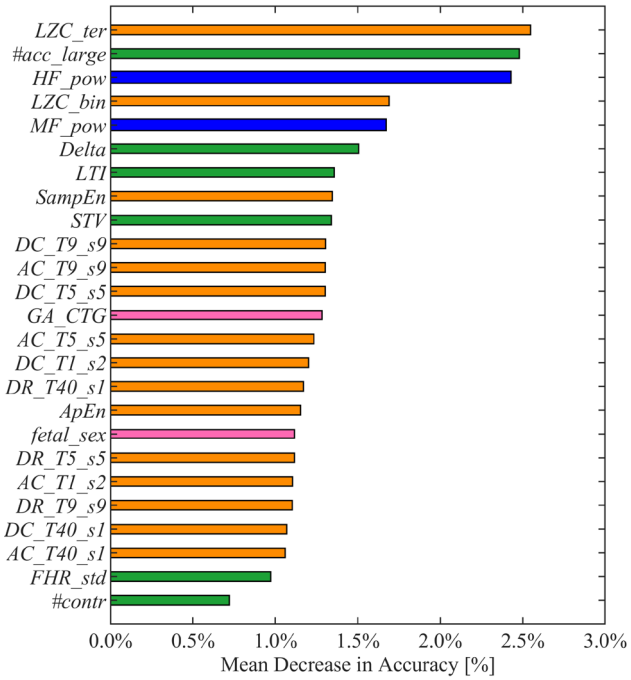


Figure 9. Feature ranked by mean decrease in accuracy in descending order from top to bottom. Mean decreases in accuracy in computed as the different between the optimal accuracy (obtained by including the entire set of selected features in the training set) and the models learnt excluding each feature alternatively. The displayed colors codes for the different feature domains: green for MTd, blue for Fd, orange for Cd, and pink for FMd.

Discussion

In this investigation, we provided evidence for the successful application of a machine learning framework for the identification of Late IUGR condition based on a single routine CTG examination. Starting from the unsatisfactory results of traditional univariate analysis (as reported in Table 8) we proposed an interpretable RBF-SVM model to be employed as screening tool in a clinical setting. The potential of Early identification of Late IUGR represents a noticeable step towards a better clinical management aimed at improving fetal outcome [34]. Discussing the model performance, the achieved values of AUC, and the associated SP and SE demonstrated the consistent ability of the proposed methodology to discriminated healthy versus Late IUGR fetuses in the training and in the validation set. This result is a consequence of the accurate tuning of model parameters (C, γ) designed to prevent overfitting. The proposed grid search for the optimal pair C, γ aimed at balancing the trade-off between the values of model variance and bias. As a general consideration, high values of the misclassification cost (C) contribute to hard margin, thus forcing

the model to a stricter interpretation of training data, potentially resulting in overfitting the training data. On the opposite, small values of gamma (γ) lead to low bias and high variance models. In this work, the selected pair consistently points to a high variance and low bias model. This translates in a separation hyperplane characterized by shaper boundaries and a strong penalization of misclassification error, suitable for the screening tool-oriented applications. Additionally, the absence of a potential bias towards overfitting is supported by the presented results on the validation set.

A crucial aspect for to the clinical application of the presented model was the possibility to interpret the data-driven results assessing the features importance. The most noticeable advantage of RFB-SVM-RFE is its peculiar insight on the individual feature contribution to classification accuracy, SP and SE. The expansion of the RBF kernel allowed a quantifiable estimation of the contribution of the different input variables, thus, to obtain a design similar to a sparse model representation [105], [106]. The feature ranking reported in Figure 9 highlighted that the combination of features from different domains is effectively enhancing the model discriminative performance compared to traditional univariate analysis. Specifically, the top three features encompassed *Cd*, *Fd*, *MTd* domains respectively. This finding supports the notion that IUGR condition is effectively impairing the fetal ANS under different aspects, thus a comprehensive set of attributes are required for an accurate determination of such pathological condition. Moreover, the presented methodology allowed to evaluate the contribution of each feature in terms of SE and SP as reported in Figure 8. If the SE contributes appeared moderately distributed among the features included in the model, this was not verified for SP. This figure of merit achieved the optimal performance only when all the selected attributes were included in the SVM framework. It is possible to speculate that the described behavior is a consequence of the grid search design. Specifically, the requirement of SE maximization allowed to achieve an adequate SP at the expense of its robustness.

An additional advantage of RFB-SVM is the opportunity to define the relationship between features as nonlinear thus, it allows to overcome the limitation of linearity imposed by traditional SVM approaches. At the same time, despite increasing the overall complexity of classification if compared to more traditional SVM implementations, the radial kernel tuning is on average of reduced complexity with respect to polynomial kernel given the fewer hyperparameters to be optimized. Lastly, RBF kernel is mathematically more stable in contrast to polynomial kernel

which tends to converge to either infinity or zero for larger degrees [98]. The proposed implementation has the advantage to generate an interpretable model without the requirement of explaining the relationship among the inputs as linear and/or strictly focusing on main effects only. Additionally, this methodology has the advantage of being highly flexible. A disadvantage of the technique is compared to black-box models is that the computational load and time largely increases with the number of inputs as several independent sets must be considered in combination with their interactions [105], [106]. While the proposed methodology is exclusively focused on linear and bivariate interactions, it holds the potentials to be extended to incorporate higher order interactions. At the same time, despite sparse modeling would theoretically apply to higher-order terms, this operation would add non negligible complexity to the method. The relationship between the degree of interaction terms and model intelligibility is topic of interest for further work.

To our knowledge, this work is the first attempt towards a CTG and quantitative feature-based discrimination of Late IUGR condition. Previous research mainly focused on the investigation of animal models [107], analyses of metabolic [108] and doppler profiles [35] of chronic hypoxia in the fetal period. Nevertheless, the underpinning and widely reported consequence of long-lasting oxygen deprivation is responsible for a delay in the maturation of the branches of ANS and their subsequent integration with the central nervous system (CNS). The impairment in ANS maturation was consistently found in this investigation by various quantitative CTG-derived parameters which have been extensively associated to the fetal ANS modulation throughout pregnancy as standalone features [43], [74]. In comparison with previous machine learning-derived and univariate results by our group in different populations of Early IUGR [44], [46], [49], it is possible to observe a consistent discriminative power of features *LZC*, *HF_pow*, and *LTI*. Specifically, the average value of each feature was greater in the control group versus Late IUGR fetuses. On the other hand, we also report an enhanced classification contribution of *SampEn*, which outperformed *ApEn*. Lastly, in the described Late IUGR population, short scale ($T = s = 5$ and $T = s = 9$) PRSA-extracted features were characterized by a greater discriminative power compared to global ones ($T = 40$ and $s = 1$). The reported finding are in accordance with the univariate results and support the hypothesis of an impaired fetal beat-to-beat responsiveness regulations in the context of nutrient restriction and chronic hypoxemia [46], [48].

Conclusion

This contribution aims at promoting the application of machine learning methodologies in the context of fetal and perinatal medicine, following the growing trend of the artificial intelligence application in medicine [33], [109]. The presented approach demonstrated the reliability of an SVM inspired framework, encompassing the automatic selection of a subset of CTG-derived features, a satisfactory classification performance in terms of AUC, SE, and SP in both the training and validation sets, and an interpretable set results suitable to be translatable in the clinical environment. Findings reported in this investigation support the importance of multivariate approaches to investigate the variety of implications resulting from a pathological condition such as Late IUGR.

Despite satisfactory and promising classification performance, improvements may be envisioned under various aspects. First, the inclusion of additional features such as the ones inspired to multiscale and fractal analysis might further contribute to classification accuracy as reported in the context of intrapartum [91]. Second, the performance of diverse machine learning approaches as well as deep learning methodologies should be investigated and compared to RFB-SVM-RFE. Lastly, the validation of the presented approach should be carried on external datasets to ultimately test the model performance as a function of different reference values of the input features. Additionally, it would be relevant to evaluate the validity of the proposed model in the context of the early insurgence of the pathology. A recent dataset of FHR indices extracted from a population of Early IUGR fetuses can be found in [9].

Chapter 2

The negative effects of excessive prenatal exposure to alcohol (PAE) and tobacco (PTE) during gestation on fetal development, birth outcomes, and subsequent neurobehavioral development have been extensively documented [110]. PAE can lead to fetal abnormalities and increased risk for multiple adverse outcomes such as preterm birth, stillbirth, low birth weight, small for gestational age (SGA) and late intrauterine growth restriction (IUGR), sudden infant death syndrome (SIDS) [111], and neurodevelopmental disorders [112]. Despite continued recommendations against maternal alcohol consumption during pregnancy, some mothers-to-be continue to do so. According to the most recent surveillance data from the Centers for Disease Control and Prevention, 1 in 10 pregnant women from 18-44 years of age reported consuming alcohol and approximately 1 in 33 pregnant women reported binge drinking [113] (> 4 drinks/occasion) within the last 30 days. Moreover, approximately 50% of pregnancies are unplanned [114], therefore risk for prenatal exposure during early gestation before the recognition of pregnancy can also lead to downstream effects on perinatal development.

During gestation, the developing fetus is vulnerable to exposures in the intrauterine environment which can influence subsequent development. Ethanol is a central nervous system depressant which crosses the placenta and can directly affect fetal brain development through several potential mechanisms. PAE decreases cortical/sub-cortical activity and causes central nervous system (CNS) damage [115], functional disruptions of CNS synaptic connections [116], and changes in myelination. PAE also alters synaptic plasticity by changing long term depression and potentiation [116] thus interfering with normal synapse formation and brain function. Prenatal exposure to maternal smoking also has the potential to affect the developing fetal brain. Acetylcholine receptors, in which nicotine is a receptor agonist, play a major role in fetal central nervous system development such as in neuronal differentiation and synapse formation [117], [118]. Activation of nicotinic acetylcholine receptors affect both spontaneous neural activity and neuronal survival [115], [118]. PAE and PTE are recognized as a perinatal modifiable risk factor responsible for reduction in birthweight, increase probability of preterm and late preterm birth, morbidities and mortality, and intrauterine fetal growth restriction (IUGR) [119], [120]. The most common etiology of the reported complications is thought to originate from a reduction in blood flow and decreased perfusion to the placenta associated with subsequent placental injury [121].

By crossing the placenta, alcohol has been hypothesized to be responsible for vascular lesion development. It has been reported that alcohol exposure results in placental vasoconstriction and lower placental weight [16], [122]. Most studies largely corroborate associations between alcohol and adverse pregnancy outcomes, but there is a lack of consensus regarding the relationship between maternal alcohol use and placental abnormality.

Regarding maternal cigarette smoking, nicotine and its stable metabolite cotinine have direct toxic effects, as these compounds readily cross the human placenta [123]. As an example, smoking decreases the flow of uterine blood to the placenta [124] through mechanisms that include vasoconstriction. As a downstream effect, the placentas of mothers smoking during pregnancy show morphological signs of reduced perfusion [125]. As a consequence, women who smoke show dramatic differences in morphology as compared to gestation-matched control placentas [126]. Taken together, these converging findings support the association between exposure to substances during pregnancy and placental abnormality.

The paucity of studies investigating perinatal exposures is closely linked to the inaccurate and imprecise available methodologies for the quantification of maternal substance consumption. This evidence is consistently reported across different scenarios of investigation. In the context of pregnancy, alcohol and tobacco consumptions are collected intermittently throughout gestation. As a consequence, techniques aimed to retrieve missing information are expected to play a crucial role towards improving the methodologies available in the field of exposure data. Moreover, dealing with data missingness on highly dimensional data requires advanced machine learning techniques rather than the simple interpolation or purely statistical methods. Given the present limitations, it becomes clear the fact that a quantitative and rigorous investigation of the effects of substance abuse in the perinatal period is still needed. Furthermore, perinatal and subsequent exposure are often studied independently one each other, whereas a more comprehensive data driven framework is expected to gain better understanding of the diverse pathways responsible for abnormalities in the development of autonomic and central nervous systems. Furthermore, the analysis of physiological signal in the context of exposures is expected to inform on the relationship between these early stressors and later neurodevelopmental outcome, via both direct and indirect pathways.

In this chapter, we will present a set of machine learning approach for the imputation and clustering of exposure data. The imputation of longitudinal data, and more in general missing information is a complex and multiparametric problem. The most common strategy assumes that each missing value in a given dataset is equally likely to be missing [19]. The former assumption is often unmet in the context of exposure collection throughout pregnancy as pregnant women might only be queried at specific time points which are likely to overlap with clinical appointments. As a consequence, data are usually missing by design and thus leading to biased estimates of substance consumptions which may ultimately confound methodologies aimed to investigate the downstream effects [20]–[23]. To address this limitation, we proposed a novel adaption of the KNN algorithm for the imputation of exposure data not missing at random bur rather by design. This assumption applies to the dataset analyzed in this report and it is consistent with previous work describing longitudinal trajectories of exposure during pregnancy [24], [25]. The remarkable design obtained with the presented analysis allowed to perform imputation on two different populations by means of the same framework.

The first part of Chapter 2 describes the work illustrated in [127] which focuses on the novel adaptation of the K-Nearest Neighbor (KNN) toward the imputation of longitudinal exposure data. The second half of Chapter 2 provides evidence of the utilization of finite mixture model to cluster the derived estimates of alcohol and tobacco consumption as reported in [128].

Keywords – **Prenatal Exposure to Alcohol and Tobacco, Unsupervised Machine Learning, Imputation, Clustering, Longitudinal exposure data**

The K Nearest Neighbor Algorithm for Imputation of Missing Longitudinal Prenatal Alcohol Data

Introduction

Accurate assessment of timing, frequency, and magnitude of prenatal alcohol exposure (PAE) in longitudinal research studies is necessary for obtaining unbiased assessments of effects of such exposure on fetal and neonatal outcome. Virtually, all existing evidence of PAE's effect is based on retrospective studies and it is often compromised by insufficient power and recall bias. Despite the importance from a public health point of view, there are currently no robust biomarkers for assessing timing and amount of alcohol exposure in the perinatal period. Thus, we often remain reliant on maternal self-report of intake. Aside from issues associated with the accuracy of self-report, there are other methodological challenges in measuring alcohol exposure in longitudinal studies [129], [130]. Recording daily intake, while providing a temporally complete set of values, involves significant participant burden and it is likely to impact consumption behavior, thus ultimately resulting in a biased estimate [131]. As a consequence, in many studies, alcohol consumption data are sampled at various times throughout pregnancy [132]. However, even when data for the specific collection time-points are complete, there is frequently missing information about intake during the intervals between samples. Addressing this missing data problem becomes critical when the exposure metrics of interest are both timing and magnitude throughout pregnancy [21].

The impact of missing data on the validity of estimates largely depends on the reasons data are missing [19]. For example, pregnant women of low socioeconomic (SES) background are more likely to access antenatal care late in pregnancy, enroll late in research studies, and, therefore, have more missing data early in pregnancy [133]. In addition, women often modify their consumption behavior following pregnancy recognition, which can happen at varying times during the first months of pregnancy. While some women tend to stop or reduce drinking immediately upon pregnancy recognition, some heavy drinkers continue to binge in the first trimester or continue heavy drinking throughout the pregnancy [21]. The accuracy of measures irrespective of the presence of missing data, such as the number of drinks consumed only on drinking days, may also provide biased overall estimates depending on when participants are interviewed. Accordingly,

new approaches for managing the missing information problem are needed and consequently addressing data imputation.

The Safe Passage Study conducted by the Prenatal Alcohol and SIDS and Stillbirth Network (PASS) was a prospective investigation on the effects of alcohol exposure on multiple fetal and infant outcomes in Cape Town, South Africa and the Northern Plains, USA [134]. In this study, alcohol data were collected using a modification of the Timeline Followback Method (TLFB) [25], in which mothers recorded drinking data on their last known drinking day and then, for the 30 days prior. While this method was deemed the best self-report system available, the approach necessarily generated a variable amount of missing data depending on the various patterns of alcohol consumption.

In the described context, the goal was to impute the drinking values on missing days using a machine learning algorithm called K-Nearest Neighbor (KNN). KNN imputes missing values using pattern recognition without any assumption on the probability density function underlying the data under investigation [135]. The KNN algorithm has been used for the imputation of missing data of diverse nature in several research studies in the healthcare and bioengineering fields [136], [137]. In this paper, we provide the methodological details of the specific application of the KNN method for PASS exposure data and the validation of these results. An accurate and reliable estimate of alcohol consumption during pregnancy is expected to provide critical insights on the effect of these substances of abuse on different domains of investigation. The utilization of KNN aims to overcome the limitation of regression models. Traditional linear approaches tend to distort the marginal distributions of covariation. As a result, extreme values contained in the tails of the distributions tend to be unrepresented thus, potentially leading to a bias in the derived estimates and ultimately to model misspecification. On the opposite, KNN imputation approaches are donor-based methods where the estimates for the imputed values are directly obtained from the measurements in the analyzed dataset. This aspect is of particular relevance for the presented investigation as alcohol data tend to be skewed towards null exposure and only a minimal portion of the population is attribute to any level of drinking throughout pregnancy. The results of the former investigation will be presented in the following sections of this chapter.

Materials and methods

The Safe Passage Study

The Safe Passage Study (PASS) was a prospective study of a cohort of pregnant women and their infants evaluating the role of prenatal alcohol exposure on incidence of adverse pregnancy outcomes including stillbirth, sudden infant death syndrome (SIDS), and fetal alcohol spectrum disorders (FASDs) of the surviving children. Between August 2007 and January 2015, 11,083 pregnant women were enrolled from antenatal clinics in Northern Plains (NP), USA and Cape Town (SA), South Africa. Women were eligible to participate in the study if they were pregnant with one or two fetuses, aged 16 years or older, were at gestational age 6 weeks or later at recruitment and spoke English or Afrikaans. Women were followed throughout the pregnancy and 1 year postnatally. Data on socioeconomic status, demographic, obstetric and medical history, periconceptional drinking and smoking were collected at the enrollment interview. Information on subsequent drinking during pregnancy was updated in study visits following enrollment. Written informed consent was obtained from all participants. Ethical approval was obtained from

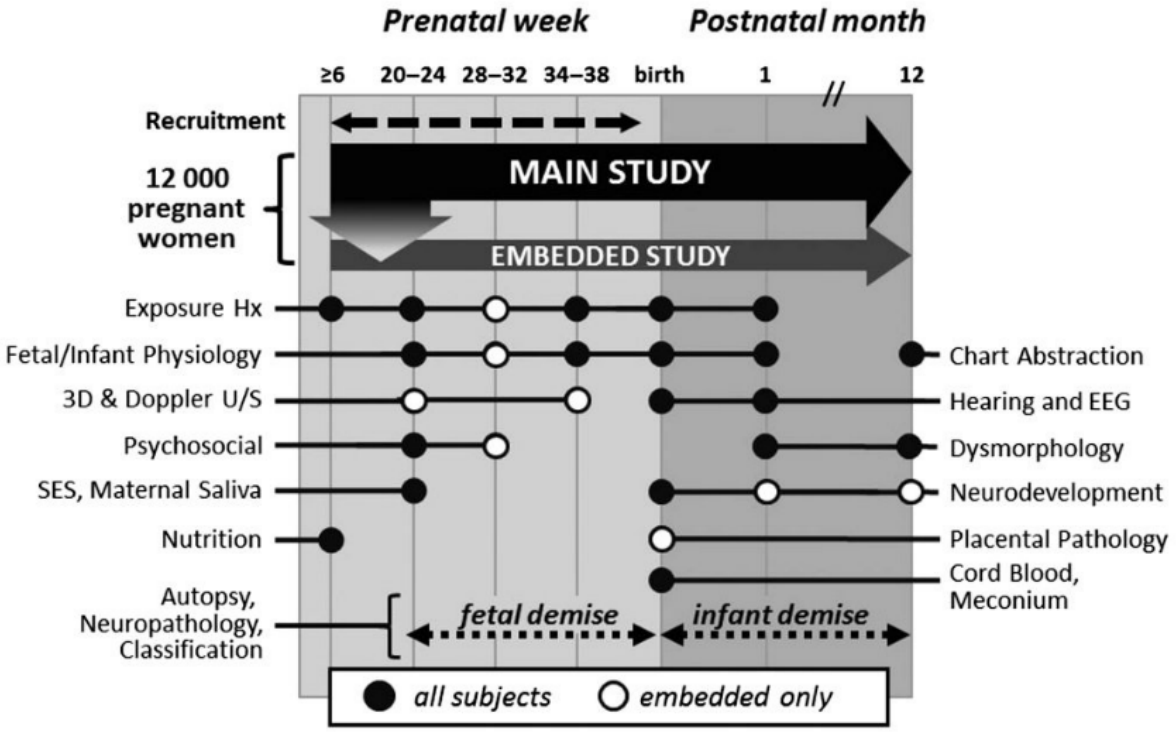


Figure 10. A schematic view of the unprecedented variety of data collected in PASS. Black dots indicate items collected for the entire sample size. Conversely, white dots code for assessments collected for a subsample of participants . Data encompassing exposure are labeled as Exposure Hx in the chart.

Stellenbosch University, Sanford Health, the Indian Health Service and from participating Tribal Nations [134]. Figure 10 reports a schematic view of the data belonging to several domains of investigation collected in PASS.

Alcohol data collection method and missing data

Alcohol exposure data were collected using a modified validated TLFB [25], which required participants to report their drinking behavior on each day ±15 days from the last menstrual period (LMP) and, at each study visit, the 30 days prior to the last known drinking day. Data were collected on the types of drinks, number of drinks, size of the containers, amount of ice in the drink, how many people drinks were shared with, and duration of the drinking episodes. These data were then used to estimate the total amount of alcohol consumed and number of standard drinks on each reported occasion [138]. Data on drinking were collected during 1-4 prenatal study visits and 1 visit postpartum. As previously shown in Figure 10, participants had a different number of prenatal visits depending their time of enrollment in the study.

Due to the nature of the modified TLFB data collection design, the number of days with missing data varied by participant as a function of the time of enrollment and number of attended visits. The number of days with missing drinking information also varied for each participant depending on the recentness of their drinking. Figure 11 shows different scenarios of how such variation emerged during the time period between LMP and the recruitment visit depending on when the last drinking day occurred. Participants who did not drink, or whose last drinking day was prior to their LMP had no missing data (Figure 11, panel A). Participants who drank but quit drinking within 30 days of the last collection period, had less or no missing data (Figure 11, panel B). Participants who continued to drink, and who reported drinking information 30 days closest to the interview date, had missing information prior to the 30-day period of reported drinking (Figure 11, panel C). In this example, if Subject Z drank often, and possibly at a higher volume, she would have a greater number of missing days than women who drink less often. Thus, a simple summation of drinks over the days will reflect a lower amount than the actual consumption and analysis using this exposure metric will result as biased.

The KNN algorithm

To address the described exposure data missingness by design, we proposed the use of KNN. KNN is a non-parametric unsupervised machine learning algorithm which can be utilized

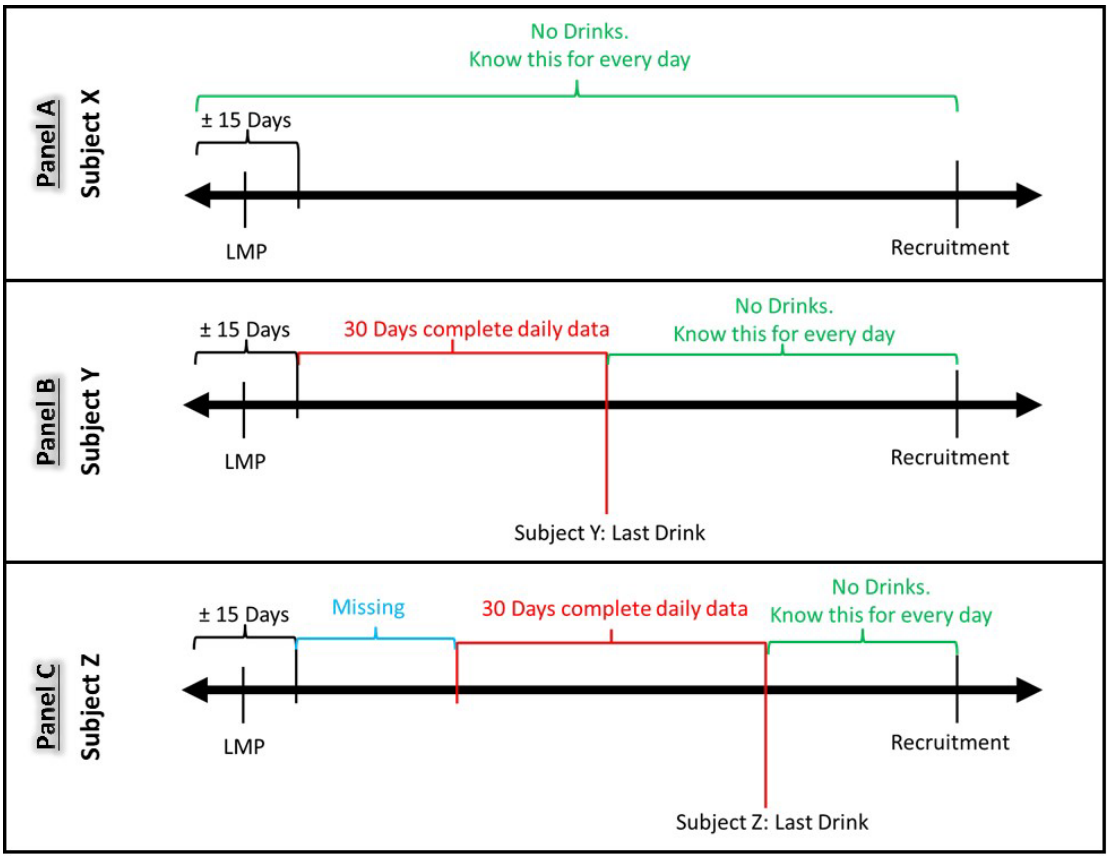


Figure 11. In the presented scenario, let's assume that subjects X, Y and Z were enrolled at the same gestational ages for their respective pregnancies (denoted as Recruitment on the x-axis). The alcohol consumption of Subject X is depicted in Panel A. Participant X is a non-drinker given no alcohol consumption is reported in the time interval from LMP and recruitment. Both Subject Y (Panel B) and Subjects Z (Panel C) did report at least an event of alcohol consumption in the same time interval. Neverthless, the timing of alcohol intake is different for the two participants, thus resulting is the absence (Subject Y) and presence (Subject Z) of data missingess. Considering Subject Y, the time interval between last alcohol intake and LMP is less or equal 30 days, thus there is no gap in the alcohol consumption information, resulting in a complete timeline from recruitment back to LMP. On the contrary, Subject Z reported her last drinking event more recently with respect to Subject Y, thus the interval between last alcohol consumption and LMP is greater than 30 days. In this latter case, we have data missing by design of the assessment instrument.

to impute missing information of a given participant based on the similar observations in the same dataset. To provide a more comprehensive understanding of the adopted approach we propose the didactical example reported in Figure 12.

Let's suppose to consider three subjects, labeled as *q*, *r* and *s* for which complete estimates of alcohol consumption were collected (e.g., number of standard drinks per day) on three different days (*d*) during pregnancy. The magnitude of alcohol consumption is identified by the values reported on the *x*, *y*, and *z* axes. On the contrary, the subject denoted as *p* is missing for alcohol consumption on the third day (displayed by the *z* axis). As a first step, data from each of the four subjects, were mapped to *d* - 1 dimensional space using data for two of the available days. In Figure 12, subjects *p*, *q*, *r* and *s* were mapped to the two-dimensional space for which all four subjects

have complete data and were labeled p', q', r' and s', respectively. To find the degree of similarity between the drinking habits of subject p and another subject, e.g. subject r, we connected p' and r' to the origin O and measured the angle between $p'O$ and $r'O$. Based on the proposed definition, the smaller the angle, the greater the similarity between two adjacent participants. In Figure 12, the angle between $p'O$ and $q'O$ is zero which means that p and q have exactly the same drinking pattern but a different magnitude. Specifically, p consumed double the amount of alcohol consumed by participant q in both day 1 (x axis) and day 2 (y axis). Considering the relationship between participant p and participant q, the angle $p'O$ and $r'O$ is larger than that between $p'O$ and $q'O$ which means that r is less similar to p than q is to p. On the other hand, the angle between $p'O$ and $r'O$ is smaller than the one between $p'O$ and $s'O$ meaning that r is more similar to p than s is to p. The applied definition of the nearest neighbors (K) adopted is based on the concept of angular distance as previously defined. For any given value of K (> 0), the k nearest neighbors of p are those K observations in the dataset that form the K smallest set of angles with p. These K participants have drinking patterns most similar to that of p and they are expected to provide a meaningful imputation of the missing data values. From a computation point of view, it is more convenient to compute the cosine distance between a set of observation, thus using this measure as a proxy of the actual angle between two alcohol trajectories the small-angle approximation. To generalize the proposed framework for any given pair of observations, let's consider the scenario in which participants a and b have drinking data for h different days, denoted by $[a_1, a_2, \dots, a_h]$ and $[b_1, b_2, \dots, b_h]$, with $h \leq d$, and an arbitrary number of missing exposure data for a. The cosine distance between to observations is defined according to Eq. 5:

$$\cos([a_1, a_2, \dots, a_h],\ [b_1, b_2, \dots, b_h]) = \frac{\sum_{i=1}^{h} a_i b_i}{\sqrt{\left(\sum_{i=1}^{h} {a_i}^2\right)\left(\sum_{i=1}^{h} {b_i}^2\right)}} \tag{5}$$

Thus, the smaller the angle between two observations, the larger the cosine distance between them within the hypothesis of dealing with angle values bounded between 0° and 90°. This condition is always verified in the proposed context as drinking data (regardless the employed estimator) are lower bounded to zero. As a result, the K nearest neighbors of a are those K subjects in the dataset that have the largest cosine distances from a.

Once the K nearest neighbors of a are identified, the weighted average of the drinking data of these neighbors for the day in which a's drinking data are missing is taken as the best estimate of the missing data. The obtained estimate represents KNN's prediction of the number of drinks subject a had on the h-ith drinking day. The weight given to each neighbor, is of the form $w_r = \sigma_r \delta_r$, where σ_r is a scaling factor and δ_r is a distance factor. We set σ_r equal to the ratio between the Euclidean distances of a and b at a given day h and the origin, which translates to the 1-norm distance in the d-dimensional space. The scaling adjustment is needed to account for similar drinking patterns characterized by different magnitude. Coming to the discussion of δ_r, it is defined as the square of the cosine distance between p and r. The distance factor assures that the neighbors closer to p have more influence on the predicted value than the ones further away in the d-dimensional space.

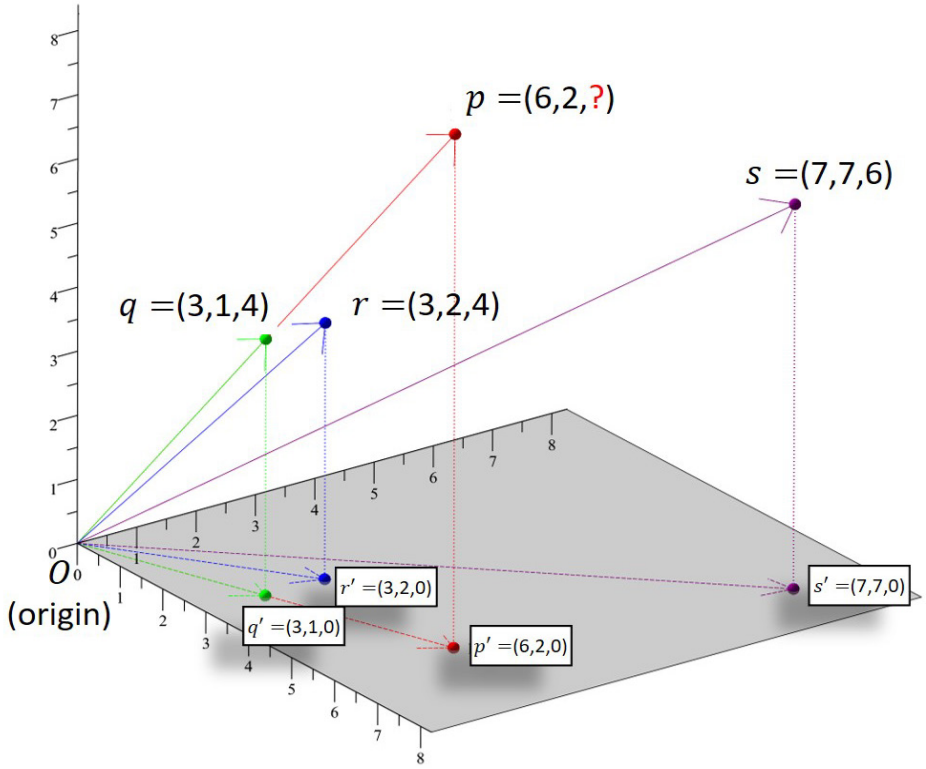


Figure 12. Participants p, q, r and s are mapped to points p', q', r' and s', respectively, in the 2-dimensional space in which no missing data are present. In this reduced feature space it is possible to compute the cosine distance among participants, thus identify the neighbors of participant p.

Data preparation

The time interval covered by the proposed imputation methods ranged from day -15 (2 weeks prior to LMP) to the maximum possible pregnancy length set to day 310 (maximum possible

pregnancy length in the available data). We used the distance between a fixed date before the start of the study (Saturday, January 1, 2000), and the beginning of pregnancy (i.e., day -15) to find the day of the week the pregnancy started. This solution was employed to temporally align each subject prior to computation of the cosine distances. The rationale behind this alignment is that the drinking behavior often varies by the day of the week [139]. We also Winsorized the outlier drinking values at 3 SD to reduce the impact of extreme values. Despite the proposed method focused on imputing the missing data for the entire population under investigation, we proposed to introduce exclusion criteria to avoid unreliable alcohol consumption interpolations. Specifically, we excluded participants with more than 200 days missing thought pregnancy or participants who did not have any data in the first trimester.

Assessment of performance and validation

To provide validation for the proposed methodology, we randomly deleted segments of length equal to 5 or 15 consecutive days from a subset of participants with no missing data in trimester 1. The first trimester was selected for validation because the proportion of women drinking and the magnitude of their drinking is highest in trimester 1, particularly for the days before pregnancy recognition. To evaluate imputation performance, we computed the root mean squared error (RMSE) for the predicted ($\hat{y}_i$) and the reported (y_i) alcohol consumption in for a given deleted segments as reported in Eq. 6

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{6}$$

where n is the length of the considered segment. Then, we calculated the overall prediction accuracy of drinking status as proportion of accurate classification. Additionally, we calculated absolute differences between actual and predicted values for drinking and non-drinking days separately.

Results

Description of missing data

Participants contributed a total of 3.2 million person-days of observation in the study, of which 0.36 million (11.25 %) person-days were missing. Based on the data collected using the TLFB method about 45% of the participants (n = 4,848) had alcohol use data for every single

day of their pregnancy while the remaining 55% (n = 5,925) had at least 1 day of alcohol-use data missing. Among the study participants 62% (n = 6,679) were drinkers, i.e., consumed at least 1 drink during pregnancy. Figure 13 shows the distribution of missing days across the study sites. On average, NP site had fewer missing data, with over 50% of the participants having 30 or fewer days of missing data. Most of the missing data in the South Africa site are from the early trimesters which largely reflects later enrollment at that site, while the majority of missing data in the NP site was found in the 3rd trimester (data not shown).

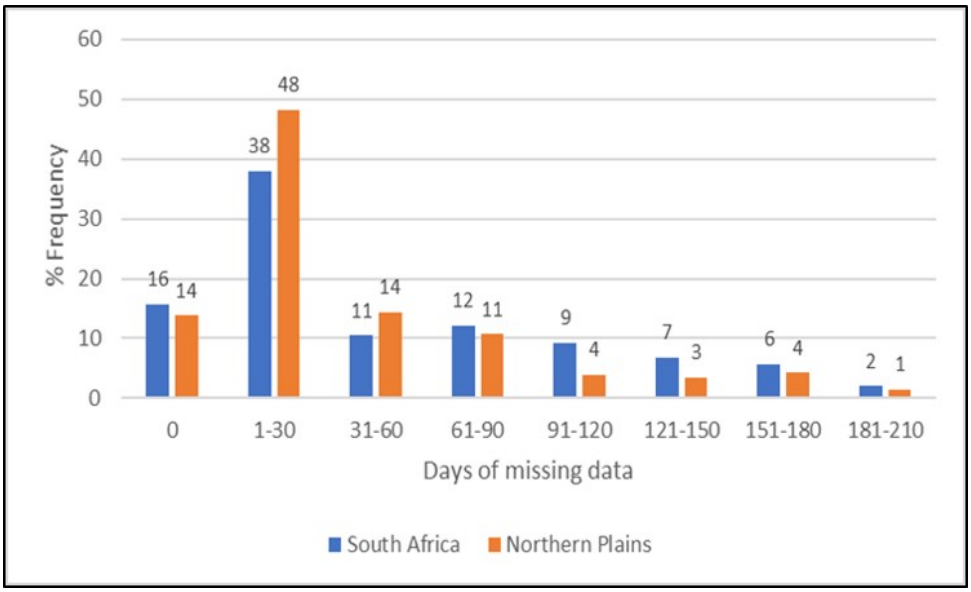


Figure 13. Distribution of missing data by study site considering 30-days binning.

Length of reference segment

The largest possible reference segment in the PASS dataset is 325 days, this corresponds to the maximum length of the pregnancy (310 days) plus 2 weeks before pregnancy. However, as mentioned in a previous section, women with complete data were more likely to be nondrinkers or light drinkers, hence using them as reference could produce an underestimate of true drinking values. The tradeoff between selecting a larger or smaller segment size is that smaller segment sizes (e.g., 7 days) allows a greater number of segments to be included as reference. At the same time, the shorter a given segment becomes, the less accurate is the algorithm's characterization of specific patterns of drinking. We determined that a reduction of segment sizes below 55 days did not increase the numerosity of available reference segments significantly, as shown in Figure 14. Thus, a segment size of approximately 2 months (55 days) retained the majority of the

participants in the reference pool without significantly diminishing the ability to identify their drinking patterns.

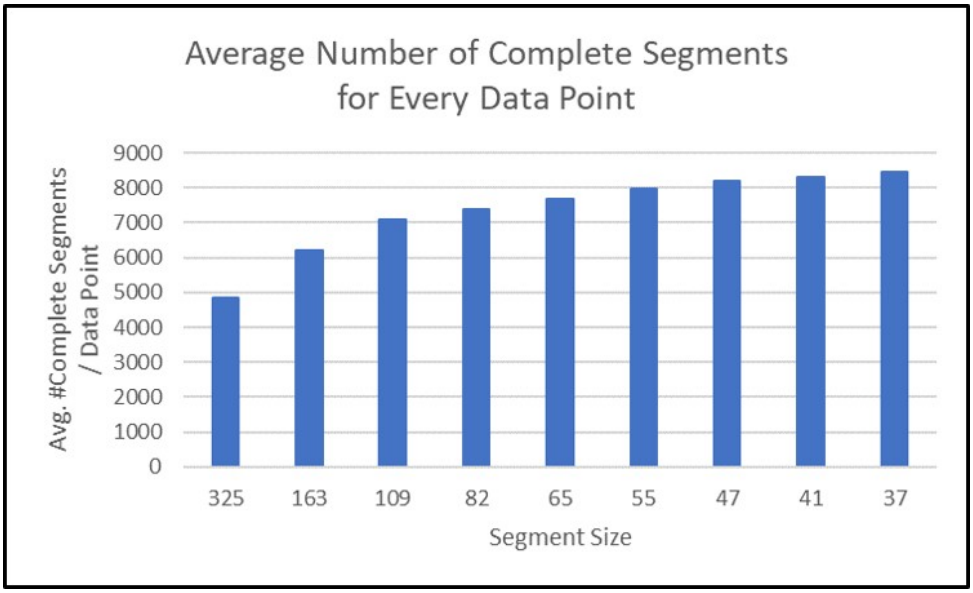


Figure 14. Numerosity of available segment to be employed as reference as a function of segment length.

Number of neighbors (K) and performance evaluation

To identify the optimal number of neighbors for carrying over the proposed imputation methodology, we varied the number of neighbors K from 1 to 10. Panel (A) of Figure 15 displays the distribution of RMSE as a function of K considering the overall population obtained by combining the SA and NP sites. For the prediction of nondrinking segments, K= 1 provided the lowest RMSE, whereas increasing values of K provided increasing lower RMSE estimates when considering the prediction of drinking segments. Results are show in the left-hand side and right-hand side of Panel (A) of Figure 15, respectively. Although the mean RMSE value for drinking participants decreased for increasing values of K, mean RMSE did not decrease substantially considering for value of K > 5. Based on the reported considerations on this figure of merit, we concluded that K = 5 provided reasonable trade-off. The classification accuracy obtained with these specifications is shown by the confusion matrices in Panel (B) of Figure 15. We found the proposed approach resulted in exact predictions of drinking status for 76% drinking segments in the combined sample. Additionally, the algorithm predicted nondrinking status accurately in 74% and 58% of the deleted segments in South Africa and Northern Plains respectively (data not shown). The discrepancies between the actual and the predicted amount of drinking for the remaining participants are reported in Panel (C) of Figure 15. Employing K=5, the described

technique predicted segments within ± 1 drinks for 80.5% of deleted segments in SA and 70.6% in the NP. The reported results highlighted the close agreement between the derived estimates even in the context of not exactly predicted segments.

Figure 16 shows the comparison between average number of drinks per participant in LMP, trimester 1, trimester 2, trimester 3 before and after imputation. Following imputation, the mean number of drinks in SA increased by 2 drinks in first trimester, whereas the increase for the Northern Plains sites was below 1 drink in the first trimester. Overall, the magnitude of increase

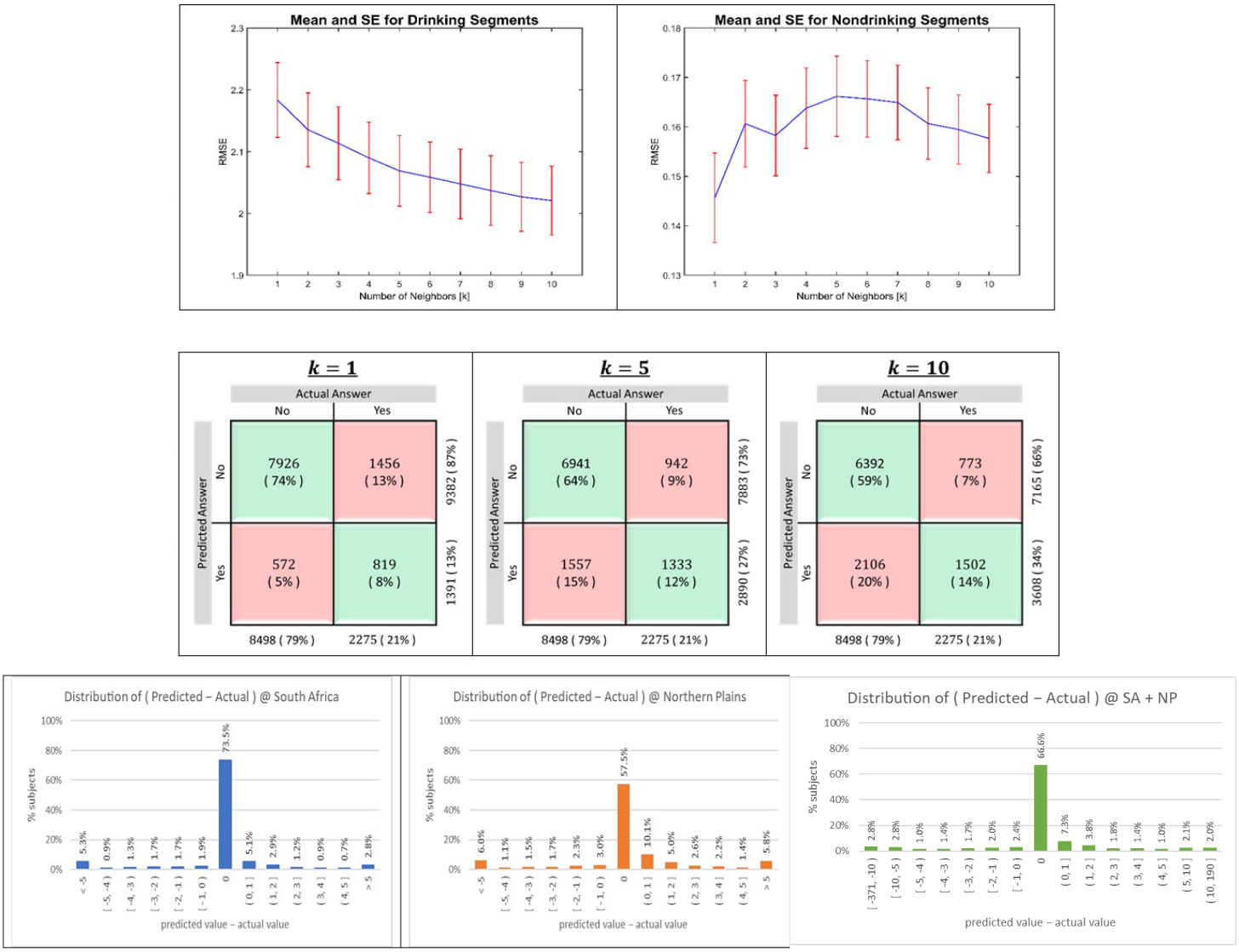


Figure 15. Summary of proposed algorithm performance. Top panel – RMSE as a function of the number of neighbors (K) considering non drinking (left) and drinking (right) segments. Middle panel – agreement between the actual and the predicted alcohol consumption. Confusion matrix for K = 5 represents the chosen model (middle). Bottom – distribution of the absolute error (in terms of number of standard drinks) between the imputed and actual values of alcohol exposure.

in terms of mean standard drinks in SA was higher than that in NP. The reported behavior could be explained by the lower amount of missing data in the NP site. Consequently, it is possible to speculate that imputation had a more significant effect of increasing average exposure in SA site. This finding can also be explained by the higher rate of exposure in the former cohort.

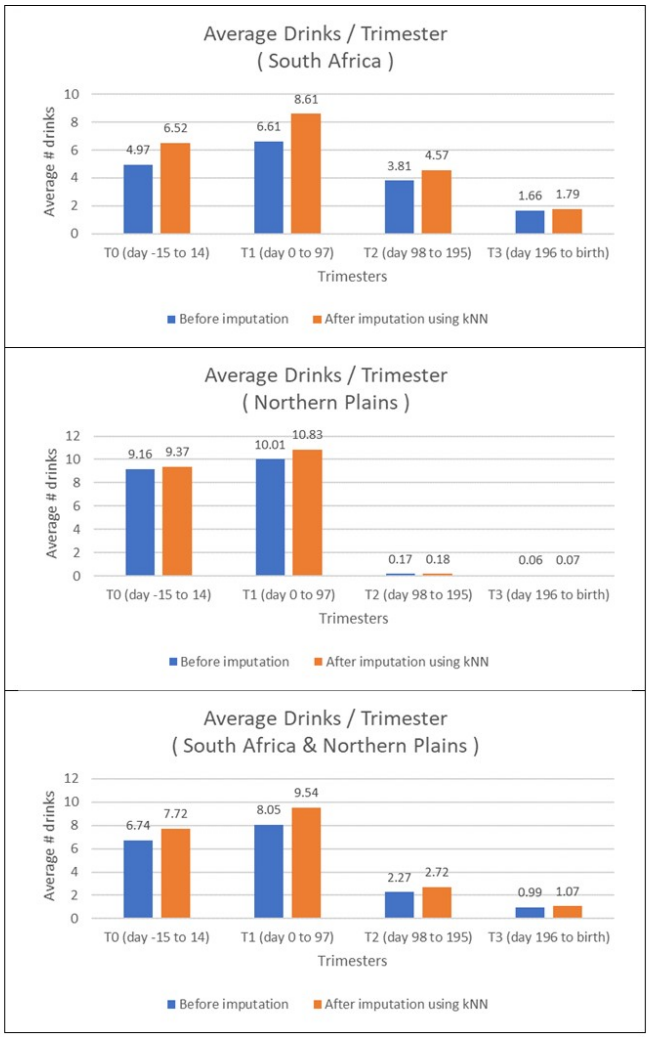


Figure 16. Comparison between the average number of drinks per trimester pre and post imputation.

Discussion

The objective of this study was to describe the application of a machine learning algorithm to impute missing daily alcohol exposure among pregnant women whose data were collected at various time points during pregnancy. The pattern of data missingness was an inherent consequence of the assessment methodology given pregnant women were asked about recent alcohol consumption during their prenatal visits, thus there were more missing data among recent drinkers. Literature defines this scenario of data missingness as either *by study design* or

equivalently *not at random*. To deal with the presented issues, we implemented an extension of a KNN algorithm for the imputation of missing alcohol consumption data in a longitudinal study among pregnant women. Moreover, we developed a validation schema to account for the absence of a 'typical/classic' reference group, i.e., training data with no missing days, a condition often met when dealing with frameworks for data imputation. To our knowledge, the present study is the first to describe this methodology. Validation of our approach showed high agreement between actual and predicted drinking values.

There is a paucity of studies dealing with techniques for the imputation of missing data as well as extensive statistical validation of the obtained results. This is especially true in the context of alcohol and drug use research studies [140]. Published work has not yet reported the performance of any machine learning method for imputation of missing alcohol data. In a simulated dataset, Hallgren et al. compared methods of imputation including complete case analysis, last observation carried forward, the worst-case scenario of missing equals any drinking or heavy drinking, multiple imputation (MI), full information maximum likelihood (FIML) and concluded that MI and FIML yielded less biased estimates [141], [142]. A recent study by Grittner et al. also found MI produced least bias based on their work in a longitudinal study in Denmark with five alcohol measurements over a period of five years [143]. However, all of the above described methods have been reported to produce an underestimate of the actual drinking level. In addition, MI models are originally recommended for imputation of a single value per subject [144], a condition often unmet in missing exposure data. As an example, to impute irregularly spaced missing longitudinal data as for the dataset analyzed in this work, complex extensions of MI would be needed [145].

There are several advantages with using a non-parametric algorithm such as the KNN for imputation of missing data. The majority of standard software packages rely on the assumption of normal distribution of multidimensional data, therefore imputation of repeated longitudinal data would result as [145]. In the two analyzed cohorts, alcohol data were collected at the daily level resulting in a high total volume of both data per participant and missing data. In the general population, alcohol consumption in pregnancy is highly skewed with the majority of the drinking concentrated in the first trimester with a subsequent drop in exposure as pregnancy progresses. In such scenarios, a nonparametric method such as KNN has the advantage of not making any

distributional assumption on the probability density function which could be employed to model the data.

In recent years, diverse KNN-like approaches has been increasingly used to impute missing data in research with high volume data such as genetics and metabolomics studies [146], [147]. In several recent reports KNN was shown to produce the smallest imputation error compared to methods such as mean and median imputation, Bayesian linear regression, K-Means, K-Medoids clustering algorithms [148], [149]. However, some studies reported that simpler methods such as mean or median replacement were as adequate as methods like KNN when imputation was followed by clustering of genetic data [150]. Additionally, some have reported slightly better performance of random forest over KNN to impute metabolomics data [151]. Another study noted improvement of performance when additional information such as SES and demographic data were included in the prediction model [152]. The validity and accuracy of imputation methods will likely vary with the data type, data structure, mechanism of missingness and amount of missing data. Therefore, future studies may need to evaluate the performance of different machine learning algorithms to impute alcohol consumption data in the context of other stressor to derive a more comprehensive profile of exposure in the perinatal period.

Cluster Analysis of Alcohol Consumption during Pregnancy in the Safe Passage Study

Introduction

Extensive literature has documented the adverse effects of prenatal alcohol exposure on birth outcomes and risk for neurodevelopmental disorders such as fetal alcohol syndrome (FAS) and fetal alcohol spectrum disorders (FASD) [153]. In a parallel fashion, several studies had focused on the investigation of magnitude and timing of exposure from a broader public health perspective. Nevertheless, since prior research has not elucidated a safe amount to drink during pregnancy, national drinking guidelines in most countries advise against consuming any alcohol during pregnancy. Despite continued recommendations against maternal alcohol consumption during pregnancy, this behavior still continues [154]. According to the most recent surveillance data from the Centers for Disease Control and Prevention, 1 in 10 pregnant women from 18–44 years of age reported consuming alcohol and approximately 1 in 33 reporting binge drinking (>4 drinks/occasion) within the last 30 days [154]. Due to the public health significance associated with alcohol consumption during pregnancy, accurate characterization of drinking is imperative to understand how these patterns relate to perinatal and early childhood outcomes. However, due to the heterogeneity of alcohol consumption during pregnancy, precise characterization of consumption behaviors in the perinatal environment is often difficult.

As previously introduced, the PASS study was specifically designed to investigate the role of prenatal drinking and smoking on peri- and postnatal outcomes in populations during pregnancy. In this paper, we provide details of a methodological approach used to derive prenatal drinking patterns from longitudinal alcohol consumption data to characterize the quantity, frequency, and timing of alcohol exposure collected in the PASS study [113], starting from the previously imputed alcohol consumption data. We utilized a finite mixture model to create groups of participants differing in their patterns of alcohol consumption. This type of approach increasingly has been used to model population heterogeneity for clustering and classification. The identified clusters will be utilized in analyses discussed in the following to investigate the effect of alcohol exposure in relation to perinatal and early childhood outcomes and early childhood outcomes. Additionally, the operation of grouping subjects based on maternal lifestyles is expected

to propose validation for a possible influence of such stressors within a broader and more comprehensive perspective.

Materials and methods

PASS study, missing data and K-NN-based Imputation

The study protocol was implemented for 11,083 pregnant women. Institutional Review Board approvals for the Safe Passage Study were obtained from sponsoring organizations at the participating clinical sites, the Data Coordinating Center, and the Physiologic Assessment Center [134]. Depending on the timing of enrollment, women participated in up to three clinical interviews during prenatal visits at 20–24 weeks, 28–32 weeks, and/or 34+ weeks. Alcohol exposure information was collected using a modified version of the Timeline Follow-Back (TLFB) method [25]. At each prenatal visit, daily alcohol consumption for the last reported drinking day and 30 days prior was collected using the modified TLFB.

As a result of the approach by which alcohol information was collected, a given participant could potentially have single or multiple segments of missing daily alcohol consumption data. A vast portion of the epidemiologic literature reports missing data as potentially introducing bias in many studies [19]. This is particularly problematic in the context of alcohol research where the mechanism of missingness is not at random, but rather a function of how the data were collected. As previously described, this created more missing days for drinkers compared with non-drinkers.

To address these issues, we imputed missing exposure data using a K-Nearest Neighbor (KNN) approach [135]. The implementation used K = 5 neighbors and it was validated considering segments of length equal to 55 days. An in-depth description of the methodology is reported in the previous section of this chapter.

Data pre-processing and feature extraction

Imputation of missing data was not performed for participants with missing alcohol data for ≥ 200 days throughout pregnancy or ≥ 90 days in the first trimester. As a result, 580 and 224 women respectively were excluded from further analysis. The remaining 10,279 subjects were included in the cluster analysis. Starting from the alcohol consumption data computed at the daily level, we extracted 6 features: the total exposure for each of the trimesters (T1, T2, and T3) expressed in terms of standard drinks and defined as: ALC_T1, ALC_T2, and ALC_T3; the total number of binge drinking events for each trimester defined as BINGE_T1, BINGE_T2, and

BINGE_T3. A binge episode was defined as the intake of ≥ 4 drinks on a given day. The rationale for deriving condensed alcohol exposure features aims at improving the interpretability of derived clusters and reduce the computational complexity. The reason behind the trimester-level granularity is of particular relevance from a public health point of view given that the majority of pregnancy guidelines are usually delivered at this level.

At this stage, women having a trimester-level alcohol consumption < 1, had that corresponding value set to zero. Then, non-drinker participants were defined as those subjects having all 6 features equal to zero (n = 5,168). For the remaining 5,915 participants, outliers for each feature were winsorized and assigned with a value equal to the mean + 4 times the corresponding standard deviations. Lastly, features were z-scored to obtain zero-mean and unitary-variance distributions.

Finite Mixture Model Clustering

Finite mixture models have been increasingly employed as a reliable and efficient methodology for data clustering, classification, and density estimation in several different fields [155]. A finite mixture model is a statistical approach that assumes the presence of unobserved groups, called latent classes, within a given population. The probability density function of each latent class can be fit with its own regression model, which can be either described by a linear or generalized linear response function. The rationale behind the application of finite mixture models relates to their ability to reveal unexplained heterogeneity that is either ignored or only partially revealed by multiple models for a priori groups. In the following, a detailed description of the method applied to the above-described population will be presented as well as its implementation in the context of this analysis.

Let $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ be a set of n independent identically distributed observations. The distribution of every observation is associated with a probability density function by one out of the G components of a finite mixture model, defined in Eq. 7 as:

$$f(\boldsymbol{x}_i; \boldsymbol{\Psi}) = \sum_{k=1}^{G} \pi_k f_k(\boldsymbol{x}_i; \boldsymbol{\theta}_k) \tag{7}$$

where $\boldsymbol{\Psi} = \{\pi_1, \ldots, \pi_G, \theta_1, \ldots, \theta_G\}$ are the parameters of the mixture model, $f_k(\boldsymbol{x}_i; \boldsymbol{\theta}_G)$ is the k-th component density for the observation $\boldsymbol{x}_i$ associated with the parameter vector $\boldsymbol{\theta}_k$, π-s are the mixing weights, and G is the numerosity of mixture components. At this point, let's suppose to set

G to an arbitrary value as no a priori information on the optimal number of clusters is available. As a result of this implementation, the remaining mixture model parameters can be estimated based on solely the considered data, provided as input to the machine learning algorithm. Such derivation of mixture parameters is usually addressed by an expectation maximization (EM) algorithm, given that direct maximization of log-likelihood is a non-straightforward procedure. In the model-based approach to clustering, by design each component $f_k(\boldsymbol{x}_i; \boldsymbol{\theta}_G)$ is associated with a group which can be intended as a cluster.

The classical assumption is that each component of a finite mixture model is constrained to having the same parametric distribution. On the contrary, we are not making that assumption. In the multivariate setting proposed in this approach, the volume, the shape, and the orientation of each probability density function can be constrained to be either equal or variable across groups. A crucial aspect in the context of mixture modelling is the choice of the optimal number of components to achieve adequate granularity to interpret the results while avoiding overfitting. To this end, we employed a modified Bayesian Information Criterion (BIC) as reported in Eq. 8:

$$BIC_{\mathcal{M},G} = 2\ell_{\mathcal{M},G}\left(\boldsymbol{x}|\widehat{\boldsymbol{\Psi}}\right) - v \cdot log(n) \tag{8}$$

where $\ell_{\mathcal{M},G}$ is the log-likehood at the EM $\widehat{\boldsymbol{\Psi}}$ for the chosen model $\mathcal{M}$ with G components, n is the sample size, and v is the number of estimated parameters. The pair $\{\mathcal{M}, G\}$ maximizing BIC is selected. At the same time, BIC has been reported to prioritize the selection of the number of mixture components needed to reasonably approximate the density, rather than the number of clusters as such. For this reason, we utilized a combination of BIC, the integrated complete-data likelihood (ICL) criterion, and likelihood ratio testing (LRT) [155] as figures of merit to select the optimal number of clusters. Once the adequate model $\mathcal{M}$ and the proper number of G components are chosen, an EM approach is carried out to provide initialization for the algorithm. In this work, an adapted version of R-package mclust has been utilized to perform finite mixture model clustering on the alcohol consumption features previously derived.

Results

As described in the Methods section, the choice of an adequate number of components (G) plays a crucial role when clustering data using a finite mixture model-based approach. Figure 17 displays the BIC values when progressively increasing the number of components, which de facto represents the optimal number of clusters into which participants will be subdivided.

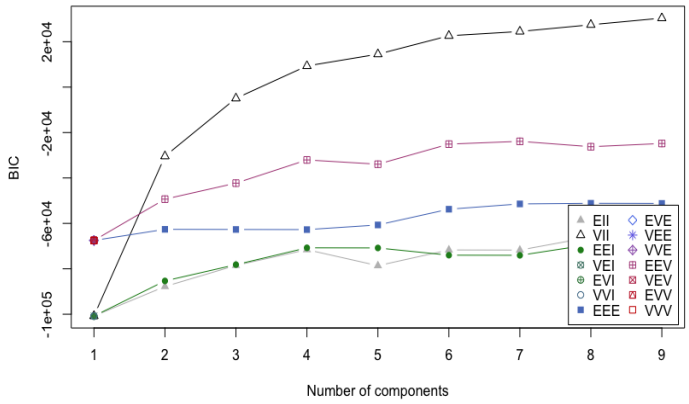


Figure 17. Performance evaluation of the different model employed to characterize the multidimensional space of features. Models differ by distribution (spherical, diagonal, ellipsoidal), volume (equal, variable), shape (equal, variable), and orientation (none, equal, variable). Maximization of BIC is reached with model VII employing a number of components G equal to 9. BIC estimates were found stable for values of G greater than 9 (data not shown).

Among all parameter combinations of the within-group covariance matrix needed to characterize the multidimensional space of the features, the model VII was chosen as fulfilling the requirement of BIC, ICL, and LRT maximization. In this analysis, the three figures of merits selected the same final model. From a geometric point of view, the within-group covariance matrixes $\sum_k$ were associated with spherical distributions as well as with an equal volume and orientation one to the others. To summarize, the pair $\{\mathcal{M} = VII, G = 9\}$ was found as optimal to model alcohol data consumption in the context of the analyzed dataset. Moreover, as shown in Figure 17, VII performed considerably better within a wide range of Gs compared to any of the other tested models. Similar results were found considering the different figures of merit described in the previous section. These convergent results allowed to strengthen the validity of the chosen model as well as the derived optimal number of clusters. Once the total number of clusters is defined, it is possible to derive the corresponding mean alcohol consumption and mean binge

episode count in T1, T2, and T3 for each group. Results are reported in Table 9 and displayed in Figure 18 (mean total number of drinks only).

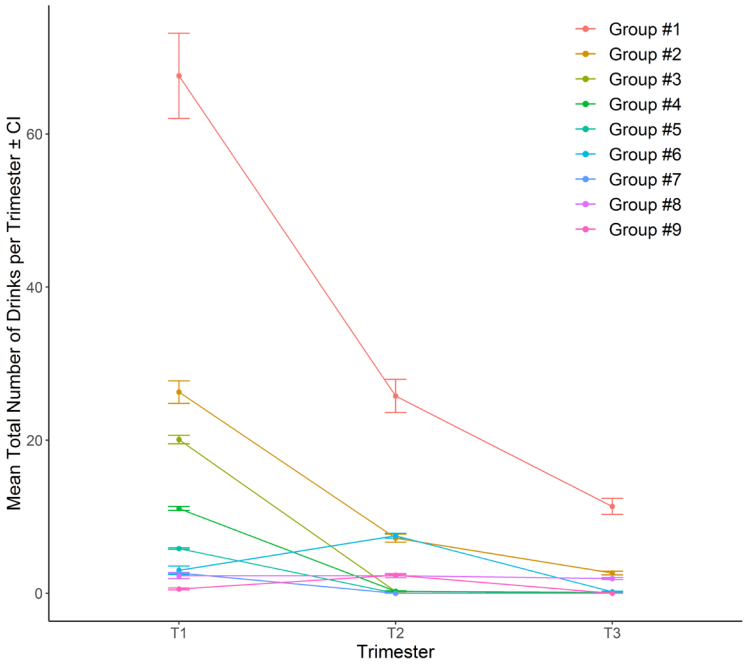


Figure 18. The average and confidence intervals (CI) number of drink per each trimester are displayed for each of the identified groups. Groups were numbered in descending order of mean ALC_T1.

The requirements of intra-cluster variance minimization and inter-cluster variance maximization imposed by finite mixture modelling were reflected in the resulting groups shown in Figure 19. To this end, it is possible to observe that the majority of participants within each group exhibit similar individual trajectory patterns of exposure. Such intra-cluster pattern matching was found in terms of both magnitude and timing of alcohol consumption. Regarding the requirement of maximization of inter-cluster variance, it is important to note the minimal overlap between each cluster as also shown in Figure 18.

Table 9. Mean ± Confidence Intervals of the alcohol consumption features within each clustered group. %drinkers indicates the group numerosity with respect to the total number of drinkers (n=5,915).

Group #	%drinkers	ALC_T1	ALC_T2	ALC_T3	BINGE_T1	BINGE_T2	BINGE_T3
1	9.50	67.60 ± 5.56	25.77 ± 2.17	11.34 ± 1.05	6.11 ± 0.48	2.67 ± 0.23	1.11 ± 0.11
2	17.48	26.28 ± 1.47	7.19 ± 0.53	2.64 ± 0.23	2.32 ± 0.15	0.64 ± 0.06	0.21 ± 0.03
3	10.55	20.08 ± 0.56	0.23 ± 0.07	0.10 ± 0.04	2.14 ± 0.04	0.00 ± 0.00	0.00 ± 0.00
4	15.74	11.07 ± 0.26	0.26 ± 0.05	0.03 ± 0.01	0.64 ± 0.03	0.00 ± 0.00	0.00 ± 0.00
5	10.85	5.84 ± 0.10	0.00 ± 0.00	0.00 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
6	4.50	2.99 ± 0.55	7.51 ± 0.32	0.19 ± 0.07	0.19 ± 0.05	1.00 ± 0.00	0.00 ± 0.00
7	19.53	2.64 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
8	6.49	2.27 ± 0.37	2.30 ± 0.27	1.91 ± 0.13	0.14 ± 0.03	0.00 ± 0.00	0.00 ± 0.00
9	5.36	0.55 ± 0.13	2.35 ± 0.10	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

When addressing the similarities among clusters it was possible to identify three main levels of cumulative exposure when dealing with magnitude only. Participants in groups #1, #2, and #3 share a high cumulative level of exposure of around 20 drinks in either T1 or T2, participants in groups #4, #5, and #6 show an average of 9 drinks throughout pregnancy, whereas participants in groups #7, #8, and #9 were exposed to a significantly lower amount of alcohol consumption. On the other hand, when focusing on timing of exposure only, participants in groups #1, #2, and #8 displayed similar alcohol consumption patterns characterized by a high alcohol consumption in either T1 or T2 only. Similar logic can be applied to individuals in groups #3, #4, #5 and #7 for which alcohol exposure is found in T1 only and subjects in groups #6, #9 reporting exposure in T2 only.

When interpreting the results in terms of both total alcohol consumption and binge drinking events, it is possible to derive more insights regarding group membership. Each cluster has a very homogenous binge drink profile characterized in terms of both magnitude and timing as summarized in Table 9. Furthermore, the reason for the partial overlap between groups can be explained by looking at the information provided by binge drinking patterns. A clear example of

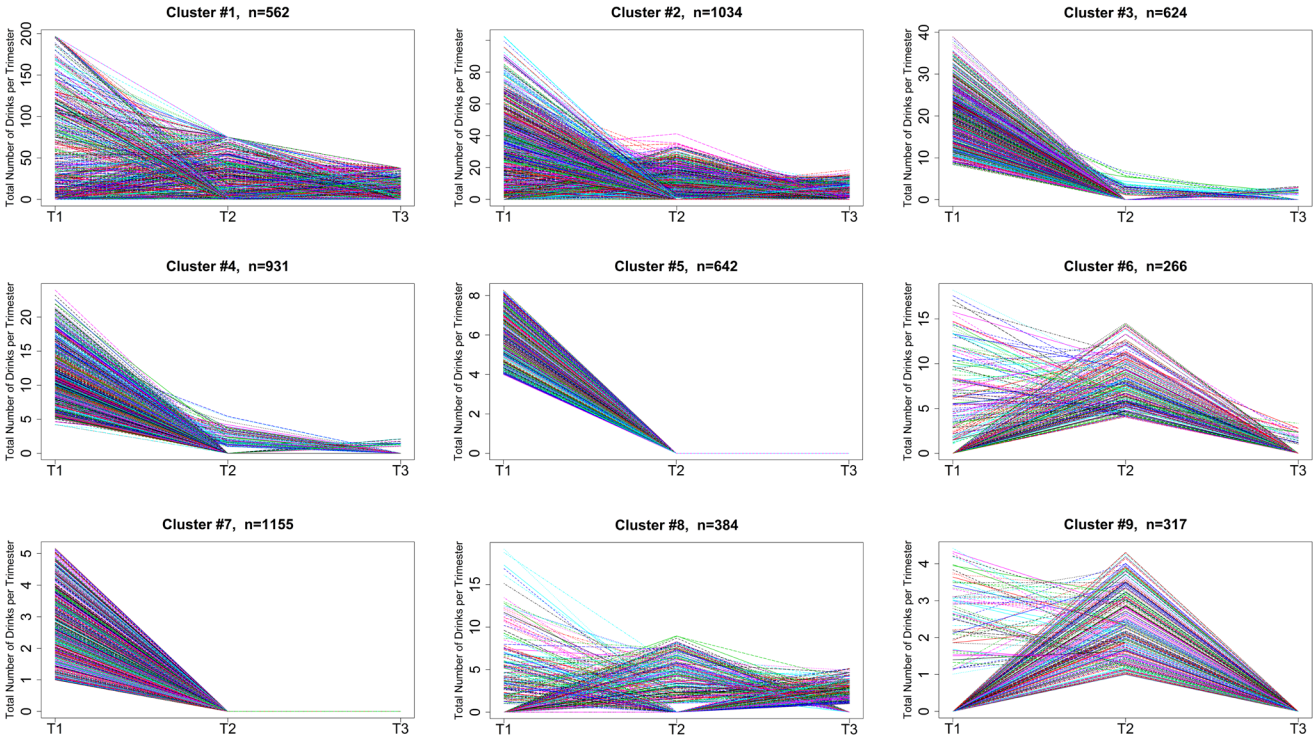


Figure 19. Individual trajectories of participants grouped in the derived clusters. Groups are numbered based on the mean total number of drinks in T1, starting with the most heavily exposed group.

this can be seen in the comparison of group 5 and group 7. A portion of the participants in the two groups share a similar ALC_T1 consumption in the 4-5 drinks category. The feature that differentiates the two groups is BINGE_T1, thus leading assignment to different group membership. The same applies to group #3 versus #4, or group #6 versus #8.

Discussion

The effects of maternal alcohol consumption during pregnancy, as well as the consequent effects on prenatal and postnatal outcomes, are still debated. A clear example of such are the contradicting results from studies investigating the correlation between alcohol exposure and birth weight [156]. The most common barrier to convergent findings is likely due to the different metrics employed to quantify alcohol intake which result in the inability to compare findings across studies. Moreover, a very limited number of studies provide a characterization of binge drinking patterns, despite documented adverse effects of heavy alcohol consumption on neurodevelopmental outcomes [157]. Lastly, the timing aspect of alcohol exposure is often neglected when comparing non-exposed women to those who drink throughout gestation [156].

In this work we aimed to identify clusters of alcohol consumption during pregnancy by means of a family of machine learning algorithm, namely finite mixture model. This family of methods are being widely used to derive computationally convenient representations of complex distributions representative of data on random phenomena. Given their inner flexibility, mixture models are being increasingly exploited as an advantageous, semiparametric way in which to model unknown distributional shapes. The main advantage of the methodology discussed in this work relies on the fact that it does not require any assumptions regarding data distribution to be clustered, thus making it suitable to deal with skewed and imbalanced distributions as those of alcohol consumption. Several approaches for dealing with skewness have been introduced, including data transformation such as the Box-Cox transformation and the Manly transformation [27].

Arguably, the most controversial methodological problem associated with mixture distributions is that of identifying the optimal number of components to summarize the distribution of the underlying set of data. To overcome the shortcomings of the EM-based approaches, Bayesian approaches have been suggested and consistently integrated in the proposed approach.

At the time, it has been reported that BIC tends to favor models with increasing number of components hence, it tends to overestimate the number of clusters [27].

Coming to the discussion of the results of mixture model analysis on PASS data, we report that the extracted clusters reflect both magnitude and timing of exposure simultaneously, thus providing a more comprehensive picture of the multiple patterns of alcohol consumption in large heterogeneous populations. This combined and concise representation of the diverse pattern of exposure has the advantage of reducing the complexity of exposure variables, thus facilitate their use in diverse contexts. The utilization of this novel characterization of alcohol exposure in the context of studies such as PASS affords the opportunity to provide a better understanding of the effect of alcohol consumption during pregnancy and as well as the possibility for obtaining deeper insight regarding the associations on between drinking consumption, perinatal physiology, and ultimately adverse outcomes.

The K Nearest Neighbor Algorithm and Cluster Analysis for Imputation and Clustering of Missing Longitudinal Prenatal Tobacco Data

Alongside the previously described in-utero alcohol exposure we investigated an often-coexisting one, namely cigarette smoking. A large portion of the literature investigates tobacco exposure in the perinatal period focusing on the acute response to maternal cigarette smoking and the consequent effects in terms of fetal responsiveness [158]–[160]. Prolonged cigarette smoking has been reported to reduce baseline average fetal heart rate, with the characteristic of the mean beat interval being negatively correlated with daily cigarette consumption [38]. Moreover, the acute response to cigarette smoking on fetal heart rate encompassed a transient decrease of the short-term fetal heart rate variability. The degree of reduction of both these parameters was positively correlated with the values of maternal nicotine in plasma before smoking [159]. Similar effects for the acute response to cigarette smoking are described in [160]. Specifically, fetal diastolic blood flow velocity increased significantly after smoking, and the pulsatility index fell significantly during the first 5 min after smoking, possibly as the consequence of fetal tachycardia. These results indicated that fetal central circulation was significantly modified following maternal smoking even in presence of an exposure of a single cigarette. Further analysis on a similar cohort highlighted a transient mild elevation of baseline fetal heart rate following tobacco consumption [160]. Nevertheless, no significant acute changes in mean fetal heart rate were found, a result in contrast with the previously reported effect of chronic tobacco exposure [159].

Coming to the discussion of the studies investigating on the non-acute effects, fetal heart rate variability was found lower in fetuses whose mothers smoked cigarettes during pregnancy [15], results are in accordance with findings reported in [38]. Additionally, authors reported a dose response relationship between amount of maternal cigarette smoked during first trimester and measures of fetal heart rate variability as well as the associated powers at different frequencies [15]. Other reports found that in-utero tobacco-exposed infants were more excitable and hypertonic, required more handling and showed more stress/abstinence signs, specifically in the central nervous system (CNS), gastrointestinal, and visual areas [161], [162]. Lastly, in a large community sample, exposure to maternal smoking was associated with increased irritability and

hypertonicity in neonates but it was not associated with neonatal response to respiratory challenge [163].

The main limitation shared among the cited studies is a paucity of data-driven approaches to characterize maternal smoking behavior in pregnancy, as the tendency is to employ arbitrary thresholds to define the different exposure groups. A very simplistic but widely used categorization of tobacco exposure consist in subdividing participants between those who smoked at least one cigarette during pregnancy and those who did not [15], [164]. An alternative strategy is to stratify participants as a function of cumulative exposure during pregnancy, thus not to account for the timing of exposure [15], [159], [161], [163], [165]. To our knowledge, the analysis in [166] is the only attempt reported in the literature to describe tobacco exposure as a function of both timing and magnitude. At the same time, quantity was categorized as either light smoking (up to ten cigarettes per smoking day), moderate smoking (11–19 cigarettes per smoking day), or heavy smoking (≥20 cigarettes per smoking day), thus not providing an indication on the trimester level. This latter limitation is a consequence of cigarette smoking data collection which often consists of a single data point in pregnancy, hence likely to result in a significant underestimation of tobacco exposure. On the contrary, the pipeline proposed in this work aims to utilize the richness of PASS dataset to derive a more comprehensive characterization of cigarette smoking exposure and ultimately cluster participants as a function of both magnitude and timing of tobacco consumption.

In the context of PASS data, prenatal smoking exposure was collected by querying participants on the frequency of their tobacco cigarette consumption by means of a graduated response options (i.e., none, monthly or less, 2-4 days per month, 2-3 days per week, 4-6 days per week, and 7 days per week) and the number of cigarettes smoked on a typical day. Thus, in contrast with what previously described for alcohol, the tobacco exposure data were collected on the weekly level and not on the daily level, hence imputed on a les granular scale than alcohol exposure data. The time interval covered by the implemented imputation ranged from week -2 (2 weeks prior to LMP) to the maximum possible pregnancy length set to 42 weeks. We employed the same the distance adjustment described for alcohol imputation. Outliers were Winsorized at 4 SD. We proposed to introduce exclusion criteria to avoid unreliable alcohol consumption interpolations. Specifically, we excluded participants with less than 5 weeks of smoking exposure data throughout pregnancy (n = 75). Similarly to the KNN approach implemented in the context of drinking data, missing smoking exposure data were imputed considering a number of neighbors K = 10.

Moreover, to evaluate the model performance we determined that a segment of length equal to two weeks retained the majority of the participants in the reference pool without significantly diminishing the ability to identify the corresponding smoking patterns. Figure 20 summarizes the algorithm performance. A minimal portion of the imputed segments were imputed with a tobacco consumption value different from the reference one. Additionally, the average difference between

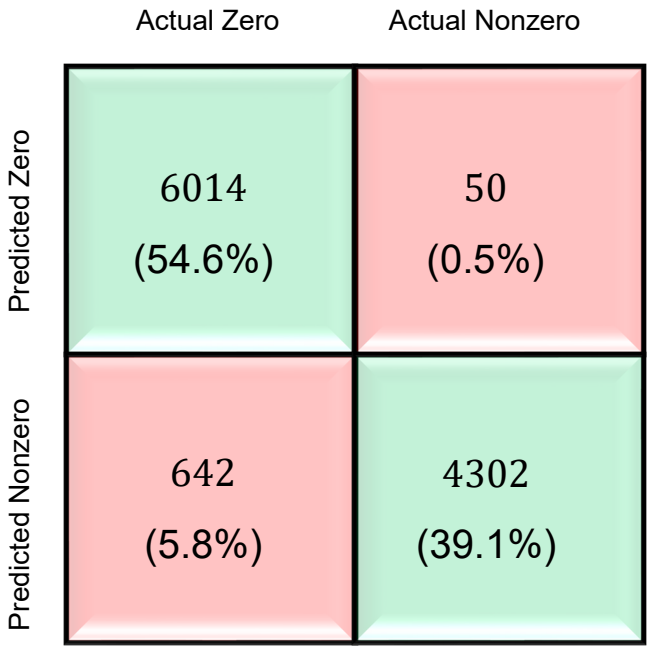


Figure 20. Summary of the performance for the proposed KNN implementation. Confusion matrix obtained considering to K = 10 and validating the model with segments of length equal to two weeks.

the actual and predicted number of cigarettes smoked weekly was ≤ 1 for the majority of the participants whose consumption could not be predicted exactly.

If we were to compare the performance of the proposed KNN implementation with respect to the results previously described for alcohol exposure data, we can notice an enhanced prediction ability in the former case. This is likely to be related to the less grained time scale by which smoking exposure data were collected. At the same time, a mismatch between the actual and predicted weekly smoking exposure has an intrinsically more pronounced influence, hence reflecting a greater degree of underestimation or overestimation of consumption.

At this stage, the remaining 11,008 subjects were included in the subsequent cluster analysis. Starting from the smoking consumption data computed at the weekly level, we extracted 3 features: the total exposure for each of the trimesters (T1, T2, and T3) expressed in terms of cigarettes and defined as: SMK_T1, SMK_T2, and SMK_T3. Participants with a trimester-level

tobacco consumption < 1, had that corresponding value set to zero. Then, non-smoker participants were defined as those subjects having all 3 features equal to zero (n = 5,155). For the remaining 5,853 participants, outliers for each feature were Winsorized and assigned with a value equal to the mean + 4 times the corresponding standard deviations. Lastly, features were z-scored to obtain zero-mean and unitary-variance distributions. The finite mixture model approach presented in the context of alcohol exposure clustering was also applied to the imputed smoking consumption data. Specifically, the optimal number of clusters was found equal to 4. This evidence reflects the more stable behavior of smoking consumption as previously described. A summary of the derived clusters is reported in Table 10 and in Figure 21, group were numbered in descending order of mean SMK_T1.

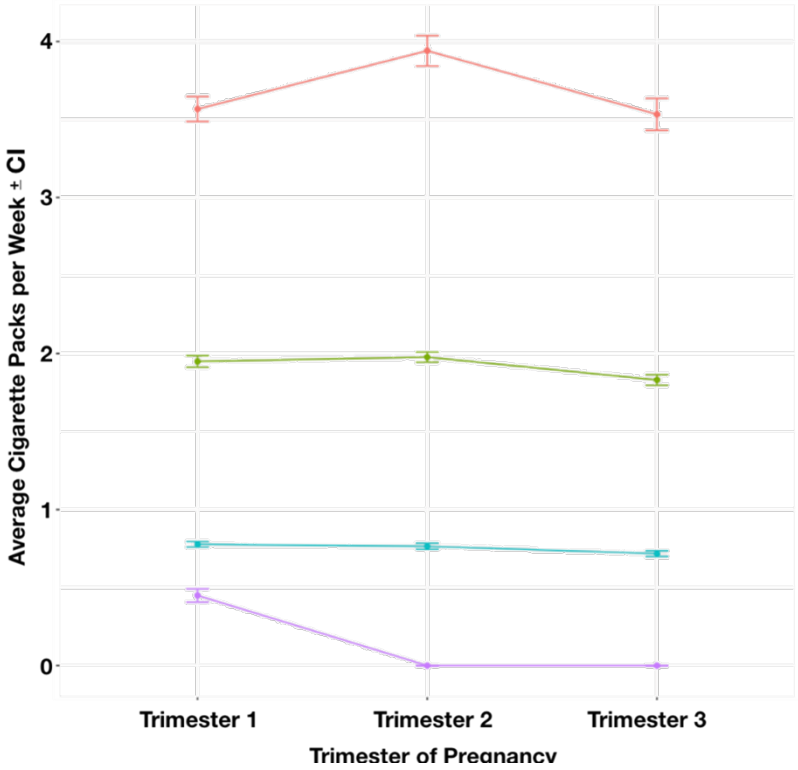


Figure 21. The average and confidence intervals (CI) number of cigarettes per week in each trimester are displayed for the derived clusters.

Table 10. Mean ± Confidence Intervals of the smoking consumption features within each clustered group. %smokers indicates the group numerosity with respect to the total number of smokers (n=5,853).

Group #	%drinkers	SMK_T1	SMK_T2	SMK_T3
1	10.50	71.84 ± 1.74	80.06 ± 2.28	72.67 ± 2.46
2	24.10	39.19 ± 0.78	39.30 ± 0.67	36.40 ± 0.70
3	55.95	14.52 ± 0.58	12.44 ± 0.61	12.12 ± 0.57
4	9.45	9.02 ± 0.84	0.07 ± 0.02	0.11 ± 0.02

In contrast with the results previously reported for alcohol, it is possible to appreciate a reduced variance of the exposure clusters. Additionally, consumption profiles are mainly differentiated by magnitude rather than timing of the exposure. Participants in groups #1, #2, and #3 share the same trajectory across trimester. The smoking consumption in these clusters remains stable throughout pregnancy even though at different magnitudes. On the other hand, participants in group #4 did not continue their tobacco consumption beyond the first trimester, a behavior probably to be associated with pregnancy recognition. Consistently with the results previously reported, we derived clusters characterized by a pronounced intercluster distance and a minimal intracluster variance.

Chapter 3

Abnormal fetal growth and alcohol and tobacco exposure are recognized as a perinatal modifiable risk factor responsible for a plethora of adverse perinatal outcome such as reduction in birthweight, increase probability of preterm and late preterm birth, morbidities and mortality [119], [120]. This chapter focuses on the contribution of advanced signal processing methodologies for the investigation of the relationship between prematurity and exposure, in terms of the investigation of possible imbalances of the cardiorespiratory controlling system.

Premature infants are clinically classified as *extremely preterm* (≤ 25 weeks), *very preterm* (≤ 32 weeks), *moderately preterm* (≤ 34 weeks), and *late preterm* (≤ 36 weeks). Prematurity has been consistently reported as a risk factor for a variety of neonatal and early infancy pathological conditions as well as adverse neurodevelopmental outcome [5], [167]–[170]. Among the illustrated degree of prematurity, late preterm infants are usually considered of subclinical relevance but they are at higher risk of morbidity and mortality than term infants [6]. Infants born between at $34^{0/7}$ through $36^{6/7}$ weeks' gestation, are characterized by less physiologically and metabolically mature physiological control systems if compared to term infants. Historically, this was the reason for which a change in terminology from *near term* to *late preterm* was originally proposed and nowadays largely accepted. Late prematurity predisposes infants to a higher risk of morbidity and mortality [6]. Specifically, the mortality rate for late preterm infants is 3-fold higher than that for term infants [171]. Despite the growing clinical awareness of late prematurity burden, late preterm newborns are the fastest growing subset of neonates, accounting for approximately 74% of all preterm births and about 8% of total births in the U.S. [171]. Late preterm newborns are frequently admitted to neonatal intensive care units (NICU) for observation during the first hours of life, even if they do not meet acute high illness criteria. Yet, infants born at 35–38 weeks in most cases do not receive any special care unless they show signs of distress in the immediate postpartum period [172]. Investigating the physiological mechanisms underlying the increased vulnerability of this population and identify early markers of risk could inform clinical care and shape public health policies for all newborn infants and ultimately create an explainable link between exposure and prematurity. Previous investigation focused on univariate methodologies for the investigation of the differences in autonomic regulation of preterm infants [172], [173]. At the same time, it is well

established that physiological systems under neural regulation exhibit high degree of complexity with non-stationary, intermittent, scale-invariant and nonlinear behaviors [31]. Additionally, in recent years it has become evident that the properties of such systems are likely to be change under different physiological states as well as pathologic conditions to reflect modification in the underlying control mechanisms [174], [175]. The quantification of diverse physiological interactions is a challenge as different systems may exhibit multiple simultaneous interactions and/or vary the strength of the couplings as a function of time. To achieve an improved characterization of the diverse human body networks, and thus study the dynamical evolution of such systems in relation to different physiological states, it is necessary to develop methodologies able to quantify the nonlinearity and transient behaviors.

At birth and in the early infancy period, the mutual influence of cardiovascular and respiratory rhythms in healthy newborns has been documented but its full characterization is still pending. The activity of many physiological subsystems presents a well-expressed rhythmic character, and often an interdependency between physiological rhythms emerges early in development. Given this premise, we propose to investigate the cardiorespiratory system interrelationships in terms of possible presence of causal or directional interplays. Specifically, we tested different methodological application to quantifies phase coupling and its directionality in a population of newborn infants born between 35 and 40 weeks of gestational age (GA). The aim is to assess whether GA at birth significantly influences the development of phase synchronization and the directionality of the coupling between the cardiovascular and respiratory system activity. Several studies indicating irregular cardiorespiratory coupling as a leading cause of several pathologies underscore the need to investigate this phenomenon in this at-risk population.

An additional approach which will be discussed in the following is an extension of the traditional Transfer Entropy measure (TE), which employs multiple lagged versions of the time series of the intervals between two successive R waves of the QRS signal on the electrocardiogram (RR series) and respiration time series (RESP). The method aims to quantify the instantaneous and delayed effects between the two processes within a fine-grained time scale. Results indicate a progressive increase in information transfer as a function of gestational age, as well as significant differences in terms of instantaneous and delayed interactions between the cardiac and the respiratory system when comparing the two TE directionalities (RR→RESP vs RESP→RR). The proposed investigation addresses the role of the different autonomic nervous system (ANS)

branches involved in the cardiorespiratory system, since the sympathetic and parasympathetic branches operate at different time scales. Our findings allowed to infer that the two TE directionalities are uniquely and differently modulated by both branches of the ANS. TE adds an original quantitative tool to understanding cardiorespiratory imbalance in early infancy. Taken together, the proposed approaches are expect to contribute to the understanding of higher risk for the documented negative outcomes in the late preterm population. Moreover, these parameters could provide a tool for the development of early markers of cardiorespiratory dysregulation in infants.

Then, we will investigate the association between the derived estimates of alcohol and tobacco consumption and physiological signals at birth. This will provide insight on the quantitative differences between in utero exposed and unexposed newborns. The derived clusters will be contextualized with physiological signals as well as a more comprehensive ensemble of prenatal exposures. This noninvasive approach to assessment integrity of the structures devoted to autonomic control allows to gain insights on the possible consequences of prenatal exposure and ultimately to unveil the potential mechanisms of underlying vulnerability.

Lastly, we will present a data driven approach for the prediction of neurobehavioral outcome. Specifically, we utilized heterogeneous data collected during the fetal and neonatal period to estimate cognitive performances in toddlers of 3 years of age. The implemented machine learning framework attains the potential of a better understanding of the relationship between diverse physiological data in the context of exposure to substances, maternal health, environmental factors, and various external stressors.

Chapter 3 investigates advanced signal processing methodologies to quantify the evolution of cardiorespiratory interaction as a function of gestational age [176]–[178]. Then, neonatal physiology is contextualized with prenatal exposure to unveil the effect of such stressors on ANS as well as CNS [179]–[181]. Lastly, we provide evidence for the successful application of a machine learning framework able to predict neurodevelopmental outcomes combining parameters extracted from physiologic and clinical data.

Keywords – **Prematurity, Cardiorespiratory Regulation, Transfer Entropy, Multivariate Modelling, Neurodevelopmental Outcomes, Recursive Partitioning, Feature Elimination**

Characterization of Cardiorespiratory Phase Synchronization and Directionality in Late Premature and Full Term Infants

Introduction

In the last few decades, knowledge about relationships between respiratory and cardiovascular systems has gradually grown with intermittent reports of investigations on associations of biological rhythms and central nervous system (CNS) activity [182]–[184]. Biological rhythms associated with peripheral autonomic system activity, on the other hand, have been thoroughly investigated [185]–[187]. Neural systems not only control cardiovascular signals, such as arterial blood pressure (ABP) and heart rate (HR). Respiratory signals, such as inter breath interval (IBI) series, are modulated by neural control as well.

The human cardiorespiratory system is highly rhythmic with oscillations in the HR tied to those in blood pressure and blood flow, but also influenced by inspiratory and expiratory rhythmic phases in respiration. Rhythmic functioning is a peculiar and ubiquitous characteristic of many physiological systems [188], [189]. Understanding the nature of the interaction between respiration and HR has informed the etiology of many adult pathologies, but many aspects of these complex interactions are still unexplained. Investigating the emergence of these interactions in infants represents an even more challenging field of study. At birth, the activation of the cardiorespiratory system is initiated but it is not fully developed either anatomically or functionally.

Several modes of interaction exist between these two subsystems: the most well-known is the HR amplitude modulation driven by respiration, referred to as Respiratory Sinus Arrhythmia (RSA). It consists of a HR increase during inspiration and decrease during expiration [190], [191]. The modulation can be categorized as a linear interaction, but nonlinear relationships have also been identified, such as phase synchronization between heartbeats and breathing cycles [192]–[195]. Many studies describe the cardiorespiratory coupling as an interaction between two subsystems which can be modeled as two weakly coupled chaotic oscillators [182], [192], [196]. The basic idea is that given two weakly coupled systems, the amplitude of their oscillations may remain uncorrelated whereas their phases do mutually perturb. Within this assumption, it becomes possible to investigate cardiorespiratory synchronization by means of a phase analysis of RR series

and respiratory signal rather than applying a classical amplitude analysis. The behavior of the cardiorespiratory system can be seen as synergetic, meaning a multi-stable system switching between several phase attractive maps, with a preference for a specific set of phase relations, which can be seen as attracting frequency ratios [194]. These different modes of interactions are not exclusive, rather they may simultaneously coexist, representing different aspects of neural regulation and acting on different time scales [197], [198].

RSA generation is associated mainly with direct brainstem modulation of the cardiac vagal preganglionic neurons and by inhibition of cardiac vagal efferent activity by lung inflation. RSA is also thought to improve pulmonary gas exchange [190]. On the other hand, the physiological mechanisms behind the phase synchronization are still not fully understood. Recent results have led to hypothesize a link with the central nervous coupling factors; these physiological circuits might coordinate cardiovascular and respiratory rhythms in the brainstem through the control of phase synchronization between nerve discharges thus improving energy efficiency [199], [200]. RSA is observed in full term infants indicating the presence of cardiorespiratory coupling even at this early life stage [201] . This coupling increases in strength and consistency with GA at birth, reflecting the transition from sympathetic to parasympathetic dominance during the postnatal period [202], [203] with an increasing influence of respiration modulating HR. Overall, immaturity in linear cardiorespiratory coupling is observed in preterm infants, mostly in the form of lower HRV in the high frequency range which can be seen as an indirect measure of RSA [204]. This developmental difference between term and preterm infants highlights that the latter are at higher risk for impaired responses to cardiorespiratory stress.

Over the years, many methods have been proposed to quantify the interaction between subsystems, ranging from traditional to more complex signal-processing techniques, e.g. cross-spectra, mutual information, phase-locking analysis, time delay stability [205], [206]. Nonetheless, such linear and nonlinear approaches provide measures of symmetric interaction and are limited in their ability to evaluate possible different causal mechanisms that may underlie cardiorespiratory behavior. Rosenblum et al. proposed a method which exploits the notion of phase synchronization of irregular oscillators: the attempt was to reveal whether the interaction is bi- or unidirectional and to quantify the degree of asymmetry in the systems' coupling [207]. The method assumes that the system under study can be modeled by coupled self-sustained systems. This approach starts from the fact that a weak coupling firstly affects the phases of the oscillators, not

their amplitudes. Thus, to quantify the coupling strength of interaction, one has to analyze relation between the phases of the systems only. These novel methodologies stem from the experience of a new field of study, called Network Physiology (NP). NP offers a frame of study to address the question of how different physiological systems interact and behave together. Moreover, it stresses the advantages of this holistic approach over the reductionist methods, which analyze every system separately. Recent publications have highlighted the new insights afforded by this novel approach [31].

In this study, we are interested in quantifying this type of cardiorespiratory interaction in a population of late-preterm infants, born between $35^{0/7}$ and $36^{6/7}$ weeks of gestation. This population is known to be at higher risk of morbidity and mortality than term infants. Even more important, in the normal clinical path, they are treated the same as term infants both in Europe and the US, unless they show signs of distress [6], [208]. Investigation into the cardiorespiratory coupling in this population is warranted given many studies documenting an impaired cardiorespiratory coupling in many pathologies, such as apnea of prematurity, and potentially involved in the pathophysiology of Sudden Infant Death Syndrome (SIDS) [209], [210]. Therefore, a set of methods focused on cardiorespiratory phase coupling may provide quantitative indices of underlying mechanisms and early identification of possible risk states.

Materials and methods

Population

The Institutional Review Boards of the New York State Psychiatric Institute and of the Columbia University Irving Medical Center (CUIMC) approved all consent and data acquisition procedures used to collect the data. Mothers signed informed consent forms prior to enrollment in the study. The newborn dataset includes 273 infants born at the Morgan Stanley Children's Hospital of New York between $35^{0/7}$ and $40^{6/7}$ weeks of GA. None of the infants were admitted to the Neonatal Intensive Care Unit and review of the maternal medical chart revealed no evidence of major illness, genetic disorders, or past/present medicated/non-medicated psychiatric complaints. All infants had a minimum Apgar score of 8 after 5 minutes of life. The participants who met inclusion criteria were recruited and tested 12-84 hours after birth. Infants were grouped based on GA: Late Preterm (LPT, $35^{0/7}$ to $36^{6/7}$ weeks GA at birth, n=66), Early Term (ET, $37^{0/1}$ to $38^{6/7}$ weeks GA at birth, n= 93) and Full Term infants (FT, $39^{0/1}$ to $40^{6/7}$ weeks GA at birth, n=114).

Signals acquisition and preprocessing

The electrocardiograph (ECG) was recorded with three leads, placed on the infant's chest (left abdomen, left and right scapula) and the signal was amplified and collected using the DATAQ Instruments ECG system (Medelex, NY, NY). A respiratory inductance belt (Ambulatory Monitoring Inc., Ardsley, NY) was placed around the infant's abdomen to measure the respiration signal. ECG and respiration signals were acquired at 500 Hz and 200 Hz respectively. Acquisition started within ~30 minutes after feeding and lasted 10 minutes, with infants sleeping in supine position. Sleep states were classified into Active Sleep (AS), Quiet Sleep (QS), Indeterminate (I) and Awake (W). The minimum length for a segment to be classified either as AS or QS was 120 seconds, I and W segments were discarded from the analysis. The sleep state coding assessment was based on respiratory variability [211] and confirmed by behavioral codes entered throughout the study to determine when infants were awake, crying, or fussy.

The R peaks were detected on the ECG with a proprietary software (Gmark, Ledano Solutions) based on the Pan-Tompkins algorithm [212] and then checked with visual inspection. The respiration signal was filtered with a bandpass filter (0.05 - 3.5 Hz), peaks of inspiration were detected with automated marking software and each record was modified for incorrect marks manually. Thresholds of acceptance for RR interval were set as 0.3-0.667 seconds, with an absolute variation between consecutive RR intervals of 10%, while for respiration thresholds were 0.5-2.5 seconds and an absolute change of 40%. Segments of three minutes length in a continuous sleep state were analyzed providing 279 QS segments and 419 AS segments.

Phase estimation

Instantaneous phase of the ECG (ϕ_{RR}) is defined as linearly increasing between an R peak and the successive one and it is computed as in Eq.9:

$$\phi_{RR}(t) = 2\pi k + 2\pi \frac{t - t_k}{t_{k+1} - t_k} \tag{9}$$

where t_k are the times of appearance of the k-th R peak. The respiratory signal was detrended and filtered with a Savitzky-Golay filter and then its instantaneous phase (ϕ_{RESP}) was computed by means of functions provided by the Data Analysis with Model of Coupled Oscillators (DAMOCO) Toolbox [207], [213], [214]. Respiration signal protophase is computed via the

Hilbert transform and the phase is derived with appropriate transformation of protophase [215]. Instantaneous phases of ECG and respiration were both resampled at 200 Hz to obtain two synchronous phase series.

Phase synchronization

Synchronization can be defined as the adjustment of the rhythms of self-sustained oscillators due to their interaction. Given the phases of the two oscillators ϕ_1 and ϕ_2, a generalized *n:m* phase locking ratio fulfills the condition expressed in Eq. 10:

$$|n\phi_1(t) - m\phi_2(t) - \delta| < const \tag{10}$$

where n and m are integers and δ is an average phase shift. Thus, in the synchronized case, phase differences should present small fluctuations around a constant, while in the unsynchronized one, differences should randomly vary for every instant of time. ϕ_{RR} and ϕ_{RESP} were defined as the first and second oscillators respectively, *n* indicates the number of heartbeats with respect to *m* respiratory cycles. Different n:m synchronization periods can be highlighted with a graphic tool called cardio-respiratory synchrogram, shown in Figure 22 panel c [192], [216]. Defining R_i $(i = 1, \dots, n_R)$ as the series of R peaks and I_j $(j = 1, \dots, n_I)$ as the series of inspiratory onsets, it is possible to compute the relative distance $\left(\Psi(t_i)\right)$ between the current respiratory onset and the successive R peaks as:

$$\Psi(t_i) = \left(\frac{R_i - I_j}{I_{j+1} - I_j} + j\right) \bmod b \tag{11}$$

Panel c of Figure 22 shows relative distances φ_i (plotted as dots) with respect to single breathing cycles ($b = 1$).

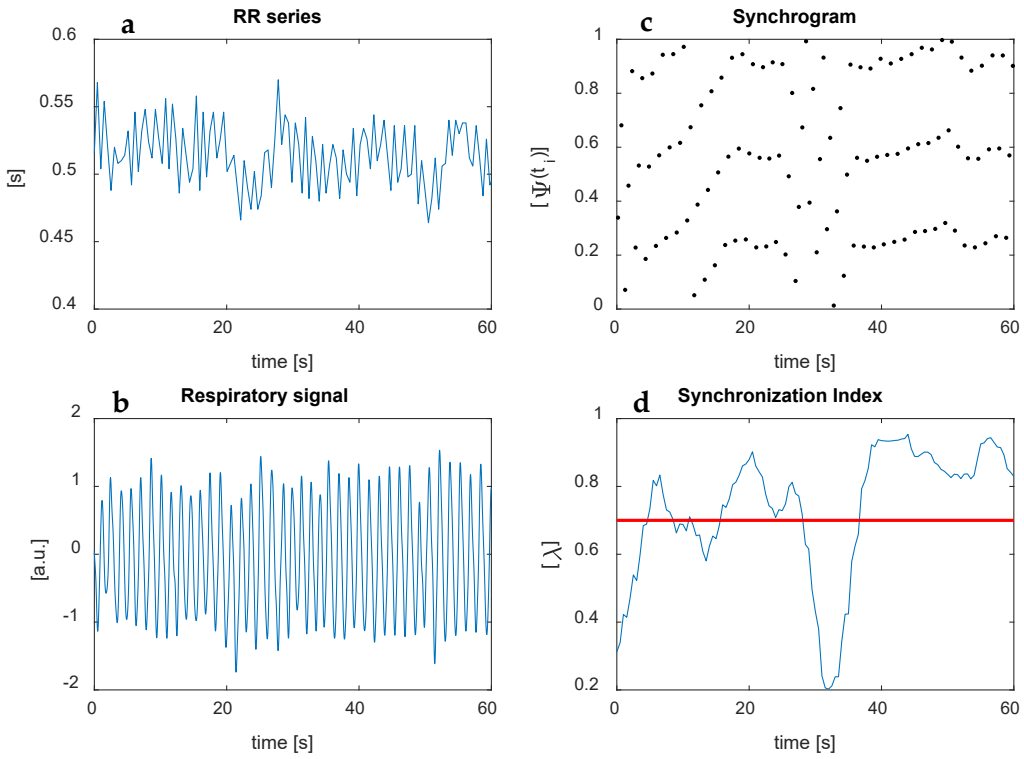


Figure 22. a): The RR series (extracted from the ECG signal) in a 60 seconds window. b): Respiratory signal for the same window described in (a). The signa l shown in (a) and (b) are obtained from a full term subject in quiet sleep. c): The synchrogram extracted by the joined analysis of RR series and Respiratory signal. The dot graph represents the relative distance between R peaks and respiratory onsets. d): The horizontal line represents the threshold for the computed synchronization index (λ). Synchronization index values above the threshold indicates epoch of synchronization between cardiac and respiratory systems.

Of the several approaches employed to quantify the level of synchronization, the method of the synchronization index λ was chosen, for its proven reliability [217]. To estimate the degree of synchronization by means of the λ index, ϕ_1 and ϕ_2 will be considered as cyclic (mod 2π). The process starts fixing a value for the phase of the first oscillator, Θ, to observe the phase of the second oscillator and compute the distribution η_i defined as $\eta_i = \phi_2|_{\phi_1=\theta}$ at each time t_i when $\phi_1 = \Theta$.

In the case of 1:1 phase locking, the values of ϕ_2 in the time points when $\phi_1 = \Theta$ are going to be scattered around a constant, due to weak noise. The distribution η_i can be characterized computing the intensity of its first Fourier mode. To strengthen the statistic, this process can be repeated with a binning-like procedure for different values of Θ.

In case of N points binning, the index Λ_l computed for the l-th bin can be expressed as:

$$\Lambda_l^2 = M_l^{-2}\left(\sum_{i=1}^{M_l}\cos\eta_i\right)^2 + M_l^{-2}\left(\sum_{i=1}^{M_l}\sin\eta_i\right)^2 \quad (12)$$

where M_l is the number of points in the l-th bin. In this work, the cyclic phase of the first oscillator was divided in 10 equally spaced bins.

The average of the obtained Λ_l over all N bins provides a measure related to circular variance that is indeed the λ index:

$$\lambda = N^{-1}\sum_{i=1}^{N}\Lambda_l \quad (13)$$

The resulting λ in Eq. 13 is bounded between 0 and 1, where 0 corresponds to absence of synchronization and 1 to synchronization in the free-noise case. This method can be then generalized in the case of *n:m* locking, simply rescaling the phases as $\phi_1 \rightarrow \phi_1/n$ and $\phi_2 \rightarrow \phi_2/m$.

In this study, the λ index was calculated on windows of 1000 samples (5 s) overlapping by 50 samples (250 ms). λ values above a threshold of 0.7 were considered significantly coupled. This value was empirically identified from visual inspection of the synchrograms of the available recordings. From the λ Index, the percentage of time spent in each particular *n:m* ratio (2:1, 3:1, 4:1, 5:1, 3:2, 5:2 etc.) and the average length of all the periods were calculated. Giving the assumption that in each instant of time the ratios of synchronization are mutually exclusive, we also estimated the percentage of time spent in a synchronized state adding together all the ratios with respect to a single breathing cycle. It results in the so-called *n*:1 synchronization and the average duration of period in a synchronized state. An example of this analysis is portrayed in Figure 22: Panels a) and b) show 1-minute segments respectively of HR and respiration, Panel c) reports the corresponding synchrogram and Panel d) the relative λ index, with a horizontal line showing the 0.7 threshold.

Directionality index

To trace the changes in coupling degree and/or directionality imposed by slow drift of the coupling parameters in the system, the index of directionality $d^{RR,RESP}$ was computed from the

time series of phase data: ϕ_{RR} and ϕ_{RESP}. In the following paragraphs we briefly summarize the model presented by Rosenblum et al. [184]. A simple model of two coupled phase oscillators is proposed, where each system can be represented by its own phase variable ϕ so that its time variation $\dot{\phi}$ can be expressed as $\dot{\phi} = \omega$, with $\omega = 2\pi/T$ being the natural frequency of the considered oscillator and T the period of oscillation. The phase space of the model can be expressed as:

$$\begin{aligned} \dot{\phi_1} &= \omega_1 + \varepsilon_1 \cdot f_1(\phi_2, \phi_1) + \zeta_1(t) \\ \dot{\phi_2} &= \omega_2 + \varepsilon_2 \cdot f_2(\phi_1, \phi_2) + \zeta_2(t) \end{aligned} \tag{14}$$

The continuous phase variables $(\dot{\phi_1}, \dot{\phi_2})$ consider the natural angular frequency of the system (ω_1, ω_2) and the random terms (ζ_1, ζ_2), accounting for amplitude fluctuations and perturbations which are intrinsic characteristics of biological systems. The coupling terms consist of 2π-periodic functions (f_1, f_2) and strength of interaction parameters $(\varepsilon_1, \varepsilon_2)$.

Given the assumption that phase variables can be estimated directly from the measured time series, it is possible to obtain an approximated reconstruction of both cardiac and respiratory oscillators to understand the causal relationship between the subsystems. In this work, the Evolution Map Approach (EMA) [184] algorithm has been used: it is capable of revealing asymmetric directionality strength from short noisy records and quantifying which of the systems under analysis influences its counterparts more strongly.

Both ϕ_1 and ϕ_2 are unwrapped phase variables, defined as continuous quantities represented on the whole real line, not limited from 0 to 2π. In this work the phase variable increments are computed using a fourth order Savitzky–Golay filter. Eq. 15 shows that to reconstruct the real weakly coupled oscillators from a single recorded realization of the process, it is necessary to fit the dependences of Δ_1 and Δ_2 over ϕ_1 and ϕ_2 upon considering phase variable increments generated by an unknown two-dimensional noisy map:

$$\begin{aligned} \Delta_1(k) &= \omega_1\tau + \mathcal{F}_1[\phi_2(t_k), \phi_1(t_k)\,] + \xi_1(t_k) \\ \Delta_2(k) &= \omega_2\tau + \mathcal{F}_2[\phi_1(t_k), \phi_2(t_k)\,] + \xi_2(t_k) \end{aligned} \tag{15}$$

Δ_1 and Δ_2 represent the phase variable increments over time computed as differences over a specific temporal window of length τ. They can be computed from phase variables ϕ_1 and ϕ_2. The deterministic part $\mathcal{F}_1$ and $\mathcal{F}_2$ of the map can be estimated as Eq. 16 shows, fitting the dependences of Δ_1 and Δ_2 over ϕ_1 and ϕ_2 with a least mean square approach. Giving the assumption that phase variables are cyclic, the most appropriate choice of a family function is the finite Fourier series:

$$\begin{aligned} \mathcal{F}_1 \approx F_1 &= \sum_m A_m \; e^{ia\phi_1 + ib\phi_2} \\ \mathcal{F}_2 \approx F_2 &= \sum_n A_n \; e^{ia\phi_1 + ib\phi_2} \end{aligned} \tag{16}$$

In this analysis the maximum order of Fourier expansions is set to 3, in the following computation $|a| \leq 3$ and $|b| \leq 3$. F_1 and F_2 can describe the deterministic (ω and $\mathcal{F}$) and stochastic (ξ) link between phase variables and their increments. They can be also seen as smoothing functions because they are able to filter out noise by means of the least square fitting.

The cross-dependency coefficients of phase dynamics of the two systems can be extracted from F_1 and F_2 as:

$$\begin{aligned} c_1^2 &= \iint_0^{2\pi} \left(\frac{\partial F_1}{\partial \phi_2}\right)^2 d\phi_1 d\phi_2 \\ c_2^2 &= \iint_0^{2\pi} \left(\frac{\partial F_2}{\partial \phi_1}\right)^2 d\phi_1 d\phi_2 \end{aligned} \tag{17}$$

The directionality index can be expressed as:

$$d^{(1,2)} = \frac{c_2 - c_1}{c_1 + c_2} \tag{18}$$

The EMA algorithm computes a normalized directionality index *d*. Index *d* varies from 1 to -1. In the case of unidirectional coupling from S_1 to S_2, *d* is 1, in the opposite case when the unidirectional coupling is from S_2 to S_1, *d* is -1. Positive intermediate values of d express a stronger or weaker S_1 to S_2 coupling strength, the negative intermediate values a coupling strength in the opposite direction (S_2 to S_1). In the case of absence of interaction when $c_2 = c_1$, d is zero.

Time and frequency domain

Traditional time domain and frequency domain parameters have been computed along with the novel approaches based on phase analysis. Time domain parameters are computed in a univariate fashion for HRV. Frequency domain analysis used a parametric approach. The autoregressive model order is set to 10. The univariate frequency analysis is performed on RR series, while the bivariate approach consisted of the cross-spectrum of RR series and respiration. The frequency bands are grouped as: Very Low Frequency (VLF) 0.01-0.04 Hz, Low Frequency (LF) 0.04-0.2 Hz, High Frequency (HF) 0.35-1.5 Hz.

Statistical analysis

Parameters were checked for normal distribution. High frequency parameters for RR series (HF_RR) and bivariate analysis (HF_biv), together with measures of phase locking (n:1 synchro %, n:1 synchro duration) did not pass the test and were transformed by means of square root operation. Hours of life at time of testing (HoL), infant sex and mode of delivery (MoD) were set as covariates for all analyses.

The effect of sleep state and GA and their interaction on time domain, frequency domain and phase locking parameters was tested with multiple 2-way Ancovas, and post hoc analyses were performed to identify differences within specific GA groups. A 2-way multivariate Ancova was performed to test the effect of sleep state and GA and their interaction on the directionality index *d* and the respiration frequency. Follow-up univariate analyses were also performed to test which parameter were driving any significant differences. All statistical analyses were performed with IBM SPSS software.

Results

Time domain and frequency domain results presented in Table 11, show increasing mean RR interval (RR_mean) and short-term HRV (RMSSD) and linear cardiorespiratory coupling (HF_RR, HF_biv) with GA, indicating an increasing cardiorespiratory coupling and autonomic control as a function of GA at birth (respectively, $p=0.001$, $p<0.001$, $p=0.007$, $p=0.008$). Subsequently, cardiorespiratory coupling was analyzed in term of phase coupling (n:1 synchro %, n:1 synchro duration) and directionality (*d*). The main results indicate no relationship between time spent in phase-synchronized state and GA group. Differences were instead detected with sleep state in all GA groups, with more frequent and longer synchronization in QS. No interaction effect

was found between the two independent variables, GA and sleep state (respectively p=0.874 and p=0.740). Results from statistical analysis are reported in Table 12 and Table 13.

The multivariate test with *d* and breathing frequency as dependent variable and GA and sleep state as independent, showed significant differences by GA and state (respectively: *F*(4.588)

Table 11. Descriptive statistics (mean ± std) of time and frequency domain parameters by gestational age (GA) in quiet sleep (QS) and active sleep (AS). The groups are: late preterm (LPT), early term (ET), and full term (FT).

state	GA group	RR_mean	SDNN	RMSSD	HF_RR	HF_biv
QS	**LPT**	0.476 ± 0.034	21.32 ± 11.88	10.20 ± 6.83	0.377 ± 0.190	0.556 ± 0.193
	ET	0.492 ± 0.042	22.87 ± 10.85	13.23 ± 7.06	0.464 ± 0.178	0.597 ± 0.192
	FT	0.507 ± 0.040	24.59 ± 11.47	15.72 ± 7.67	0.497 ± 0.175	0.663 ± 0.196
AS	**LPT**	0.482 ± 0.040	28.80 ± 12.19	11.60 ± 5.40	0.292 ± 0.122	0.379 ± 0.127
	ET	0.486 ± 0.044	30.93 ± 12.90	12.29 ± 6.09	0.307 ± 0.137	0.395 ± 0.133
	FT	0.497 ± 0.050	31.83 ± 13.87	13.40 ± 6.51	0.321 ± 0.135	0.413 ± 0.131

Table 12. Descriptive statistics (mean ± std) of the square root of the percentage of time in synchronized state and the average synchronization duration and number of subjects by gestational age (GA) in quiet sleep (QS) and active sleep (AS).

state	GA group	n:1 synchro %	n:1 synchro duration	N Subjects
QS	**LPT**	0.436 ± 0.209	2.99 ± 1.02	19
	ET	0.454 ± 0.184	3.03 ± 1.01	41
	FT	0.473 ± 0.170	3.00 ± 0.90	58
AS	**LPT**	0.185 ± 0.124	1.82 ± 1.04	54
	ET	0.218 ± 0.127	2.07 ± 0.84	62
	FT	0.221 ± 0.139	1.92 ± 0.97	68

Table 13. 2-way Anova of the square root of the percentage of time in synchronized state and the average synchronization duration by gestational age (GA) and sleep state along with their interaction effect.

Parameter		p-val	Partial η^2
n:1 synchro %	**state**	0.000	0.357
	GA group	0.342	0.007
	state * GA group	0.874	0.001
n:1 synchro duration	**state**	0.000	0.214
	GA_group	0.559	0.004
	state * GA group	0.740	0.002

= 4.151, *p* = 0.003; Wilks' Λ = 0.946 and F(2.294) = 56.582, p = 0.000; Wilks' Λ = 0.722). No significant interaction was detected between GA and state (F(4.588) = 0.130, p = 0.972; Wilks' Λ = 0.998). Results are reported in Table 14. The test of between-subjects' effects highlighted how each dependent variable differed based on the independent variables. Results are reported in Table 15. To account for multiple ANOVAs, we used a Bonferroni correction accepting statistical

significance at $p < 0.025$. This test showed that sleep state significantly influenced both *d* and respiratory frequency, while GA at birth influenced significantly only *d*. Directionality index *d* decreased with GA at birth. In QS all the three GA groups show dominant influence of breathing on HR ($d<0$), but this influence grows with GA (*d* becomes more negative). On the other hand, in AS a balanced relationship is present in LPT ($d\approx0$) and it moves toward a dominant relationship from breathing to HR in FT ($d<0$). Post-hoc tests of the significant ANOVAs showed that among the GA groups, LPT were significantly different from ET which are also significantly different from FT.

Previous studies suggested a dependence between *d* and breathing frequency, arguing that the cardiac influence on respiration is weak and frequency independent, while the coupling from the respiration to HR is strong compared with the strength of the cardiac influence for low respiratory frequencies, and of similar entity for higher frequencies [218]. The same group proposed a threshold of 0.6 Hz to investigate this mechanism in infants. Figure 23 shows the directionality index histograms of the three GA groups at breathing frequency ≥0.6 Hz (in each panel on the right) and at breathing frequency <0.6 Hz (in each panel on the left). This figure illustrates the occurrence of this bimodal influence: concurrently with breathing frequency <0.6 Hz, an increased polarization toward negative values of d occurs, especially in ET and FT. Nonetheless, the significant evolution of d with GA (p=0.001) is not accounted for by this mechanism, given that respiratory frequency does not change significantly with GA at birth (p=0.881). A graphical summary of the tested quantities is reported in Figure 23 and Figure 24.

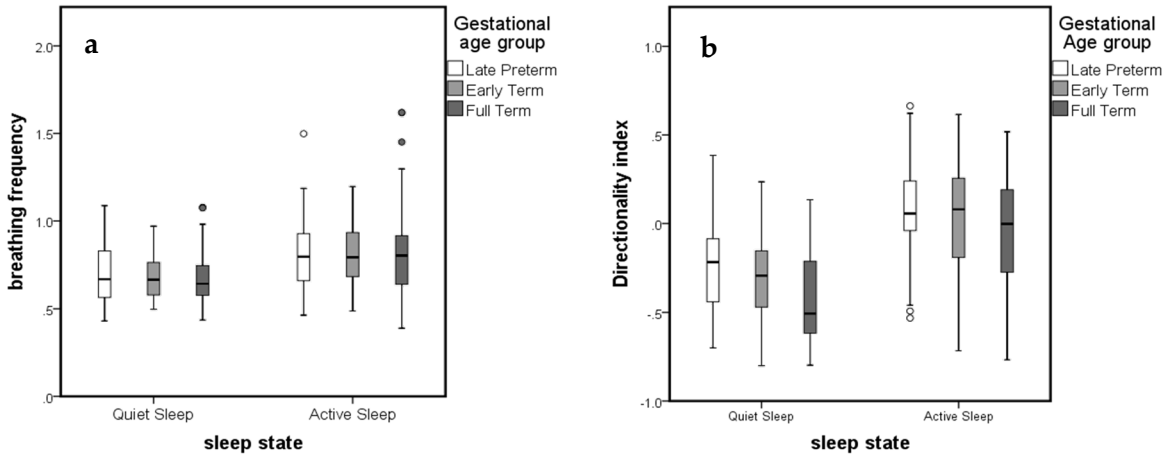


Figure 23. a): Boxplots show breathing frequency by gestational age (GA) and sleep state. Breathing frequency is not significantly different by gestational age at birth but is significantly different by sleep state (p-value<0.001). b): Boxplots show directionality index by gestational age and sleep state. Directionality index is statistically different both by gestational age at birth (p-value=0.001) and by sleep state (p-value<0.001). Neither breathing frequency nor directionality index show any interaction effect between gestational age and sleep state.

Table 14. Descriptive statistics (mean ± std) for directionality index and frequency of respiration by gestational age (GA) and sleep state and number of subjects.

state	GA group	Directionality index	resp freq	N SUBJECT
QS	**LPT**	-0.240 ± 0.292	0.699 ± 0.181	19
	ET	-0.307 ± 0.266	0.676 ± 0.126	41
	FT	-0.417± 0.259	0.668 ± 0.149	59
AS	**LPT**	0.083 ±0 .262	0.815 ± 0.200	55
	ET	0.044 ± 0.306	0.808 ± 0.180	62
	FT	-0.028 ± 0.284	0.818 ± 0.237	68

Table 15. Multivariate model analysis of directionality index and frequency of respiration by gestational age (GA) and sleep state along with their interaction effect and post hoc analysis.

Parameters		p-val	Partial η^2	Post hoc
Directionality index	**GA group**	0.001	0.049	LPT>ET>FT
	state	<0.001	0.273	
	GA group * state	0.816	0.001	
Resp freq	**GA group**	0.881	0.001	LPT>ET>FT
	state	<0.001	0.097	
	GA group * state	0.857	0.001	

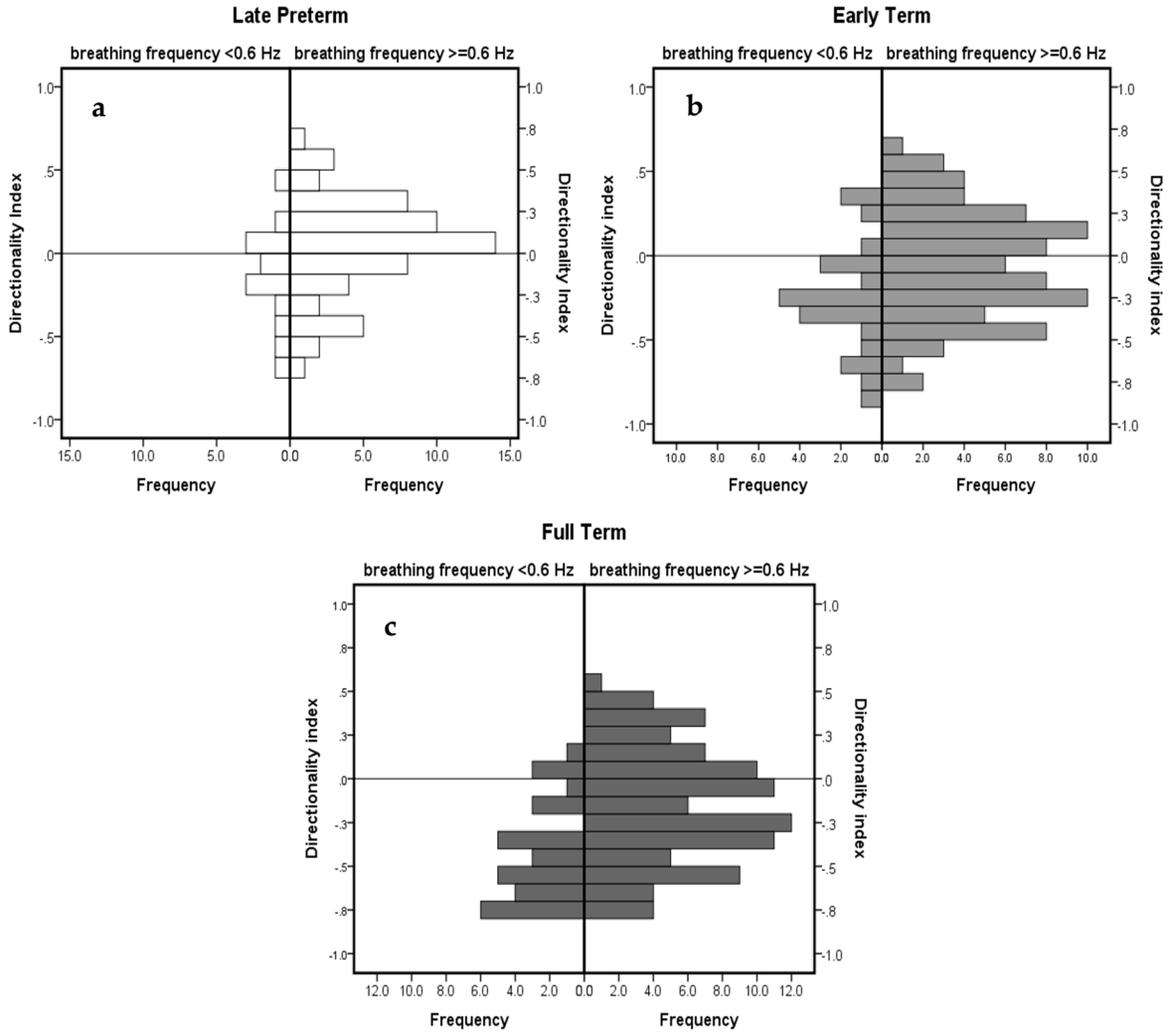


Figure 24. a): Histograms of directionality index distribution are plotted, on the left for late preterm subjects with a breathing frequency < 0.6 Hz, and on the right for late preterm subjects with a breathing frequency ≥ 0.6 Hz. b): analogous histograms for early term. c): analogous histograms for full term. Histograms for subjects with a breathing frequency < 0.6 Hz in all three panels show a progressive shift of directionality index towards negative values, indicating a growing unidirectional interaction characterized by a driving effect of respiration over heart rate.

Discussion

The goal of this investigation was to address the emergence of directionality in cardiorespiratory phase synchronization within the last weeks of pregnancy. In particular, we aimed at evaluating whether infants born between the 35th to 40th week of gestation exhibit preferential directionality interactions between the cardiac and the respiratory system. To this end, we analyzed HRV and respiration of healthy newborns as a function of gestational age at birth. Previous analyses on our study population explored the use of traditional time and frequency

domain and entropy analyses [219], [220], which suggested that LPT and ET autonomic control of HR is significantly less mature and their cardiorespiratory regulation less developed when compared to FT. However, previous approaches solely investigated linear coupling between HR and breathing; and, due to the type of analyses used in those previous studies, no inferences were possible regarding the directionality of the interaction between the cardiac and the respiratory system.

In the current study, time domain and frequency domain analyses confirmed previous findings, with increasing HRV and linear coupling with of increasing GA. We then investigated cardiorespiratory coupling from a phase relationship perspective. This concept is based on the finding that cardiorespiratory systems are intimately linked with one another. Consequently, illnesses or distress may disrupt this complex connection. Our results confirmed the dependence of phase-locking with sleep state, as already found in previous studies in infants and adults [221], [222]. Moreover, results indicated that GA at birth differences are not apparent in the duration of phase locking, but rather in the directionality of the relationship between the cardiac and the respiratory system. Previous work highlighted a change in the quantity of synchronization during the first months of life [217], [221]. Thus, it was surprising to discover that the quantity of synchronized time did not change significantly within the last weeks of GA. This might be due to the fact that cardiorespiratory system mainly develops in the months following birth. In summary, these differences between previous and current findings might be related to differential influences of intra-uterine versus extra-uterine life (rather than of GA at birth).

Regarding directionality, in both sleep states a GA-related shift occurs toward a stronger driving influence of respiration on HR. Previously, another research group reported a similar developmental shift but in full term newborns followed during the first months of life, suggesting that the maturation occurring in the last stage of pregnancy continues in the first months of life until a final set point is reached in later childhood [218]. In the current study, in QS lower values of d were observed in all GA groups signifying a stronger influence of breathing on HR, as previously demonstrated with other coupling measures [220].

Mrowka et al. [217] have hypothesized that directionality depends on respiratory frequency. In particular, respiration rate would act as a low pass filter, i.e. below a set respiratory frequency directionality is mainly from respiration to HR, whereas above a certain respiratory frequency threshold the interaction becomes bidirectional. The underlying mechanisms for this

low pass filter effect could be related to lower information transmission to the cardiac oscillator caused by reduced information from the vagal nerve to the atrial pace-maker cells when respiration frequency is above a certain threshold, such as 0.6 Hz. [218].

In this study, breathing rate does not change significantly in this specific GA window (35-40 weeks); thus, the change in directionality with GA at birth cannot be explained solely because of breathing frequency. One explanation for this phenomenon could be that the threshold for the low pass filter effect is still adapting between 35-40 weeks GA. At a younger GA, for instance LPT, the value for the cutoff frequency might be lower when compared to more mature conditions, such as FT. Given that this cutoff frequency is potentially related to vagal nerve regulation, these findings would be consistent with previous studies showing immature vagal function in premature infants. Morphological studies demonstrate a rapid developmental increase in number of myelinated vagal fibers with postconceptional age (PCA), and by 40 weeks after conception, total fiber counts were comparable to those of adolescents. Interestingly, the number of total myelinated vagus fibers in preterm infants (≤38 weeks PCA) was found to be significantly smaller than for the term or adolescent age groups [223]. The question remains as to physiological meaning behind this type of nonlinear coupling. One possible explanation could be that this type of synchronization involves an energy consumption benefit, namely, the reduction in intrathoracic pressure during inspiration increases cardiac filling and consequently cardiac output. However, this hypothesis needs further investigation.

In conclusion, it is well known that the last weeks of gestation are crucial regarding development of central nervous system interconnections that are responsible for cardiorespiratory regulation. Thus, we hypothesize that the directionality imbalance shown by FT is an important step in the process of autonomic maturation, given that it evolves toward the unidirectional interaction observed later in childhood and adulthood [218]. The differential directionality of the cardiorespiratory profile in LPT and ET infants, as compared to FT, might provide insight into mechanisms underlying the increased risk of biological and developmental pathology in these GA groups. Clark et al., using similar measures of cardiorespiratory interaction, found that coupling increases with chronological age postnatally, but the rate of increase was not affected by GA at birth [224]. This indicates that development of cardiorespiratory autonomic control is a postnatal age-dependent phenomenon. The increased occurrence of apparent life-threatening apneic events

(ALTE) and sudden infant death syndrome (SIDS) in premature infants during the first months of life, suggests a longer window of adaptation associated with several high-risk conditions.

It has been shown that the increased incidence of SIDS in premature infants is related to an underlying cardiorespiratory vulnerability. Defining factors underlying this increased vulnerability in preterm infants remains an important public health concern [225]. Previous studies have investigated regulatory systems associated with infants' responses to threatening stimuli, such as hypoxia, asphyxia or hyperthermia. This phenotype can increase risk vulnerability to exogenous stressors, such as the prone sleeping position.

This paper proposes a novel approach to address the clinical question of late preterm higher morbidity and mortality in the context of the new discipline of Network Physiology. Looking at different physiological systems as dynamically interacting could shed light on the process of horizontal integration at the level of organ to organ interaction required to maintain an optimal health status. This work is convergent with the goals proposed by Ivanov et al. to develop building blocks for the creation of an atlas of dynamic organs interactions [226]. Moreover, the unique contribution of this study is the early developmental aspect in the newborn period given the dynamic physiologic relationship within rapidly developing organs. One major focus in our future investigations will be testing the use of different types of surrogates to optimally define a threshold of significance for λ, the quantitative measure of phase coupling, rather than using a fixed value. Further studies will extend the focus on the interactions at different time scales and the investigation of the evolving relationships among the heart, lungs and the brain [31]. Our major conclusion is that infants born even only 1-4 weeks early show irregular cardiorespiratory characteristics with respect to full term, suggesting a crucial role for last weeks of pregnancy in the maturation of the interaction between the cardiovascular and the respiratory systems. Moreover, the proposed measures of cardiorespiratory coupling provide a tool to assess maturity of cardiorespiratory regulation, thus serving as a potential biomarker for risk stratification.

Transfer Entropy Modelling of Newborn Cardiorespiratory Regulation

Introduction

Premature birth and related complications are the leading cause of death under 5 years of age across the world [227]. According to the March of Dimes, in the United States, the percentage rate of preterm birth in 2019 was 10.00%, marking the third consecutive year of increase after 7 years of decline [228]. Epidemiological studies have shown that late preterm (LPT: $34^{0/7}$ - $36^{6/7}$ weeks of gestational age (GA)) infants have significantly more medical problems, resulting in markedly increased hospital costs compared with full term infants (FT: $39^{0/7}$-$40^{6/7}$ GA) [229]. Data from a population study from 2006 to 2014 in the US showed that LPT birth rate was 6%, while early term (ET: $37^{0/7}$-$38^{6/7}$ GA) rate was 26.9% [230]. Late preterm and early term birth are associated with adverse neonatal outcomes, such as higher incidence of respiratory distress syndrome, temperature instability, hypoglycemia, hyperbilirubinemia, apnea, feeding problems, as well as higher rates of re-hospitalization and a two-fold increase in Sudden Infant Death Syndrome (SIDS) [171], [231], [232]. Limited sleep state regulation, frequent episodes of apneas, periodic breathing, altered pulmonary function, bradycardia, and diminished autonomic control of heart rate (HR) have been documented in these populations [176], [233]–[235].

Starting from a concept introduced by the new field of Network Physiology, the human organism can be viewed as a network of integrated and interacting physiological systems [226], [236]. Thus, given the described adverse conditions related to imbalances of both cardiac and respiratory systems, investigation of risks associated with late prematurity should include a focus on the dynamic interaction in the cardiorespiratory network. Regulation and autonomic control of respiratory and cardiovascular interactions are crucial for the maintenance of homeostasis during sleep [237]. In adults, many studies have shown evidence that cardiorespiratory imbalance is associated with obstructive sleep apnea and heart failure, resulting in higher sympathetic tone and potentially ultimately triggering life-threatening events [238]. Similarly, it has been reported that nocturnal perturbations of cardiac and respiratory systems in newborns play a crucial contributory role [239]. Despite the high clinical relevance there is a paucity of research data about the mechanisms related to cardiorespiratory interactions early in life when the primary control systems are still developing.

Many approaches have been proposed in the past to address the complex interaction of the cardiorespiratory system, from simple time and frequency domain measures [240] to more complex ones, such as those based on information theory [198], [206], [241]. These, in particular Transfer Entropy (*TE*), are progressively gaining interest as model-free approaches which quantify directional interaction between subsystems and are thus sensitive to both linear and nonlinear interactions. In prior publications the existence of several co-existing forms of cardio-respiratory coupling [197] has been shown, and our group has also addressed this topic analyzing cardiorespiratory interaction with regards to entropy and phase locking [176], [220]. In the current report, we propose a new application of *TE* measure to provide an estimation of information transfer between the cardiac and the respiratory system at various lags. The focus on the timing of such interactions will augment descriptive approaches for assessing cardiorespiratory interplay at various time scales. Specifically, we are interested in characterizing such system crosstalk in a population of LPT, ET, and FT infants. This investigation aims to provide insight into developing control systems involved in the cardiorespiratory regulation and how prematurity affects this complex interaction. This could inform interventions aimed at reducing risk for morbidities and mortality in this population.

Materials and methods

Lagged Transfer Entropy

For our proposed framework, we modelled a dynamical system composed of two interacting sub-systems ($M = 2$), whose visited states can be represented by discrete-time stationary stochastic processes, namely X and Y. In this context, *TE* aims at evaluating the information transfer by the past states of the process X about the present of the process Y, that is not already provided by the past states of Y (directionality $X \rightarrow Y$) and vice versa (directionality $X \rightarrow Y$) [242].

We define x_n, y_n as the stochastic variables representing the present states of the processes X and Y at a given time point n, with $n < N$ and N = length of the signals, and $x_{1:n-1}$, $y_{1:n-1}$ the vectors of their respective past states.

TE is defined accordingly to Eq. 19:

$$TE_{X \rightarrow Y} = \sum p(y_{1:n}, x_{1:n-1}) log \frac{p(y_n | x_{1:n-1}, y_{1:n-1})}{p(y_n | y_{1:n-1})} \quad (19)$$

where the sum incorporates all states visited by the subsystems.

Similarly, the formulation of *TE* can be expressed in terms of the difference between two Conditional Entropy (*CE*) terms as shown in Eq. 20:

$$TE_{X\to Y} = H(y_n|y_{1:n-1}) - H(y_n|x_{1:n-1}, y_{1:n-1}) \quad (20)$$

The previously reported *TE* formulations encompassed an aggregate measure of information transfer which is not candidate-specific, where candidate refers to one of the elements contained in the vectors employed to reconstruct the past of processes *X* or *Y* at the instant *n* defined as $x_{1:n-1}$ and $y_{1:n-1}$.

In this work, to disambiguate the contribution of different candidates towards the estimate of *TE*, we employed the approach described in [243]. Given the *TE* formulation expressed in Eq. 20, we computed *TE* based on a sequential procedure for non-uniform conditioning, where the conditioning vector is updated progressively by selecting the candidate which reduced the most uncertainty in explaining the target variable. The initial set of candidates was defined including a predefined maximum number of past states, i.e., $\Omega = \{X_{n-1}, X_{n-2}, ..., X_{n-Lmax}, Y_{n-1}, Y_{n-2}, ..., Y_{n-Lmax}\}$. In this work, the maximum number of candidates (L_{max}) was set equal to 10. Candidates were progressively selected among the elements of Ω as described in Faes et al., 2014 [243]. Once the selection procedure has terminated, the vectors of candidates for both *X* and *Y* processes were produced and defined as $Vk = [V_k^X, V_k^Y]$. Thus, they were suitable to be employed as conditioning vectors for further *TE* estimations. Given the reported notation, Eq. 20 can be rewritten employing conditioning vector formulation as reported in Eq. 21:

$$TE_{X\to Y} = H(y_n, V_k^Y) - H(V_k^Y) - H(y_n, V_k) + H(V_k) \quad (21)$$

The final step to *TE* estimation relied on the computation of probability density functions to approximate the interrelationship between *X* and *Y*, based uniquely on single realizations of the two processes. The practical estimation of terms in Eq. 21 was based on the previously defined embedding vector (V_k) and it employed a Nearest Neighbor (*NN*) estimator. The combination of non-uniform embedding and *NN* estimator (*NN NUE*) has been reported to be optimal for *TE* estimation [244]. Furthermore, the statistical significance of computed *TE* was assessed using surrogate data implemented by time shift procedure. In this analysis, the number of employed

surrogate series was equal to 100, with a maximum allowed time shift of 20 samples. The significance threshold was set above the 95th percentile of the surrogate series distribution.

TE has been usually employed for a global measure of information transfer between time series. This work proposes a novel approach towards the quantification of the instantaneous and delayed effects among two interacting systems based on *TE* notion. This application lies its foundations on the previously described *TE* implementation, yet it considers several lagged versions of the original series [178].

In order to quantify $TE_{X\to Y}$ at a lag value (τ) equal to one, the target series was shifted forward of one sample so that $x(n)$ was aligned with $y(n-1)$. The lagged version of TE proposed in this approach aims at quantifying the information provided by the past of *X* on the shifted portion of the process *Y*, that is not already provided by the past of *Y* as reported in Eq. 22:

$$TE_{X\to Y} = \sum p(y_{1-\tau:n-\tau}, x_{1:n-1}) log \frac{p(y_{n-\tau}|x_{1:n-1}, y_{1-\tau:n-1-\tau})}{p(y_{n-\tau}|y_{1-\tau:n-1-\tau})} \quad (22)$$

The underlying idea is to quantify the source series effects on the target and the instantaneous and delayed effects between the two processes. The previously described computational implementations for *TE* are again employed for this proposed lagged version. For this analysis the maximum lag between *X* and *Y* series was set to 15. The statistical significance of *TE* estimations for each lag was tested with surrogate data, as previously described.

Validation

To provide validation of the proposed methodology, lagged TE was computed based on a dynamical system composed by $M = 2$ stochastic dynamic subsystems, namely *X* and *Y*, defined by Eq. 23:

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + 0.07 Y_{n-1} + U_n$$
$$Y_n = C_1 X_{n-1} + C_2 X_{n-2} + C_3 X_{n-5} + V_n \quad (23)$$

where U_n and V_n are independent white noises with zero mean and unitary variance. The autoregressive parameters a_1, a_2, C_1, C_2, and C_3 were set as described in Faes et al., 2014 [243]. Process *X* simulates a self-sustained rhythm with a characteristic frequency centered at $f = 0.1$ Hz. Parameters and delay settings were chosen to simulate two different types of interaction: one which is, lasting and strong taking into consideration the directionality $X \to Y$, the other which is transient

and weak for $Y \rightarrow X$. The simulated series length was set to $N = 300$ points, and the total number of generated series was equal to 100 for each lag, ranging from the unlagged version (lag = 0) of the series to their maximum lagged version (lag = 15). As previously described in the Methods section, the statistical significance of TE estimations for each lag was tested with surrogate data. Two one-way ANOVAs were performed to test the interaction between the fixed factor lag and each dependent variable ($TE_{X \rightarrow Y}$ or $TE_{Y \rightarrow X}$).

Experimental protocol and data preprocessing

For this analysis our dataset included 268 infants born at the Morgan Stanley Children's Hospital of New York between $35^{0/7}$ and $40^{6/7}$ weeks of GA. No participating newborn was admitted to the Neonatal Intensive Care Unit, and there was no evidence of major illness, genetic disorders. Also, there was no past/present medicated/non-medicated psychiatric complaints in the mothers. A minimum Apgar score of 8 after 5 min of life was required. Mothers signed informed consent forms prior to enrollment in the study. The Institutional Review Boards of the New York State Psychiatric Institute and Columbia University Irving Medical Center approved all consent and data acquisition procedures.

Participants who met inclusion criteria were tested 12–84 hours after birth (mean and standard deviation of hours of life = 48 ± 12 hours). Infants were grouped based on GA: LPT (N = 67), ET (N = 91), and FT (N = 110). Within ~30 min after feeding, infants were put supine to sleep and data acquisition lasted 10 min. ECG and respiration signals were acquired at 500 Hz and 200 Hz, respectively. ECG was recorded with three leads, placed on the infant's chest (left abdomen, left and right scapula) and the signal was amplified and recorded using the DATAQ Instruments ECG system (Medelex, NY, NY). A respiratory inductance belt (Ambulatory Monitoring Inc., Ardsley, NY) was placed around the infant's abdomen to measure the respiration signal. Sleep states were classified into active sleep (AS) and quiet sleep (QS) based on respiratory variability and confirmed by behavioral codes entered throughout the study to determine when infants were awake, crying, or fussy [211], [245]. The R peaks were detected on the ECG with proprietary software (Gmark, Ledano Solutions) based on the Pan–Tompkins algorithm and subsequently checked by visual inspection. The respiration signal was bandpass filtered (0.05–3.5 Hz). The thresholds of acceptance for RR interval were set as 0.3–0.667 s, with an absolute variation between consecutive RR intervals of 10%, while for respiration thresholds were 0.5–2.5 s (absolute change 40%). Segments with more than 5 % rejected samples were discarded from

further analysis. The RR series was then defined so that $RR(n)$, was the time interval between the n-*th* R peak and the successive one at a time $(n+1)$-*th*. Similarly, the n-*th* sample of resampled respiration series $RESP(n)$ was obtained by resampling the original respiration series at the onset of the n-*th* R peak which coincides with the time previously defined for $RR(n)$. Within the same sleep state, segments of 300 consecutive RR intervals (RR) and 300 respiration samples ($RESP$) were identified. The resulting series, $RR(n)$ and $RESP(n)$ with $n = 1, \ldots, 300$, were normalized to zero mean and unit variance to be employed for further analysis. The segments length was chosen based on previous studies, reporting 300 samples as appropriate for a reliable TE estimation as fulfilling the requirement of stationarity [220], [243]. The described preprocessing pipeline was necessary to avoid potential bias in the further analysis. Specifically, the effect of nonstationarities over entropy measures and estimators due to artifacts has been extensively shown in [246].
The total number of analyzed segments was 661, 392 in AS and 269 in QS. The MuTE toolbox was employed for computing Transfer Entropy [247].

Two-way ANOVAs tested the effect of fixed factors lag (0:15) and GA (LPT, ET, FT) on $TE_{RR\rightarrow RESP}$ and $TE_{RESP\rightarrow RR}$ for each sleep state. Sex, mode of delivery (MoD), and hours of life (HoL) were included in the statistical model as covariates. Significance for fixed factors as well as their interactions were tested. A series of post-hoc tests were performed: simple main effects and specific group differences. Statistical analysis was conducted with IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Results

Validation data

The top panel of Figure 25 shows $TE_{X\rightarrow Y}$ as a function of lag, computed on a simulated dataset. The behavior of $TE_{X\rightarrow Y}$ is in accordance with the simulated interaction between subsystems X and Y. Specifically, $TE_{X\rightarrow Y}$ exhibits a strong and stable influence of process X over process Y for lags 0 to 5, where the information transfer between the two series is expected to be maximum given that X_{n-1}, X_{n-2}, and X_{n-5} are effectively contributing to modulate the target series Y. The rapid reduction in TE at lags equal to 6 and 9 are consistent with the set delays. Specifically, at lag 6, the past state X_{n-5} of process X cannot be included in the conditioning vector anymore, given the chosen maximum candidate delay $L = 10$. Accordingly, a net decrease in TE is noticed when passing from lag = 5 to lag = 6. Analogous reasoning applies when moving from lag = 8 to lag =

9. Lastly, from lag = 10 on, the mutual influence in the directionality $X \rightarrow Y$ becomes negligible given the loss of interaction between the two sub-systems, thus resulting in *TE* estimates close to zero. Statistical analysis reveals a significant effect of lag over $TE_{X \rightarrow Y}$ (p-value < 0.001). Bonferroni post-hoc tests showed significant differences of lag 0-5 vs lag 6-15, lag 11-15 vs lag 0-10, and lags 6, 7, 8, 9, 10 are significantly different from each other.

Distribution of selected *RR* candidates included in the conditioning vector V_k^X referring to *TE* at lag = 0 is displayed in the top panel of Figure 26. The frequency of selected candidates is in accordance with the simulated interaction delay between the two series, namely X_{n-1}, X_{n-2}, and X_{n-5}.

With regard to $TE_{Y \rightarrow X}$, values are stable across all the lags and noticeable lower when compared with estimates for $TE_{X \rightarrow Y}$, as presented in the bottom panel of Figure 25. The absence of any $TE_{Y \rightarrow X}$ significant differences by lags reflects the weak and transient influence of the information transfer for this directionality. Uniformly, only Y_{n-1} results the preferred candidate as

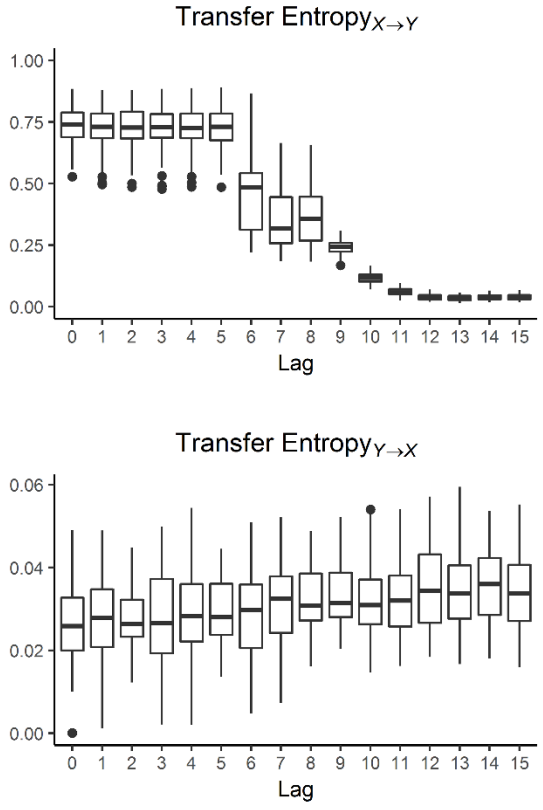


Figure 25. TE estimates for the directionality X→Y (top) and Y→X (bottom) computed on simulate dataset. $TE_{X \rightarrow Y}$ exhibit a marked influence as lags are progressively increasing. On the contrary, no influence of lags over $TE_{Y \rightarrow X}$ is detected.

it effectively contributing to modulate the target series *X* as a standalone past sample of series *Y*, as shown in the bottom panel of Figure 26.

Cardiorespiratory data

Using the same approach previously described for the simulated dataset, *TE* estimations across lag and GA, as well as *RR* and *RESP* candidate distributions were computed. Additionally, the statistical significance of *TE* estimations for each lag was tested employing surrogate data. The subsystems' interaction for the directionality RR→RESP exhibited a long-lasting and steady effect of cardiac system modulation over the respiratory system, as shown in the top two panels of Figure 27. No differences were found when comparing $TE_{RR \rightarrow RESP}$ across lags in the interval 0-9, consistently for both QS and AS.

Analogous behavior was displayed in the interval 10-15. In contrast, post-hoc tests revealed significant differences comparing each lag in the interval 0-9 vs 10-15. Significant GA group differences were found when considering estimates of $TE_{RR \rightarrow RESP}$ in AS only. Specifically, we observed an average increase in information flow across GA. This was confirmed by the post-hoc

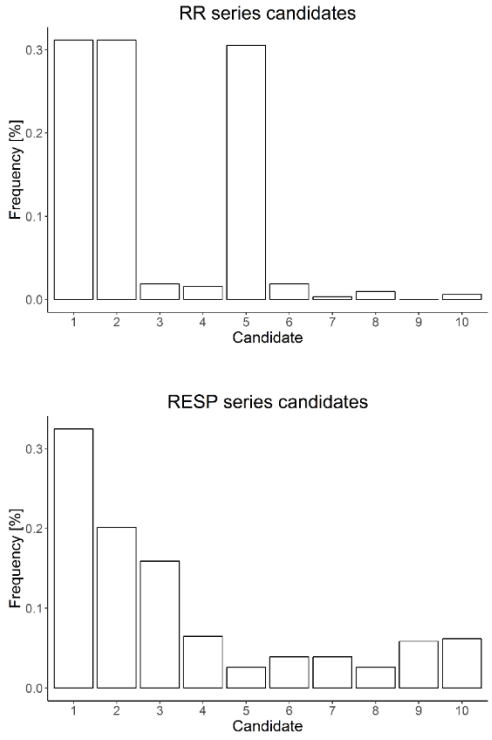


Figure 26. X (top) and Y (bottom) series candidate distributions employed for TE computations considering lag = 0. RR series candidate distribution shows a proper selection of the simulated delays. On the other hand, RESP candidate distribution reflects the weak modulation effect of series Y on X.

test comparing LPT vs ET ($p < 0.001$), LPT vs FT ($p < 0.001$), and ET vs FT ($p < 0.001$). *RR* candidates employed in estimating $TE_{RR \rightarrow RESP}$ at lag = 0 exhibit a similar frequency of selection for both QS and AS. Moreover, when investigating the role of GA for candidate frequency of selection, no differences are found across age.

In contrast to what previously reported for $TE_{RR \rightarrow RESP}$, $TE_{RESP \rightarrow RR}$ exhibited, the current study demonstrated a decrease in information transfer from *RESP* to *RR* as lags were progressively increasing in both QS and AS states. As confirmed by statistical analysis and shown in the two bottom panels of Figure 27, no significant differences among lags were found for lags > 3. Given this finding we limited the analysis to a restricted poll of lags, specifically 0-3, with the aim of avoiding over-representing similar class distributions in the successive analysis. The statistical analysis performed on the subgroups of lags for $TE_{RESP \rightarrow RR}$ showed a significant effect of lag as a fixed factor. Specifically, each lag was consistently different from each other for both QS and AS. Analyses among GA groups report no differences for $TE_{RESP \rightarrow RR}$ in AS. Significant differences were evident in QS. A summary of statistical results is reported in Table 16. The candidate analysis reported a behavior characterized by a prevalent selection of Y_{n-1} and Y_{n-2} candidates for both sleep states as displayed in Figure 28.

Given the differences of $TE_{RR \rightarrow RESP}$ and $TE_{RESP \rightarrow RR}$ as a function of GA, we decided to investigate the role of breathing rate for our model. The rationale for investigating breathing frequency was based on previous studies showing differences by sleep state (higher breathing rate in AS) but not by GA [176]. Thus, we hypothesized that breathing rate is partly mediating the interaction between sleep states and TE.

In this analysis, we first tested (Sobel test) a model having sleep states as the independent variable (IV), *TE* as the dependent variable (DV) and breathing frequency as a mediator (M). Partial

Table 16. Results of two-way ANOVA tests on cardiorespiratory data.

Directionality	Fixed factor	Quiet sleep	Active sleep
RR→RESP	Lag	< 0.001	< 0.001
	GA	0.120	< 0.001
	Lag*GA	1.000	0.897
RESP→RR	Lag	< 0.001	< 0.001
	GA	0.050	0.183
	Lag*GA	0.627	0.953

mediation analysis quantifies the decrease in correlation strength among two factors, once a specific mediator is introduced in the model. Specifically, when considering $TE_{RR \rightarrow RESP}$ as DV, the correlation between IV and DV was significant (p-value < 0.001) and the mediation effect of breathing frequency was equal to 11%, similar results were obtained considering $TE_{RESP \rightarrow RR}$ as DV (p-value = 0.001 and 14%). On the other hand, when testing GA as the DV, no significant mediation was reported in either TE directionality.

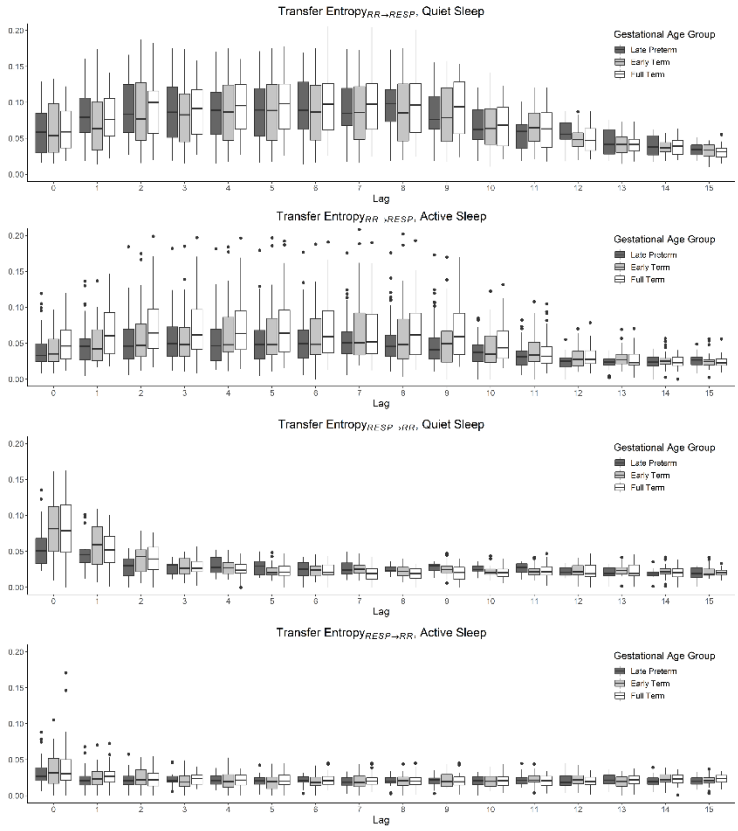


Figure 27. X (top) and Y (bottom) series candidate distributions employed for TE computations considering lag = 0. RR series candidate distribution shows a proper selection of the simulated delays. On the other hand, RESP candidate distribution reflects the weak modulation effect of series Y on X.

Discussion

The goal of this study was to analyze the maturation in terms of the evolution of the cardiorespiratory networks in terms of information flow dynamics in a population of newborns during sleep. Gaining insight on such interactions attains the potential for assessing individual differences in neonatal control mechanisms and vulnerability for the reported higher morbidity

and mortality rates in LPT and ET newborns [230]. Investigation of the neurophysiological

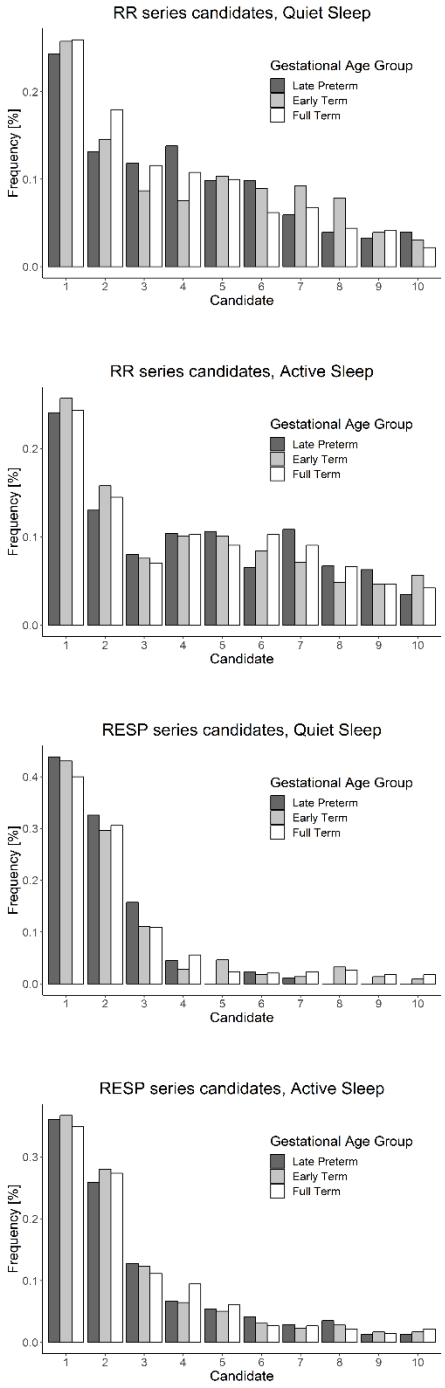


Figure 28. RR (top two panels) and RESP (bottom two panels) series candidate distributions employed for TE computations of cardiorespiratory data. RR candidates are selected uniformly among all possible candidates. On the opposite considering the respiratory signal, the first two possible RESP candidates are selected with higher probability.

mechanisms responsible for cardiorespiratory regulation is challenging, due to their intrinsic complexity and the requirement to employ noninvasive monitoring. Quantitative analysis of cardiorespiratory interactions in the newborn nursery represents a valuable investigation tool. Moreover, the derived parameters provide a window of opportunity to observe non-invasively the interaction between sympathetic and parasympathetic nervous systems and their capability to timely respond to internal and external challenges.

In this report, we propose *TE* as an optimal method to overcome the limitations of traditional linear association-based methodology to investigate the above-mentioned interaction [244]. The advantages of the proposed approach are multiple. Firstly, it is model-free, i.e., it does not require any a priori assumption regarding the systems generating the observed data. This is crucial in the neonatal context, given that control systems at birth are still developing and the typical cardiorespiratory models for adults thus cannot be applied. Even in the context of integrated system physiology in adult subjects, general models cannot be directly employed but often need to be modified and adapted in accordance with the observed dynamics. Secondly, *TE* is a measure encompassing the dynamics of information transfer and thus it provides an indication of directionality [242]. This is particularly important as it is well documented that respiratory and cardiovascular rhythms influence each other due to central as well as peripheral nervous mechanisms of interaction [31]. Despite these two important advantages there is one intrinsic limitation of traditional TE techniques, namely the reliance on a single global measure of information. Thus, *TE* lacks information about different time scales of information flow across subsystems. Gaining an understanding of the different time scales at which the vagal and sympathetic mechanisms operate would augment the description of ANS modulating action over cardiorespiratory interaction [243].

To specifically address this issue, this paper proposes an extension of traditional *TE* formulation. It complements the estimates of magnitude and directionality of information flow with that of timing between the two coupled processes. Accordingly, *TE* was calculated on several lagged versions of the original series. To provide validation for the proposed methodology, computation of lagged *TE* was performed based on a dynamical system composed of two stochastic subsystems simulating two different types of interaction: a lasting and strong one considering the directionality X→Y, paired with a transient and weak one for Y→X. The

validation procedure performed on simulated data confirmed the ability of lagged *TE* to track the information transfer at different time scales.

The resulting application of lagged *TE* on neonatal cardiorespiratory data showed two distinct interaction profiles as a function of directionality: a fast and quickly decaying information transfer from RESP to RR, and a slower but more stable transfer from RR to RESP. Convergent findings with regard to the directionality from RESP to RR were obtained by Faes et al. in 2014 [243], showing that the fast information flow from breathing to HR is associated with the respiratory sinus arrhythmia. Furthermore, the timing of activation of the information flow profile is comparable to the known latencies of activation for the sympathetic and parasympathetic arms of the nervous system. Specifically, the sympathetic branch intervenes on a slower time scale but its effect on the target system lasts longer whereas the parasympathetic has a punctate, yet rapidly vanishing action. Thus, the reported lagged *TE* dynamics might reflect that information transfer directionalities are driven by different autonomic branches of the ANS [248]. This is relevant in the context of possible approaches for quantification of sympathetic activation. A state of sympathetic hyperactivity has been in fact reported as associated with an increase in cardiovascular morbidity and mortality [249], [250]. Thus, while several heart rate variability parameters can assess parasympathetic activity, consensus about quantification of sympathetic nervous system activity is still pending.

Regarding the differences as a function of GA, we reported a significant decrease of information flow in LPT from RESP to RR in quiet sleep, and for LPT and ET a reduced information flow from RR to RESP in active sleep. Interestingly, there was no difference in the candidates selected for the reconstruction of the past states. In line with our findings, previous work addressing other forms of cardiorespiratory interaction in newborn population highlighted that the direction of coupling between cardiovascular and respiratory systems varies with age over the first six months of life, with a tendency to change from a nearly symmetric bidirectional interaction to primarily unidirectional mode from *RESP* to *RR* [184], [251]. These findings strengthen the assumption that GA-related differences are due to intrinsic differences in the interactions between subsystems, given the reported candidate selection employed for the reconstruction of RR series past vector. The reported GA-related results constitute a plausible explanation for cardiorespiratory differences in the newborn period and might pave the way to a possible explanation for the increased risk of LPT and ET populations. To extend these findings,

we explored the role of breathing rate on the modulation of the information flow. Our partial correlation model confirmed the role of breathing rate as a mediator for the interaction between sleep states and *TE*, but not between GA and *TE*. These results mirror our previous findings regarding cardiorespiratory interactions. We previously reported the absence of a change in breathing frequency as a function of GA. However, a significant modification of directionality of the cardiorespiratory coupling had been observed [176].

One limitation of the presented investigation is the absence of arterial blood pressure (ABP) included in the model. The availability of this additional signal would lead to a more comprehensive investigation of the complex physiological interactions of the cardiovascular systems as a function of state and age [252]. Lastly, larger scale studies are needed to investigate neonatal ANS regulation in the context of diverse factors, e.g., ethnicity, socio-economic status, maternal conditions, psychosocial stressors.

In conclusion, the utilization of a lagged version of *TE* might lead to a novel approach to investigate physiologic networks, selectively assessing horizontal information transfer at different time scales. This particular investigation of the interaction between the cardiac and respiratory systems aimed at characterizing the different regulatory profiles of the two branches of the ANS and at ultimately providing an indication of altered patterns of physiological behavior. Findings presented in this paper are convergent with previous published findings [241], [243]. The novel contribution of this study is the characterization of the dynamics of the cardiorespiratory network across sleep states and gestational ages. Ultimately, application of *TE* in assessment of network physiology affords an opportunity for early risk stratification in a high risk population [31]. In the near future, we envision the application of *TE* methodology for the characterization of other interacting subsystems such as brain-brain and brain-heart to provide a more comprehensive picture of the complex mechanisms characterizing neonatal development.

Influence of Prenatal Alcohol and Tobacco exposure on Neonatal Vagal Tone in Response to Head-Up Tilt

Introduction

As extensively reported in this Ph.D. thesis, prenatal exposure has been shown to play a crucial role in shaping the infant's ability to face internal and external stressors and may also contribute to autonomic nervous system (ANS) failure to maintain homeostatic control following physiologically challenging events [253]. In this approach, we focused our attention on the effects of PAE and PTE in relationship with the ANS response to tilt maneuver in a cohort of newborns assessed at birth. Associations between exposure and increased two- to five-fold risk of SIDS has been reported [118]. The underlying hypothesis describes abnormalities in the control of homeostatic functions by the developing brainstem [22]. Exposure to nicotine and alkaloid in cigarettes may contribute to Harper et al.'s proposed model, which states that exposed infants are less able to arouse and respond to potential life-threatening physiological challenges [254].

Graded head-up tilt has been widely used as a test to observe and quantify autonomic control. The physiological response to tilt encompasses a shift of the sympathovagal balance toward a sympathetic activation and parasympathetic withdrawal [255]. Several indexes have been proposed to quantify sympathetic and parasympathetic activity. However, there is no current consensus for standardized measures of vagal activation/withdrawal in infants [256]. In addition, newborns spend most of their time asleep, either in active or quiet sleep state while on the other hand, tilt is performed in wake state in adults. Time domain quantification of vagal response to head up tilt in unexposed infants has been reported by Myers at al. [257]. A measure of beat-to-beat variability (RMSSD) showed significant decreases following head-up tilts, regardless of sleep state. Another method employed to quantify vagal activity is the computation of High Frequency (HF) spectral component of Heart Rate Variability (HRV) Power Spectral Density (PSD) [255], [68]. An indirect but quantifiable effect of respiration on beat-to-beat vagal control on HR is termed Respiratory Sinus Arrhythmia (RSA) [185]. RSA is determined by both tonic and phasic processes, with the tonic branch more related to vagal control. Medullary respiration control centers and pulmonary stretch receptors also contribute to RSA. Thus, neural and mechanical contributions are physiologically coupled and may not be easily dissociated. Furthermore, the

relative contributions to RSA may vary across behavioral states, with a prevalence of ANS control during rest/sleep state [185].

This investigation entails a novel application of parameters of autonomic function control to characterize the immature neonatal ANS, whose cardiorespiratory interactions are not yet stable. Multiple measures of vagal withdrawal were investigated: traditional time/frequency domain estimates along with novel complexity measures. The analysis was performed on two cohort of infants, unexposed newborns and those exposed in utero to both alcohol and smoking. The combined use of these indexes suggested different regulatory dynamics in the two groups and opens to novel perspective in physiological and clinical risk assessment.

Materials and methods

The participants analyzed were a subset of the maternal/infant dyads enrolled in the Safe Passage study. They were recruited at Tygerberg Hospital and affiliated clinics in Bellville, Cape Town, South Africa. Infants were routinely discharged from the hospital less than 24 hours after delivery, and they subsequently returned within 48-96 hours for assessment. At this visit, nurses recorded electrocardiogram (ECG), respiration, blood pressure (BP), and cortical brain activity (EEG) during sleep. Written informed consent to record these physiology signals was obtained as part of the consent for the main study. Ethical approval was obtained from Stellenbosch University, Sanford Health, the Indian Health Service, and New York State Psychiatric Institute. Data recordings were approximately 1-hour long and occurred approximately 30 minutes after eating. The standardized procedure consisted of a 10-minutes baseline period and three rapid (~3–5 seconds) 45° head-up tilts, while the infant was in the prone position [258] as it may accentuate the response evoked by tilting. Each tilt session is organized in: 30-sec just prior tilt (B); 15-seconds block after the infant reached head-up position (I); three per 30-seconds (T1, T2, T3) before returning to flat position and repeated three time during 1-hour recording. The data analyzed in this work include the first tilt session out of the three performed and the block labelled as I was not considered due to its lack of stationarity.

ECG was recorded at 500 Hz and respiratory tracings were simultaneously collected using a respiratory inductance belt, which was placed around each infant's chest and acquired at of 20 Hz. Participants were divided in: i) control group (CG): no prenatal exposure to alcohol or tobacco smoke; ii) exposure group (EG): high continuous prenatal exposure to both substances (the

previously defined cluster #1 for alcohol and smoking) throughout pregnancy. Specifically, the mothers in EG had on average more than 25 drinks per trimester and smoked more than 20 cigarettes per week throughout gestation. Inclusion criteria for the cohorts in this study were: full-term infants (gestational age (GA) at birth ≥37 weeks), birthweight ≥2500 g, South African mixed ancestry, no resuscitation at birth or admission to the Neonatal Intensive Care Unit (NICU), no prenatal exposures to drugs of abuse, no evidence of diabetes, hypertension, or pre-eclampsia during pregnancy in maternal medical records.

The total number of infants in the CG is 15 (3 M – 12 F), Gestational Age (GA) is 39.7 ± 1.07 weeks while in the EG the total number is 13 (6 M – 7 F), GA is 39.2 ± 0.83 weeks. Infants sleep states were classified into Active Sleep (AS) and Quiet Sleep (QS), or Awake (A) by an automate algorithm [211]. In this study, the reported analysis was performed exclusively on periods classified as QS.

The ECG signal was filtered to remove electrical line noise using a 50 Hz notch filter. The RR peaks were detected on the ECG with the Pan-Tompkins algorithm. An adaptive filter was then applied to remove ectopic beats or artifacts in the extracted RR series. RR intervals with a duration <0.300 s (200 bpm) or >0.667 s (90 bpm) as well as adjacent intervals with a relative increase >15% were discarded from further analyses. The respiration signal was band-pass filtered in the range 0.05-2 Hz. The Inter Breath Intervals (IBI) series, extracted based on respiration onsets, was corrected with an approach similar to RR series. Breath-to-breath intervals <0.5 seconds (120 breaths/minute) or >2.0 seconds (30 breaths/minute) were excluded.

To investigate the response elicited by tilt maneuver we investigated the performance of a set of estimators quantifying vagal activation and/or withdrawal. Given the lack of consensus for which measures are most suitable to interrogate the response elicited by tilt, four different estimators along with their performances have been investigated in this study:

- RMSSD [s] (time domain): the square root of the mean of the sum of the squares of differences of RR series, a time-domain measure of short-term components of HRV [68].
- HF_power [s^2] (frequency domain): the RR series was resampled at 5 Hz and the associated Power Spectral Density (PSD) was obtained using a non-parametric approach. The spectrum was computed using 20-seconds overlapping Hamming windows. The HF frequency of HRV spectrum was defined as the frequency associated to the power in the

window [mean breathing frequency -0.1 Hz; mean breathing frequency +0.1 Hz] where mean breathing frequency is the inverse of mean IBI [68].

- Quadratic Sample Entropy (QSE) [bits] (complexity domain): this measure overcomes the limitations related to the univocal definition of the tolerance parameter r when computing Sample Entropy (SampEn) [259]. QSE is defined in Eq. 24 as:

$$QSE = SampEn(m, r) + log(2r) \tag{24}$$

where r is optimally varied until a priori defined number of matches are observed given the embedding dimension m. In this study, m is set to 1 and initial value of r is set to 0.1*standard deviation of analyzed RR series.

- RSA_amplitude [s] (phase domain): is computed based on the method reported by Bartsch in [198]. In this work, the phase of the respiratory signal ϕ_r is computed using the DAMOCO toolbox [15]. The protophase of respiration signal is obtain using Hilbert transform and further corrected to obtain ϕ_r. The series of relative heartbeat distances obtained considering windows of two consecutive breaths $\phi_r \in [0, 4\pi]$, is fitted with a least-square sinusoidal approach. The RSA_amplitude is defined as the amplitude of the derived sinusoid.

The extraction of the four parameters provide a set of variables for every participant in each of the four tilt blocks, e.g. RMSSD_B, RMSSD_T1, RMSSD_T2, and RMSSD_T3. The distribution of each parameter was tested for normality by means of Shapiro-Wilk test and values of skewness and kurtosis were computed. The variables RMSSD, HF_power and RSA_amplitude failed the above described normality tests thus, they were log transformed (ln_RMSSD, ln_HF_power, and ln_RSA). The statistical analysis was performed using repeated measures ANOVA. This test aims at evaluating the within-subject differences comparing the four time points corresponding to tilt blocks in CG and EG separately. As a consideration for further analysis, the overall extracted parameter distribution (regardless tilt blocks) can be considered approximately normally distribute, thus the requirement of normality for repeated measures ANOVA stands for both CG and EG.

Results

A summary of the trend for the four parameters, F and p-value of Greenhouse-Geisser tests are reported in Table 17. Each parameter was analyzed having tilt block as within-subjects factor. Reults highlighted a significant within-subjects effect of tilt block on all extracted parameters in the CG. Left panel of Figure 29 shows the mean trend of the extracted parameters for CG.

Table 17. Mean ± std of parameter for each tilt block and within-subjects results of repeated measures anova for cg and eg. across-subject differences are found in CG only as a consistent result for each of the tested parameters.

Control Group (N=15)				**Exposed Group (N=13)**			
B	T1	T2	T3	B	T1	T2	T3
ln_RMSSD							
-4.27 ± 0.69	-4.50 ± 0.62	-4.59 ± 0.68	-4.58 ± 0.67	-4.17 ± 0.53	-4.19 ± 0.60	-4.18 ± 0.49	-4.20 ± 0.57
p-value=0.001		F(1.83, 25.55)=9.29		*p-value*=n.s.		F(2.11, 25.30)=0.05	
ln_HF							
-10.46 ± 1.45	-10.76 ± 1.27	-11.04 ± 1.48	-10.92 ± 1.41	-10.26 ± 1.03	-10.26 ± 1.16	-10.20 ± 0.99	-10.19 ± 1.15
p-value<0.05		F(2.40, 33.52)=3.22		*p-value*=n.s.		F(2.10, 25.19)=0.07	
QSE							
3.72 ± 0.74	3.57 ± 0.58	3.44 ± 0.66	3.46 ± 0.73	3.89 ± 0.53	3.80 ± 0.58	3.74 ± 0.50	3.78 ± 0.52
p-value<0.001		F(2.58, 36.09)=7.66		*p-value*=n.s.		F(1.86, 22.37)=0.71	
ln_RSA							
-3.93 ± 0.77	-4.28 ± 0.74	-4.35 ± 0.84	-4.29 ± 0.77	-3.99 ± 0.61	-4.06 ± 0.72	-3.95 ± 0.61	-3.90 ± 0.67
p-value<0.05		F(1.69, 23.66)=4.24		*p-value*=n.s.		F(1.80, 21.53)=1.02	

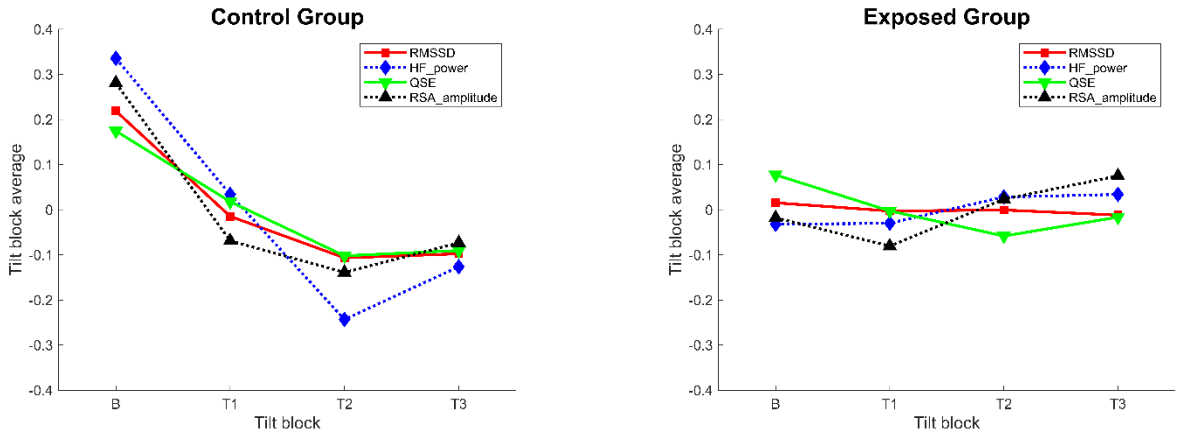


Figure 29. Left and right panel show the mean values of the four extracted parameters and their trends when plotted based upon the tilt blocks. CG show a consistent behavior across all indexes, characterized by a decrease from B to T3. Group means in EG do exhibit a flatter tendency. The mean value for a given index (obtain by grouping means in B, T1, T2, T3) has been subtracted from mean values reported in Table I to provide a comparable scale across parameters.

The computed parameters show a consistent decreasing in vagal magnitude in the comparison of BvsT1 and BvsT2. On the other hand, the vagal activation pattern in segment T2 and T3 exhibit a flat trend. To investigate the statistical significance of the reported differences between pairs of tilt blocks, a series of post-hoc tests were performed. The pairwise comparisons

did not show any difference between BvsT1, BvsT3, T1vsT2, T1vsT3 or T2vsT3. However, a significant difference was found when comparing BvsT2, RMSSD p-value=0.01, ln_HF p-value=not significant (n.s.), QSE p-value<0.01, ln_RSA p-value<0.05. The comparison between CG and EG did not show any main effect associated with groups when performing repeated measures ANOVA with group as a factor.

The right panel of Figure 29 depicts the group trends for EG which is profoundly different compared to the behavior reported for CG. Vagal tone estimators do exhibit a stable trajectory throughout the tilt session, a slight but not significant decrease is shown by RSA in the comparison BvsT1. The absence of any difference is confirmed by the results of repeated measures ANOVA. Differences in baseline values between the groups are not significant due to large individual variability and small number of subjects.

Discussion

In this work we investigated different measures of vagal tone aimed to quantify the ANS response elicited by head up tilt in two population of unexposed and heavily exposure to prenatal maternal alcohol and smoking consumption. Results for the former group evidenced a physiological vagal withdrawal in response to head-up tilt. This behavior was consistently found for all four measures. In contrast, the absence of a vagal withdrawal was demonstrated for the prenatally unexposed infants. The results for the traditional time and frequency parameters are consistent with findings by Myers et al. [257], which showed a significant decrease in beat-to-beat variability as consequence of tilt and with previous studies [255] showing a vagal withdrawal in a population of adults in response to tilt. Complexity analysis performed in this study shows, for unexposed infants, a significant within-subject decrease of QSE. This trend is similar to previous findings reporting Approximate Entropy (ApEn) decreasing as a response to vagal blockade [260]. The decrease in complexity on such short-term scale (m=1) may indicate a shift towards increased regularity of RR series due to vagal withdrawal and increase prevalence of sympathetic system. Lastly, the quantification of vagal tone by means of RSA is still largely debated. Polyvagal theory by Porges [261] states that successful vagal regulation is marked by RSA suppression or withdrawal, which is thought to facilitate an organism's ability to cope with challenging state. On the other hand, high correlation between RSA measures and breathing frequency have been reported, suggesting a substantial equivalence of the two quantities. In this study, absence of

correlation between the above-mentioned parameters has been reported for each tilt block. Instead, significant correlation between RSA and HF power is consistent. Given these findings it is not possible to conclude if the RSA could provide a reliable estimation of vagal tone estimation or it may eventually mimic the behavior of HF power modulated by respiratory frequency. At the same time, RSA trend is consistent with the reported estimators of vagal magnitude.

The results for EG show consistent lower vagal withdrawal in response to tilt challenge for every extracted parameter. In literature, prenatal exposure to tobacco smoking has been reported to effectively impair receptors in the medullary 5-HT system, as well as alterations in cardiorespiratory control mechanisms and abnormalities in the pathogenesis of the parasympathetic systems [118]. It is possible to speculate that illustrated difference in vagal tone comparing CG and EG may effectively arise as a consequence of prenatal exposure and may contribute to the decreased infant's ANS capability to maintain homeostatic control when expose to a direct physiological challenge. The reported findings seem to confirm the altered heart rate response to tilt as described for other groups of infants at high risk for SIDS [262].

The reported absence of group differences in the comparison between CG and EG is possibly related to the limited database analyzed and remarkable variability between subjects. In future analyses we propose to apply this same approach to a larger cohort of infants with the aim of investigating the individual effects of alcohol and smoking as well as the possible dose response effect consequent of prenatal exposures.

Predicting Toddler Neurodevelopmental Phenotype: a Comprehensive and Contextualized Approach

Introduction

It is well-established that the gestational environment is strongly tied to future risk for neurodevelopment sequelae [263]. However, there is limited research exploring the relationships between diverse fetal and maternal prenatal exposures in relationship with early life adversities. Comorbid in-utero exposure to teratogens and adverse psychosocial factors may have overlapping pathways with additive, multiplicative, or possibly even opposing effects on adverse neurodevelopmental outcomes. Therefore, a more comprehensive and quantitative approach it is imperative to understand how these factors interact in shaping early brain-behavior development in a more rigorous and quantitative way. Starting from the results discussed this Ph.D. thesis, we proposed to define a machine learning framework to predict neurobehavioral outcome in the context of perinatal exposure. The innovative combination of heterogenous features is expected to inform on the diverse mechanisms responsible for shaping early life environment, and ultimately be predictive of subsequent development.

This approach is supported by the previously findings showing that maturation of the central and autonomic nervous systems occurs at the most rapid pace during the fetal period, leaving the fetus vulnerable to exposures in the intrauterine environment during a critical developmental period [264]. Adaptations to adverse gestational exposures can potentially program the fetal brain to alter typical developmental trajectories, which can influence subsequent postnatal neurodevelopmental functioning [265], [266]. To our knowledge, there are no longitudinal studies examining the interactive role of multiple prenatal environmental and maternally mediated exposures on early brain and childhood neurobehavioral development. However, there is a wealth of literature highlighting the importance of the prenatal environment in shaping later neurodevelopmental outcomes. Several studies have also examined associations between maternal distress during pregnancy and infant brain-behavioral development [266]. For example, a recent meta-analysis demonstrated in-utero exposure to depression, anxiety, and stress were associated with childhood social-emotional problems such as behavioral dysregulation and difficult temperament [267]. Other prior research has highlighted the importance of social determinants of

health during pregnancy such as economic stability, social support, educational attainment, living conditions, and access to healthcare amongst vulnerable populations in influencing maternal-child outcomes [267].

Several approaches have previously implemented to examine the effects of prenatal exposures on infant brain development including structural magnetic resonance imaging (MRI), functional MRI, or diffusion tensor imaging [18], [268]. Imaging the perinatal brain is challenging due to high costs and significant issues relating to movement artifacts, resulting in studies with small sample sizes. Electroencephalography (EEG) is a practical and non-invasive tool which can be used to assess brain activity in the neonatal period. The extracted EEG power can be utilized as a measure of brain maturation due to age-related changes in power across the frequency spectrum which are postulated to reflect decreases in synaptic density, cortical, and subcortical gray matter that underlies neural pruning to increase functional specialization [269]–[274]. Although there is significant heterogeneity in prior studies, in the context of the current study we sought to move beyond identifying linear brain-behavior associations and implemented a data-driven approach with feature selection and a non-parametric machine learning model to predict several neurodevelopmental outcome groups.

Regression models play a crucial role in generating comprehensible and interpretable prediction and classification rules among diverse modes of investigation, such as brain-behavior relationships. Although appealing in simplicity, traditional linear approaches are often unsuitable to describe complex relationships embedded across multiple levels of analysis. Input-output relationships are often nonlinear, thus more sophisticated methods are required to overcome the limitations of a simple regression modeling [33], [109]. Generalized additive models are flexible statistical methods in which a given target is predicted by an unknown nonlinear smooth function capable of aggregating several predictor variables. Among generalized additive models, decision tree-based methods afford powerful analytic approaches suitable for both classification and regression. Decision tree-based methods partition the feature space into a set of rectangles and a simple model is then subsequently fit in each one [275]. The described stratification technique has been reported to closely mirror human decision making, thus a major advantage of tree-based methods is the low complexity and high interpretability of the classification results. At the same time, trees generally do not exhibit the same level of accuracy if compared to more advanced machine learning techniques. As a consequence, novel methodologies capable of combining

several decision trees have been developed to substantially improve the performance of generalized additive models [276]. The described family of models is particularly suitable in the context of psychological research where a large number of variables can be measure on a limited set of participants.

In the current report we applied machine learning methodology to predict toddler neurodevelopmental phenotype from prenatal environmental factors, maternally mediated exposures, and neonatal EEG spectral power at birth. By implementing this rigorous approach to predict neurodevelopmental risk based solely on prenatal factors and neonatal EEG markers, we sought to identify outcome-specific markers, risk factors, and protective factors as future targets for timely intervention strategies to support high-risk mother-infant dyads.

Materials and methods

Participants were a subset of infants enrolled in the Safe Passage Study conducted by the Prenatal Alcohol and SIDS and Stillbirth (PASS) Network. The present study took place in the Western Cape Province of South Africa. The final sample consisted of 224 maternal-infant dyads with information regarding maternal prenatal depression, anxiety, socioeconomic status, prenatal exposure history, and neonatal EEG data (95 females; gestational age at birth: 39.23 weeks ± 0.96. Infants were 12 – 100 hours old at the time of the EEG assessment (61.07 ± 22.55 hours of life) and had no prenatal exposure to psychiatric medications. Participants were all born between 37 weeks and 41 weeks gestation. Written informed consent to record infant brain activity using electroencephalography (EEG) was part of the consent for the Safe Passage Study. Separate informed consent was obtained from a parent or legal guardian for each participant for developmental follow-up assessments. Ethical approval was obtained for both time points from the Health Research Ethics Committee of Stellenbosch University and the New York State Psychiatric Institute.

Maternal-infant charts were abstracted to obtain maternal age at delivery, gestational age at birth, delivery mode, biological sex, and any maternal medical conditions during pregnancy. The infant's age at the EEG study was determined by their gestational age at birth and postnatal age (in hours) at the time of the EEG study. All participants in the present analysis (n=224) had complete data for all covariates. Information regarding maternal mental health during pregnancy were obtained at 20 – 24 weeks' gestation. Depressive symptoms were measured using the

Edinburgh Postnatal Depression Scale (EPDS) [277], [278], a maternal depression screening tool which has been validated in English and Afrikaans [133]. Maternal anxiety symptoms were measured using the State-Trait Anxiety Inventory (STAI) [279], an anxiety screening tool to distinguish anxiety symptoms from depressive symptoms which has also been validated in both languages. Through study specific case reports forms, participants indicated demographic and socioeconomic information including race, maternal educational attainment, household crowding (persons per room in household), prenatal care during pregnancy, and marital status. PAE and PTE were extracted in accordance with the methodologies described in Chapter 2.

EEG data were acquired during the newborns' natural sleep using a hybrid system of a 28-lead high-impedance electrode net (Electrical Geodesics) and a miniature amplifier recording device (ATES). Prior to screening data for artifact for each second using the thresholds described below, raw data for the entire minute were filtered for line noise and any ECG artifact was subtracted from each channel. A 16,000-point finite-impulse response 4 Hz wide notch filter was applied at the line noise frequency and its first harmonic (60 and 120 Hz, or 50 and 100 Hz, depending on the site), with 36 to 100 dB power reduction within the notches. ECG artifact was removed from each channel using a recently developed method that mimics the procedure for ballistocardiogram removal from EEG recorded during MRI [272]. Ballistocardiogram artifact is removed by using a simultaneously recorded ECG signal to precisely identify the times of each R wave peak, and then to signal average the EEG over small windows centered on those times, deriving a channel-specific template that is then subtracted from each channel. EEG power spectra were then computed for 60-second epochs using the Welch method, taken each second. Data were demeaned and a Hanning window was applied prior to computing the FFT for each second. To determine the leads and times contaminated by movement-related or other sources of electrical artifact, we applied multiple criteria on a second by second basis to data from each lead. Criteria were as follows: standard deviation of voltage less than 50 μV and greater than 0.001 μV; sample-to-sample change less than 50 μV; absolute value of voltage less than 300 μV; log-log spectral slope of raw data between 20 and 120 Hz less than -0.1 (to screen for muscle artifact). If more than 5 leads had artifact during any one second, that second was excluded. Remaining data were re-referenced to the average over all leads at each sample. Finally, minute by minute power was the average of the squared FFT's over the accepted seconds, requiring at least 30 acceptable seconds per minute for each lead as a minimum inclusion threshold. Minute by minute EEG power was

aligned with simultaneous sleep state codes and averaged over AS and QS minutes within each study. Average absolute EEG power was calculated for 12 scalp regions (left frontal-polar, right frontal-polar, left frontal, right frontal, left central, right central, left parietal, right parietal, left temporal, right temporal, left occipital, and right occipital) for 3 Hz wide bins from 1 Hz – 48 Hz. Due to a restricted sample size in QS as well as previous results, subsequent analyses focused on data in AS only.

The following developmental assessments were administered between 31 and 42 months of age (38.2 ± 3.0 months):

- **Bayley Scales of Infant Development III Screening Test (BSID-III)** is a rapid assessment of cognitive, language, and motor functioning in infants and young children in order to determine if a child's development is within normal limits and identify risk for developmental delay [280].
- **Brief Infant Toddler Social Emotional Assessment (BITSEA)** is a 42-item parental report measure of social-emotional development, behavioral problems, and delays in competence58. The BITSEA also had good discriminative power to differentiate children with and without autism spectrum disorder (ASD) using item level questions that directly relate to autism specific behaviors [281], [282].
- **Modified Checklist for Autism in Toddlers Revised with Follow up (M-CHAT-R/F)** is a, 20-item guardian/parental report checklist validated to screen children for early signs of ASD or developmental delay [283]–[285].

To examine the predictive power of several prenatal environmental factors and quantitative brain activity at birth in discriminating toddler neurodevelopmental phenotypes, firstly we implemented an unsupervised clustering methodology namely hierarchical analysis to derive subgroups based on shared cognitive or other neurodevelopmental domains [286] to disambiguate neurodevelopmental profiles. We used the following developmental features as inputs to the cluster analysis: social-emotional problems derived from the BITSEA, social-emotional competence derived from the BITSEA, autism risk composite score derived from averaging the (standardized) MCHAT and BITSEA autism risk scores, and cognitive ability derived from the BSID-III. Outliers 2.5 standard deviations from the mean were winsorized prior to running the cluster analysis. All features were z-scored after adjusting for chronological age at assessment via

linear regression to extract residuals. We performed simultaneous evaluation of several clustering methods to aid in the determination of the most appropriate method and number of profiles. Several statistical parameters were evaluated to ensure maximum intracluster distance and minimum intercluster variability. The internal validation is comprised of three parameters: connectivity, silhouette width, and the Dunn index [287]. Connectivity is a measure of intracluster distance and intercluster homogeneity whereas the silhouette width and Dunn index are both non-linear measures of cluster compactness and separation. Based on our evaluation of the above-mentioned parameters, we implemented hierarchical clustering with average linkage as metric for intracluster distance. The optimal number of clusters was found to be equal four. The procedure of extracting a more comprehensive profile of neurodevelopmental outcome is expected to provide an enhanced description of potential abnormal trajectories of development.

Next, we sought to combine the prenatal variables described in this section and neonatal EEG features to predict subsequent neurodevelopmental phenotype in toddlerhood. Prior to the construction of predictive models, recursive feature elimination [288] was implemented to reduce the dimensionality of the EEG power features to be used in subsequent analyses, remove the least predictive features, and derive the optimal number of features by testing several subsets of attributes. Specifically, RFE consisted of a 10-fold cross-validation, repeated 3 times restricting the numerosity of EEG features to a maximum of 20 in order to achieve clinical interpretability and keep the complexity limited. To predict the multivariate neurodevelopmental cluster groups from EEG and prenatal features we employed recursive partitioning analysis [288]. The rationale behind the proposed methodology was aimed to overcome the limitations of classical statistical methods such as linear and logistic regression. Specifically, a scenario as the one presented in this investigation may be described as the so called “small n large p” case, where the number of predictor variables p is greater than the number of subjects n. In this context, traditional parametric approaches are no longer applicable as they are unable to successfully deal with collinearity among variables and feature space of high dimensionality [288]. The performance evaluation of recursive partitioning analysis consisted of 10-fold cross-validated repeated three times.

Results

Four neurodevelopmental profiles were derived from the previously described cluster analysis: Autism-Risk (n=41), Cognitive-Risk (n=81), Social-Emotional Risk (n=48), and Low-Risk (n=54). A graphical representation of the feature distributions in these clusters is shown in Figure 30.

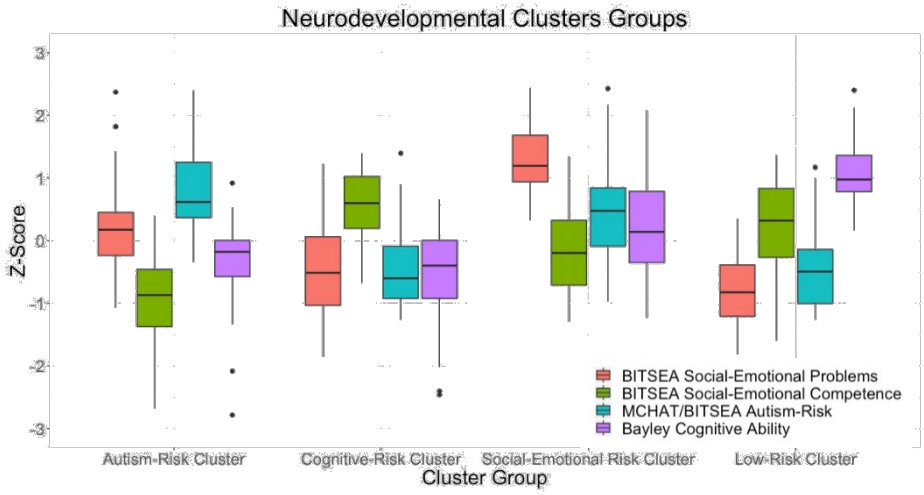


Figure 30. Neurodevelopmental Cluster Group.

These groups are statistically significantly differentiated by social-emotional problems ($F(3,223) = 107.80$, $p < 0.001$), social-emotional competence ($F(3,223) = 45.28$, $p < 0.001$), autism-risk scores derived from the BITSEA ($F(3,223) = 37.17$, $p < 0.001$), autism-risk scores derived from the MCHAT ($F(3,223) = 2.79$, $p < 0.050$), and by cognitive ability ($F(3,223) = 53.53$, $p < 0.001$). There was no significant main effect of cluster group on gestational age at birth, gestational age at follow-up, sex, delivery type, prenatal care, parity, maternal age at delivery, maternal marital status, maternal education, maternal prenatal anxiety, maternal prenatal alcohol use, or maternal prenatal tobacco use. Household crowding during pregnancy, maternal prenatal depression, maternal prenatal recreational drug use, and maternal gestational hypertension were significantly different across neurodevelopmental cluster groups.

Among the initial 192 EEG features we identified 17 informative EEG features to be utilized in for the subsequent modeling. These consisted of a wide range of EEG frequencies in the right-frontal, left-temporal, left-central, and left occipital electrode regions. The retained quantitative EEG power features are reported in Table 18.

Table 18. Results of the feature selection procedure applied to the neonatal EEG measures.

EEG Features	
Region	Frequency Bins (Hz)
Right Frontal	37 – 39, 40 – 42
Left Temporal	1 – 3, 7 – 9, 31 – 33
Left Central	1 – 3
Right Central	31 – 33, 34 – 36, 43 – 45
Left Occipital	4 – 6, 13 – 15, 22 – 24, 28 – 30, 31 – 33, 37 – 39, 40 – 42, 49 – 51

At this stage, we trained the recursive partitioning model with the reduced set of EEG features, maternal and neonatal characteristics. The results of the feature importance ranking are reported in Figure 31. It is noticeable that the large majority of EEG features were associated with an enhanced predictive power if compared to demographic and maternal-related variables such as depression, prenatal maternal state anxiety, prenatal maternal trait anxiety, maternal educational status at the time of delivery, prenatal care visits, infant's sex, prenatal substance use, and maternal gestational hypertension. The validation model had 81.25% accuracy (95% CI, 75.51%, 86.14%)

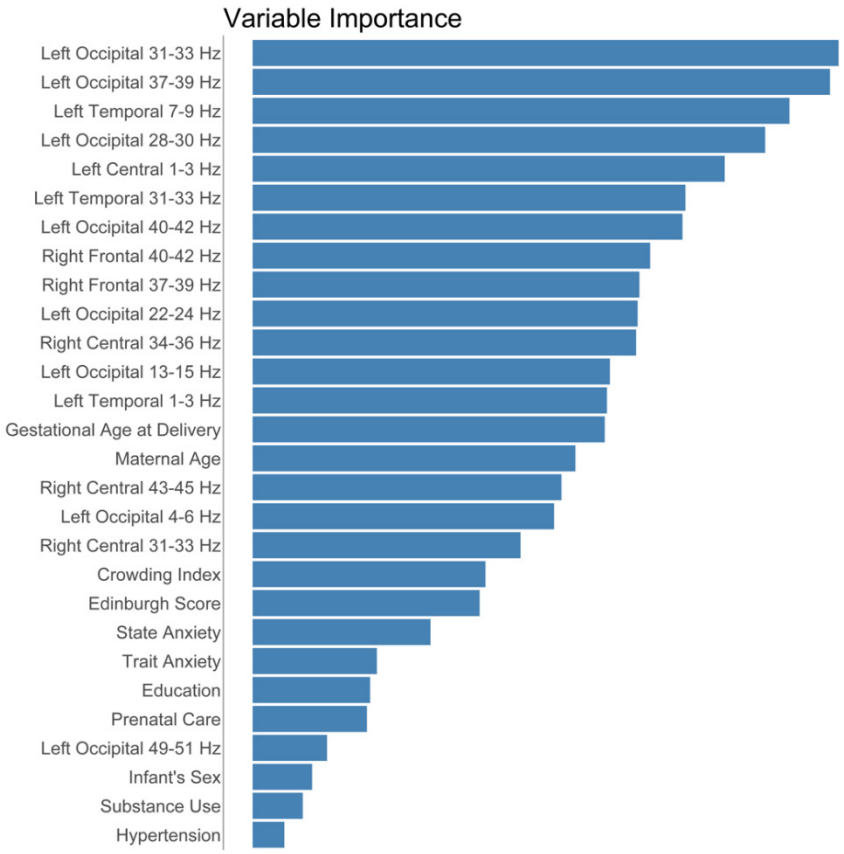


Figure 31. Variable Importance in Predicting Neurodevelopmental Outcomes

in predicting the multiclass neurodevelopmental phenotype cluster groups. Balanced accuracy for each neurodevelopmental cluster group was: autism risk: 86.84%, cognitive-risk 85.58%, social-emotional risk 82.86%, and low-risk 93.69%. Additional measures of performance are reported in Table 19.

Table 19. Prediction model performance.

	Neurodevelopmental Cluster Groups			
	Autism-Risk	**Cognitive-Risk**	**Social-Emotional Risk**	**Low-Risk**
Sensitivity	78.05 %	80.25 %	70.83 %	94.44 %
Specificity	95.63 %	90.91 %	94.89 %	92.94 %
Positive Predictive Value	80.00 %	83.33 %	79.07 %	80.95 %
Negative Predictive Value	95.11 %	89.04 %	92.27 %	98.14 %
Prevalence	18.30 %	36.16 %	21.43 %	24.11 %
Detection Rate	14.29 %	29.02 %	15.18 %	22.77 %
Detection Prevalence	17.86 %	34.82 %	19.20 %	28.12 %
Balanced Accuracy	86.84 %	85.58 %	82.86 %	93.69 %

Discussion

The present report has several important implications for research examining prenatal factors and potential physiological markers of neurodevelopment. Firstly, we proposed a novel data-driven approach to overcome the limitation of traditional linear methodologies. The rationale was the examine the associations between prenatal exposures and neurodevelopment in toddlerhood by a more sophisticated and comprehensive framework, able to combine EEG spectral power during the active sleep state at birth in the context of perinatal factors. By combining heterogenous data source on prenatal environment with neonatal EEG spectral power in a recursive partitioning analysis model, we predicted multiclass neurodevelopmental outcomes in toddlerhood with 81.25% accuracy. Sensitivity and specificity for each neurodevelopmental outcome class (autism-risk, cognitive-risk, social-emotional risk, and low-risk) exceeded 70% with balanced accuracy exceeding 82%. The model performed the strongest in predicting children in the low-risk cluster (balanced accuracy 93%) and weakest in predicting children in the social-emotional risk cluster (balanced accuracy 82%). Few studies have attempted to identify markers for neurodevelopmental-risk in community samples or in infants at-risk due to pregnancy related exposures. Several studies have demonstrated associations between EEG power during infancy

and subsequent autism risk [289]. Infants at high familial risk for ASD have been show to demonstrate reduced frontal power at 3-months of age, no hemispheric asymmetry in frontal alpha activity at 6-months with differential alpha asymmetry trajectories from 6- to 12-months of age, and decreased power across all frequency bands at 6-months of age with different trajectories of change in power from 6- to 24-months of age [290], and frontal and central hyperconnectivity at 14-months of age [291]. In a recent study sample of low-risk controls, high-risk infants without an ASD diagnosis, and high-risk infants with an ASD diagnosis, the estimated summed delta, low alpha, high alpha, and beta EEG power at 6-months of age significantly predicted ASD diagnosis [292]. However, patterns of brain activity may be fundamentally different in children with familial autism-risk than in children with an unknown etiology or other risk factors who later receive an autism diagnosis. We previously demonstrated decreased beta and gamma power in the frontal-polar, temporal, and parietal regions in newborns was associated with increased autism risk in toddlerhood [293]. However, to our knowledge no other published studies have investigated broad neurodevelopmental-risk in a community sample. Interestingly, PAE and PTE features were listed among the least predictive variables. It is possible to speculate that the EEG derived parameters implicitly account for the modifications induced by exposure on neonatal CNS functioning, thus they were able to explain a more consistent portion of the variability among included participants.

Additionally, through characterization of neurodevelopmental risk using cluster analysis, we demonstrated that prenatal maternal depression, prenatal maternal anxiety, household crowding during pregnancy, and prenatal maternal recreational drug use were significantly associated with neurodevelopmental outcomes in toddlerhood. Specifically, prenatal maternal depression was associated with increased neurodevelopmental vulnerability across all three domains, autism-, cognitive-, and social-emotional risk. Prenatal state anxiety was associated with increased social-emotional risk only whereas prenatal trait anxiety was associated with both increased social-emotional and autism-risk. Household crowding was associated with increased autism-risk and cognitive-risk whereas recreational drug use during pregnancy was associated with increased social-emotional risk only.

Our findings are partially consistent with a recent meta-analysis, which demonstrated associations between prenatal depression and prenatal anxiety and social-emotional problems in offspring with odds ratios of 1.79 and 1.50 respectively [294]. This paper also found stronger effects when socio-demographic risk was highest. A World Health Organization (WHO)

systematic review examined the effects of household crowding on health. In this report, the majority of studies demonstrated an association between crowding, a marker of poverty, and adverse mental health outcomes in adults. The opportunity to derive an intelligible link between selected features and physiological correlates might be conceptualized as an opportunity to exclude biased in variable selection.

Although we demonstrated several associations between prenatal factors, EEG features, and neurodevelopmental outcomes, there are also several limitations we must address. The utilization of recursive partitioning methodologies aimed to exploit a family of non-parametric technique able to deal with a relative imbalanced participant to variables ratio. Additionally, random forest and tree-based framework have been successfully applied to model complex and heterogeneous interactions. Two major limitations for the proposed machine learning framework are the requirement for a complete data set, which can significantly restrict sample size when data cannot be imputed, and overfitting. To address the former limitation, imputation methods ought to be considered to deal with data missingness. At the same time, the potential validity of the derived results is not independent of sample size, as previously described in this chapter. The issue of limited samples size also plays a crucial role towards overfitting and machine learning performance. To this end, we proposed a 10-fold cross-validation schema repeated three times. Nevertheless, the assessment of classification performance on a totally independent database would support the derived framework and strengthen the validity of the proposed approach. Additionally, in the current study we only examined the effects of prenatal exposures and predictive power of EEG in term-age infants, which may vary for those born preterm. Moreover, we relied on a combination of screening level assessments of social-emotional development, cognitive ability, and autism-risk to define neurodevelopmental phenotype. We did not have clinical outcome data, such as ASD diagnosis. In future reports, we plan to examine the predictive potential of more sophisticated EEG-derived measures such as coherence, brain activity in the quiet sleep state, and additional long-term neurodevelopmental or clinical outcomes. Lastly, since this is the first study to examine several interactive prenatal factors and physiologic markers, replication is imperative.

Conclusions

The goal of the illustrated Ph.D. thesis was to design a novel methodological framework for a comprehensive and rigorous characterization of the diverse phases of the perinatal period as well as the interrelationship between pregnant women, fetuses, and newborns. To pursue this vision, we tested the applicability of an ensemble of computational approaches capable of highlighting the emergence of potential alterations in diverse crucial pathways throughout the perinatal period. Specifically, machine learning and data analytics tools were trained employing features extracted from advanced signal processing techniques. Prediction of pathological states in the fetal period, data imputation, non-parametric clustering of maternal substance exposure, network physiology analysis, and risk assessment at birth are the main examples of the utilization of the proposed methodologies. Furthermore, an additional innovation of this Ph.D. thesis lies in the analytic framework which proposed the integration of heterogenous data sources. Physiology based features derived from quantitative data such as fetal, neonatal, and maternal biological signals were contextualized with qualitative data such us maternal lifestyle, metrics quantifying exposure to substances, and psychosocial factors. The described approach aimed to go beyond the traditional and widely employed univariate methodologies which tend not to conceptualize the perinatal period as a continuum. To this end, we proposed a novel and rigorous contextualization of this fundamental life period based on multisource information fusion towards advanced monitoring. The envisioned application of the designed framework is to enhance the current state of the art and to be ultimately incorporated as a tool to support and complement clinical practice.

In Chapter 1, we provided evidence for the successful application of multivariate techniques for the detection of intrauterine growth restriction (IUGR). We proposed to develop a set of models for the in-utero detection of the pathology in contrast with the current at birth assessment. Specifically, as a proof of concept of an impactful and clinically relevant application of artificial intelligence in the field of fetal monitoring, we compared the validity and performance of several machine learning techniques for the classification of healthy fetuses versus fetuses affected by abnormal growth profiles. Starting from the preexisting knowledge coming from established physiology based heart rate features, we investigated different machine learning

models for the prediction of both early and late phenotypes. The rationale for the utilization of the illustrated physiology based heart rate features starts from the evidence of their previous application as standalone indices for the discrimination of healthy versus IUGR fetuses [44], [46], [84]. However, throughout the years it has become clear that a single index cannot be descriptive of all pathophysiological processes taking place in the pregnancy period thus the need for multivariate analysis of FHR emerged as evident.

We found Random Forests and Support Vector Machines to best perform in the context of early and late IUGR identification respectively. The need of machine learning model adaptation as a function of gestational age of the developing fetus reflects the changes in fetal autonomic regulation throughout pregnancy itself. Additionally, we reported a considerable variability among subjects. To address this issue, we proposed an approach to remove the influence of gestational age on the physiology based heart rate features which were then fed into the design models. The use of robust linear regression to derive residuals employed in machine learning frameworks allowed to contextualize the differences between healthy fetuses and IUGR independently of age of assessment, thus reflecting a true difference in autonomic nervous system regulatory profiles. This solution becomes crucial for the translation of such model in clinical practice as gestational age at examination may vary substantially depending on several non-modifiable factors. The contribution achieved in this Ph.D. thesis aimed at promoting the application of machine learning methodologies in the context of fetal and perinatal medicine, following the growing trend of the artificial intelligence application in medicine [33], [109]. Findings reported in this investigation support the importance of multivariate approaches to investigate the variety of implications resulting from a pathological condition such as IUGR.

In Chapter 2 we introduced an ensemble of methodologies for the precise and robust quantification of maternal substance consumption during pregnancy. The accurate estimation of maternal substance exposure is expected to contextualize the potential fetal and neonatal increased risk for adverse outcome. From a methodological perspective, the integration of poorly estimated exposure features is expected to underestimate and undercount their contribution to the quantification of potential states of risk in the perinatal period. To effectively contextualize fetal and neonatal wellbeing with information on maternal habits in the perinatal period, we proposed the utilization of advanced data-driven computational methodologies. The rationale behind the

solutions discussed in Chapter 2 is consequent of the lack of rigorous approaches for the quantification of longitudinal exposure to alcohol and tobacco.

The novel adaption of the KNN algorithm implemented in this Ph.D. thesis exhibited satisfactory performance for the imputation of exposure. The nature of the derived approach originates from neighbor-based methods where the imputed value is either measured for a single neighbor or the average for multiple neighbors. An additional advantage of the illustrated methodology is the absence of constrain to fit the available data within a given parametric distribution, thus resulting in decreased variance of estimation. Following the process of imputation, we utilized methodologies of unsupervised clustering to improve interpretability of longitudinal exposure data. Furthermore, from a methodological perspective, clustering can be interpreted as methodology for data dimensionality reduction, thus reducing computational burden of subsequent machine learning and artificial intelligence approaches. In this investigation we provided evidence for the successful application of finite mixture models to extracted clusters reflect both magnitude and timing of exposure simultaneously, thus providing a more comprehensive picture of the multiple patterns of alcohol consumption in large heterogeneous populations. Arguably, the most controversial methodological problem associated with mixture distributions is that of identifying the optimal number of components to summarize the distribution of the underlying set of data. To overcome the shortcomings of the EM-based approaches, Bayesian approaches have been suggested and consistently integrated in the proposed approach. The utilization of the innovative characterization of alcohol exposure is expected to provide a better understanding of the effect of alcohol consumption during pregnancy and as well as the possibility for obtaining deeper insight regarding the associations on between drinking consumption, perinatal physiology, and ultimately adverse outcomes.

In Chapter 3 we discussed the condition of late prematurity. The effects of exposure in the context of abnormalities in fetal growth and development may play the role of factors able to elicit modification in the fetal homeostasis, thus contributing to earlier in gestation delivery. To contextualize the increased risk of mortality and morbidities in this population, we utilized methodologies inspired to network physiology and multimodal signal processing. Specifically, the differences among late, early, and full term were unveiled by the utilization of methodologically advanced signal processing techniques capable of modelling the cardiorespiratory interaction in

this population of newborns. Transfer Entropy revealed a pronounced immaturity of cardiorespiratory interaction in both directionalities of interaction in the late preterm group. The quantitative approach proposed in this Ph.D. thesis demonstrated the validity of a quantitative analysis of cardiorespiratory interactions as a potential tool for risk assessment in the perinatal period.

An additional approach for early life risk stratification was introduced in Chapter 3. In a nutshell, we provided evidence for the use of the previously derived exposure clusters in relationship with fetal and neonatal physiology for the prediction of potential adverse outcome in early childhood. The combination of heterogenous data sources weighted and updated based on longitudinal fetal and neonatal physiology, maternal conditions, and diverse environmental conditions is the envisioned approach for a multimodal monitoring framework able to promptly anticipate adverse conditions in the perinatal period. A combination of features collected during pregnancy and at birth such as neonatal EEG, fetal, neonatal, and maternal chart abstractions, home environment, and in-utero substance use was able to successfully predict neurodevelopment outcome data collected at 24-37 years of age. This finding strongly aligns with the developmental-origins hypothesis, supporting the evidence that perinatal experiences and exposure have a profound repercussion on the early stages of life.

The framework presented in this Ph.D. thesis represents a first attempt toward a more comprehensive and cross-sectional conceptualization of the diverse aspects characterizing the perinatal period. We proposed a solid set of computational tools inspired to machine learning, data analytics, and advanced signal processing for a quantitative and rigorous profiling of potential states of increased risk or adverse outcome. The envisioned advancement of the presented research is the application of artificial intelligence for a comprehensive and longitudinal investigation of the perinatal period. Innovative implementations of neural networks offer the potential to start designing novel approaches to patient care which are expected to be equal or ultimately better than the current standard adopted in clinical practice. These methodologies are expected to facilitate the automatic analysis of raw physiological signals. This concept is the summarized in the so-called waveform artificial intelligence. the technological advancements in the field of wearables are expected to facilitate a waveform-based framework for continuous monitoring. At the same

time, the absence of interpretable extracted indices from the collected signals may pose a serious limitation the use of these discussed tools in clinical practice.

In conclusion, this Ph.D. thesis proposes an innovate quantitative framework expected to facilitate the design of a methodological and rigorous sustainable network for perinatal health monitoring trained on heterogenous data. The envisioned application is to promote healthy pregnancy, safe childbirth, and reduce adverse outcome by informing monitoring solutions for risk assessment with novel dynamical indicators of perinatal health.

Ongoing Developments and Future Work

Artificial intelligence methodologies

While in this Ph.D. thesis we were able to derive new insight on the complex mechanisms typical of the perinatal period utilizing machine learning techniques, we only scratched the surface of the potential offered by the rapidly developing field of artificial intelligence. If purely black box approaches posed a substantial obstacle for the application of the aforementioned methodologies, more recent advancement in the direction of model interpretability as well as intelligibility paved the way for their broader application in the field of engineering applied to medicine. Despite the utilization of black box approaches might ultimately results in accurate prediction of the target outcome, little room is left to uncover the physio/pathological mechanisms underlying the systems under observation. At the same time, recent methodological advancement of neural networks not only to achieve enhanced performance if compared to the clinical gold standard but allowed to gain insight on the network-derived features contributing to prediction. We are currently applying convolutional neural networks for heterogenous data towards the early and timely identification of potential conditions of increased risk of pregnant women, fetuses, and newborns in the perinatal period. Specifically, we are interested in a solution able to merge heterogenous data inputs. To achieve this design, we are working on a hybrid architecture able to combine discrete time series, extracted features, and clinical data. One of the main challenges to be address is the depth of the intersection between the different branches of the architecture. The more the differential networks are interconnected, the greater is the degree of the generated information flow.

We are in the preliminary testing stages of the aforementioned approach focus on i) non invasive prediction of the insurgence of gestational diabetes mellitus (GDM) as well as prediction of hypo/hyperglycemia events, ii) risk stratification of maternal cardiovascular health (CVH) during pregnancy and the first year post delivery, and iii) identification of newborns at higher risk for sudden infant death syndrome (SIDS). The combination of several of the tools developed in this Ph.D. is expected to inform on maternal, fetal, and neonatal longitudinal risk profiles when paired with a neural network architecture able to take advantage of the wealth of high dimensional heterogeneous data source. Specifically, an ensemble of recurrent networks offers the potential for

a longitudinal risk assessment able to account for the highly dynamical evolution of pregnancy and early stages of life.

Replication, novel data acquisition modalities, and the role of wearables/nearables

In parallel to the development of new data analytics we are actively working on applying several of the developed tools on various cohorts. As first step, we envision applying the tools for the detection of the IUGR condition in the Safe Passage Study. As more extensively described in the previous sections, this study has recruited approximately 12,000 women in two diverse sites, namely in South Dakota, United States and South Africa. Pregnant women enrolled in the study were followed up at multiple time points during pregnancy. For the purpose of replication, the investigation of the IUGR condition in a longitudinal and dynamical fashion, would inform on the accuracy and reliability of the developed tools as well as the stability of the selected features. Additionally, the wealth of clinical and demographic data collected in the context of the study is expected to complement the results obtained in the outlined investigation. Furthermore, the same set of data could potentially be exploited for the purpose of addressing prediction of maternal CVH, SIDS, and several other maternal as well as fetal outcomes. Additionally, we envision the addition of fetal information in the framework for the prediction of neurodevelopmental trajectories. A secondary approach for the validation of the developed techniques sees the utilization of data collected in a purely clinical setting. We are closely working with various Departments at New York Presbyterian Hospital to replicate the presented results utilizing data collected during routine examinations. This would inform on the efficacy of the described models in the context of a less controlled environment and with a set of reduced features.

More recently, accelerated by the SARS-CoV-2 outbreak we have been extensively working on the deployment of several wearable monitoring solutions in the context of preexisting studies. In particular, we have been focusing on the conceptualization of alternative strategy for the remote collection of physiology data throughout pregnancy. As an example, we are currently testing a wearable device able to collect fetal ECG in a noninvasive way, at any time during the day and longitudinally throughout pregnancy. Specifically, we became interested in the relationship between maternal and fetal physiology during sleep. Thus, pairing the fetal monitoring device with an additional wearable to monitor maternal sleep allows us to generate an inexpensive solution to gain insight on circadian rhythms, and interrogate biological system in a network physiology perspective. To support this approach, we are actively developing a framework for the

automatic classification of fetal sleep states as well as the validation of an automatic sleep monitor for pregnant women via transfer learning, starting from the wealth of methodologies available in the context of adult research.

Appendix I

The presented Supplementary Material provides a more in-depth description of the steps towards the estimation of parameters employed in this study starting from the FHR signal. The analyzed FHR recordings were sampled at 2 Hz and they were measured in beats per minutes (bmp).

Time domain parameters:

Each FHR value coming from Philips fetal monitor Avalon FM30 was transformed in equivalent RR interval and expressed in milliseconds for the computation of time domain parameters.

Short Term Variability (STV) **[ms]**: it quantifies FHR variability on a short time scale. Considering an interbeat sequence of 1-minute duration, STV is defined as:

$$STV = \text{mean}[|T(i+1) - T(i)|]_i = \frac{\sum_{i=1}^{23}|T(i+1) - T(i)|}{23}$$

where T is the average FHR computed by diving the FHR recording in nonoverlapping windows of 5 consecutive FHR values (2.5 s for a sampling frequency of 2 Hz). STV is computed in a window of 1-minute duration so that 24 T values are obtained for each window. The corresponding STV estimate is obtained by averaging the differences between adjacent $T(i)$ values, having accelerations and deceleration excluded [40].

Interval Index (II): it provides an estimated of short term variability scaled by STV, defined as:

$$II = \frac{\text{std}[|T(i+1) - T(i)|]_i}{STV}$$

where $i = 1, \dots, 23$ are the total number of T FHR values recorded in 1 minute [43].

Delta **[ms]**: considering a window in time of 1-minute duration, Delta is defined as:

$$Delta = \max\big(T(i)\big) - \min\big(T(i)\big)$$

where $i = 1, \dots, 24$ which are the total number of T FHR values recorded in 1 minute [58].

Long Term Irregularity (LTI) **[ms]**: is defined as the interquartile range of the distribution $m(j)$ which is defined as:

$$m(j) = \sqrt{T^2(j+1) + T^2(j)}$$

where $j = 1, \dots, 71$ which are the total number of T FHR values in a 3-minutes window. LTI quantifies FHR variability on a longer time scale with respect to the previously reported time domain indices excluding accelerations and decelerations [58].

Frequency domain parameters:

Each FHR value coming from GE Corometrics 170 was transformed in equivalent RR interval and expressed in milliseconds for the computation of frequency domain parameters.

PSD Power Spectral Density (PSD) is a widely employed tool for HRV frequency analysis as it can quantitatively measure the periodic oscillations related to neural control activity, namely autonomic nervous system (ANS) modulation over the cardiac system.

In the context of this analysis, the PSD estimation for FHR was performed based on autoregressive (AR) modelling (parametric spectral estimation). The AR model utilized to mimic FHR dynamic is defined as:

$$\widehat{FHR}(j) = \sum_{i=1}^{p} a_i \cdot FHR(n-i) + w_j$$

where $w_j \sim WGN(0, \sigma^2)$ (White Gaussian Noise), p is the model order (from 8 to 12), and a_i are the model parameters. The modelled FHR windows ($\widehat{FHR}(j)$) are of duration equal to 3 minutes ($j = 1, \dots, 360$).

Model parameters are calculated recursively by means of the Levinson-Durbin algorithm. Once the proper model order is defined, so that the model parameters are determined, PSD is defined as:

$$PSD(f) = \frac{\sigma^2 \Delta}{\left|1 - \sum_{k=1}^{p} a_k e^{-j2\pi k f \Delta}\right|^2} = \frac{\sigma^2 \Delta}{A(e^{j2\pi f \Delta}) A^*(e^{j2\pi f \Delta})} = \left.\frac{\sigma^2 \Delta}{A(z) A^*\left(\frac{1}{z^*}\right)}\right|_{z = e^{j2\pi f \Delta}}$$

where Δ is the mean value of $\widehat{FHR}(j)$ in seconds and $A(z)$ is the z-transform of the transfer function of the AR process previously defined. Through this parametric approach, FHR signal undergoes an automatic decomposition into a sum of sinusoidal contributions, themselves identified by their corresponding central frequencies and the associated power [43].

In the context of frequency analysis applied to FHR it is possible to identify three specific frequency bands of interest, namely LF [0.03-0.15] Hz; MF [0.15-0.5] Hz; HF [0.5-1 Hz]. The

corresponding associated power is computed by integrating the PSD power in these intervals, thus obtaining ***LF_pow*** **[ms²/Hz];** ***MF_pow*** **[ms²/Hz];** ***HF_pow*** **[ms²/Hz].** Lastly, ***LF/(MF+HF)*** is obtained as: *LF_pow/(MF_pow+ HF_pow)* [43].

Nonlinear domain parameters:

Each FHR value coming from GE Corometrics 170 was transformed in equivalent RR interval and expressed in milliseconds for the computation of Approximate Entropy and Lempel and Ziv complexity.

Approximate Entropy (ApEn) **[bits]**: quantifies a signal regularity by assessing the occurrence rate of patterns by comparing the patters themselves to a reference one of length m. Pattern similarity is defined based on a threshold r [61].

Given a sequence of N data points $u(i)$, $i = 1, \dots, N$ the algorithm creates sequences $x_m(i)$ (based on window length m) and it computes for each $i \leq N - m + 1$ the quantity expressed as:

$$C_i^m(r) = \frac{1}{N}\{\text{count of } j \leq N - m + 1 \mid \text{distance } [x_m(i), x_m(j)] \leq r\}$$

Approximate Entropy (ApEn) is defined as:

$$ApEn(m, r) = \lim_{N \to \infty} [\Phi^m(r) - \Phi^{m+1}(r)]$$

where $\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln\left(C_i^m(r)\right)$.

In the context of finite time series of length N as for FHR, ApEn can be written as:

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r)$$

In this work, ApEn was estimated by considering non-overlapping windows of duration equal to 3 minutes, with the following parameter setting: $m = 1, r = 0.1, N = 360$ and named ***ApEn(1,0.1)*** [46], [61].

Lempel and Ziv complexity (LZC) **[bits]**: the first step towards its formulation encompasses the definition of the quantity $c(n)$ which measures the number of different sub strings and the rate of their recurrence in a given time series. According to the Information Theory, in turn it assesses the minimum quantity of information needed to define a binary string. LZC quantifies the rate of new patterns arising as signal evolves [45].

Suppose the number of symbols in the alphabet A is α and the length of sequence is equal n. The upper bound for $c(n)$ is given by:

$$c(n) < \frac{n}{(1-\varepsilon_n)\cdot \log(n)}$$

where

$$\varepsilon_n = 2\frac{1+\log\left(\log(\alpha n)\right)}{\log(n)}$$

and $\log(x)$ means the logarithm of x to the base α.

When n is large enough ($n \to \infty$), $\varepsilon_n \to 0$ so as a result:

$$b(n) = \lim_{n\to\infty} c(n) = \frac{n}{\log(n)}$$

where $b(n)$ is the asymptotic behavior of $c(n)$ for a random string.

The normalized complexity is thus defined as:

$$C(n) = \frac{c(n)}{b(n)}$$

In order to obtain LZC estimation for FHR time series, the latter requires to be transformed into a symbolic sequence according to a binary and/or a ternary coding procedure.

Binary coding: given a FHR series $x(N)$, the sequence $y(N)$ is built by assigning 1 when the condition $x(n+1) > x(n)$ is verified for $n = 1, \ldots, N$. On the opposite case of signal decrease, $y(n)$ is assigned to 0 when the condition $x(n+1) \leq x(n)$ is met.

Ternary coding: given a FHR series $x(N)$, the sequence $y(N)$ is built as in the binary coding case with the additional condition of signal invariance which is defined as $x(n+1) = x(n)$ and coded with the symbol 2 [45].

Additionally, in the context of recorded time series a factor p is introduced to define the minimum quantization level for a symbol change in the coded string (e.g. $y(n) = 1$ if $x(n+1) > x(n) + x(n)\cdot p$.

In this work, LZC was estimated by considering non-overlapping windows of duration equal to 3 minutes, with the following parameter setting: binary coding and $p = 0$ and named ***LZC(2,0)***. The choice of $p = 0$ reflects the current value for the quantization level, which is actually ± 0.5 bpm [84].

Each FHR value coming from GE Corometrics 170 was expressed in bpm for the computation of Acceleration Phase Rectified Slope and Deceleration Phase Rectified Slope in order to be concordant with the common definition of acceleration and deceleration in fetal heart rate monitoring

Acceleration Phase Rectified Slope (APRS) **[bpm]**: the computation of Phase Rectified Signal Averaging (PRSA) curve (which APRS is extracted from) starts from considering a time series x_i of length N ($i = 1, \dots, N$) as FHR in this work. The first step towards PRSA computation is the determination of the so-called anchor point (APs). In this context, APs are defined as the time series points x_i fulfilling the following inequality:

$$\frac{1}{M}\sum_{j=0}^{M-1} x_{i+j} > \frac{1}{M}\sum_{j=1}^{M} x_{i-j}$$

where the parameter M is employed as the upper frequency bound for the periodicities to be detected by PRSA methodology.

After APs being detected, windows of length $2L$ are built symmetrically with respect to each AP. Given the fact that the majority of APs are temporally close one each other, the resulting windows are effectively overlapping. An additional specification for the parameter L is that it should be larger than the period of slowest oscillation to be detected [69].

The PRSA curve X_i is obtained by averaging the derived windows synchronized in their APs. After obtaining the PRSA curve, it is useful to summarize its characteristics by extracting different parameters. An example of such is APRS defined as:

$$APRS = \left.\frac{\partial X_i}{\partial i}\right|_{i=AP}$$

The parameter APRS is a descriptor of the average increase in FHR amplitude and the time span of such increase event.

In this work, the considered signals x_i is the whole available FHR recording, thus resulting in a single APRS value. The parameters M and L are equal 40 and 200 respectively [46], [69].

Deceleration Phase Rectified Slope (DPRS) **[bpm]**: the computations are analogous of those previously reported for APRS apart for the definition of APs which are defined as the time series points x_i fulfilling [59]:

$$\frac{1}{M}\sum_{j=0}^{M-1} x_{i+j} < \frac{1}{M}\sum_{j=1}^{M} x_{i-j}$$

Appendix II: Quantification of Acceleration and Deceleration Capacities in Late Fetal Growth Restriction

Introduction

In recent years, antepartum fetal surveillance has been improving in predicting fetal pathological states also thanks to development of novel quantitative approaches for the investigation of Fetal Heart Rate (FHR). The novel approach of integrating clinical knowledge and signal processing tools emerged as promising for the diagnosis of the pathological fetal condition defined as Intrauterine Growth Restriction (IUGR). In this context, the investigation of FHR variability is the best proxy for the assessment of Autonomic Nervous System (ANS) system development, whose impairment is different depending on the Gestational Age (GA) at which the pathology emerges. Among the employed signal processing techniques, Phase Rectified Signal Averaging (PRSA) has been largely exploited for the identification of healthy and IUGR fetuses [46], [47], [70]. Nevertheless, PRSA depends on several parameters (s, T, and L) to set the range of FHR oscillations to be detected. Different parameter implementations have been proposed, e.g., setting s=T, s=T+1 [47], [70], or T=1 [46] Furthermore, there is a lack of consensus on PRSA implementation and scarce evidence of its predictive power as a function of pregnancy progression. In this work, we investigate the performance of PRSA-derived parameters, namely Acceleration Capacity (AC) and Deceleration capacity (DC) in a population of late IUGR fetuses (>36 weeks of gestation). The aim is to validate previous results obtained primarily in populations of early IUGR (<32 weeks of gestation), with the ultimate expectation of a more in depth understanding of the relationship between ANS development/impairment and the prognostic information of PRSA.

Materials and methods

Participants and data collection

Antepartum FHR recordings were collected at the Azienda Ospedaliera Universitaria Federico II, Napoli, Italy, during daily routine monitoring. Data were recorded using Philips cardiotocography (CTG) fetal monitor Avalon FM30 connected to a computer. The device employs an autocorrelation technique to compare the demodulated Doppler signal of a given

heartbeat and the subsequent one. The resulting resolution for beat detection is below 2 ms. The derived CTG signal consist of a series of FHR values sampled at 2 Hz and expressed in beats per minute (bpm). Additionally, each FHR sample is accompanied by an indication of signal quality: optimal, acceptable, or insufficient based on the results of autocorrelation technique previously described. The cohort analyzed in this work comprises 102 late Intrauterine Growth Restriction (IUGR) fetuses, and 160 healthy fetuses matched for GA at the first CTG examination. Fetuses in the healthy and IUGR populations underwent a routine ultrasound examination at around 34 weeks GA which did not exhibit any alteration in fetal growth, abnormalities in the doppler velocimetry, middle cerebral artery, and ductus venosus. Once subsequently admitted for a CTG recording at 37.54 ± 0.77 (mean ± std) weeks, fetuses in the healthy group did not shown any abnormality in the FHR trace. On the contrary, IUGR fetuses (admitted at 36.94 ± 0.59 weeks) did present irregularities in their CTG recordings. The suspected diagnosis of IUGR condition was confirmed by a concurrent ultrasound examination which showed alteration in both growth and doppler profiles. Each prenatal fetal condition was verified after delivery to confirm group membership previously suspected at the CTG timepoint. The GA weeks at the time of CTG assessment did not differ between the two groups and the length of the CTG recordings considered in this study is 40 minutes in both populations.

Phase Rectified Signal Averaging

PRSA methodology as well as PRSA-derived parameters have been shown to provide an estimate of ANS regulation of FHR under a variety of physiological and pathological fetal conditions [46], [47]. The rationale for the application of such technique in this context relies on its robustness and partial insensitivity to abrupt changes in the system, miss-detection of heartbeats, and signal losses common in CTG recordings.

In a nutshell, a set of anchor points is determined from the FHR series. Focusing on the derivation of AC parameter, each point satisfying the condition reported in the equation below is included into the accelerations' list (AAC):

$$A_{AC} = \left\{ t : \frac{1}{T}\sum_{i=0}^{T-1} FHR[t+i] > \frac{1}{T}\sum_{i=1}^{T} FHR[t-i] \right\}$$

where T is an integer used to set the timescale.

A window of length 2L samples is centered at the location of each anchor point. Then, the derived windows are aligned and averaged, obtaining the PRSA series. This computed series is used to calculate AC accordingly to (2):

$$AC = \frac{1}{2s}\sum_{i=1}^{s} PRSA[L+i] - \frac{1}{2s}\sum_{i=0}^{s-1} PRSA[L-i]$$

where s is the scale and L determines the span of PRSA series. DC is computed accordingly to the equation above but considering a different PRSA series obtained by modifying the direction of inequality in the definition of AAC accordingly. As a consequence, DC values are negative by definition. For simplicity, in this work only absolute values of DC are reported to provide and facilitate their comparison with AC. To summarize, T sets the order of the low-pass moving average to detect anchor points, L determines the length of the PRSA curve which in turn sets the limit for the period of the slowest oscillation to be detected, and s selects the oscillatory components that most affect AC and DC.

Statistical analysis

FHR values meeting the criterion to be considered as anchor point according to AAC but associated with insufficient quality (as defined by the acquisition systems) were discarded from AAC. Values of AC and DC were computed by setting the timescale and scale parameter to be equal, namely s=T. The pairs of s and T were computed on a continuous range from 1 to 10, and thereafter from 15 to 60 for increasing values in step of 5. A between group analysis was performed to compare the trends of AC and DC between the two groups. Differences in group comparisons were evaluated by means of the t-tests by adjusting the threshold for significance according to a Bonferroni correction for multiple group comparisons.

Results

AC and DC parameters were characterized by an increasing trend across the considered pairs of s and T values, for both groups. Such results are displayed in the left and right panels of Figure A for AC, and DC respectively. It was possible to observe a steeper increase of both AC and DC for smaller timescale (1 ÷ 10), followed by a less pronounced and positive trend for the higher scales (15 ÷ 60).

AC and DC exhibited an adequate ability in discriminating healthy and pathological fetuses as confirmed by statistical analyses. Both AC and DC distributions are statistically different between the two groups and across the entire span of s=T considered in this analysis. On note, the separation between the two groups appeared to be enhanced at the higher times scales (15 ÷ 60). This latter finding was especially evident for the parameter DC.

Discussion

AC and DC parameters were found to be on average lower for the late IUGR fetuses with respect to the healthy ones. These results are in agreement with the findings reported in [47], [70], where a similar analysis was performed on early IUGR groups at two different GA. In their studies, the increasing trend of the parameter T (and consequently s) was associated with larger average difference and standard deviation for both AC and DC in the comparison between the two groups. We observed this behavior in our cohort as well (Figure A). Specifically, we found a statistically significantly difference between healthy and late IUGR fetuses for both capacities in the entire spectrum of T values considered excluding T=1 as consistently reported in [70] (despite a different

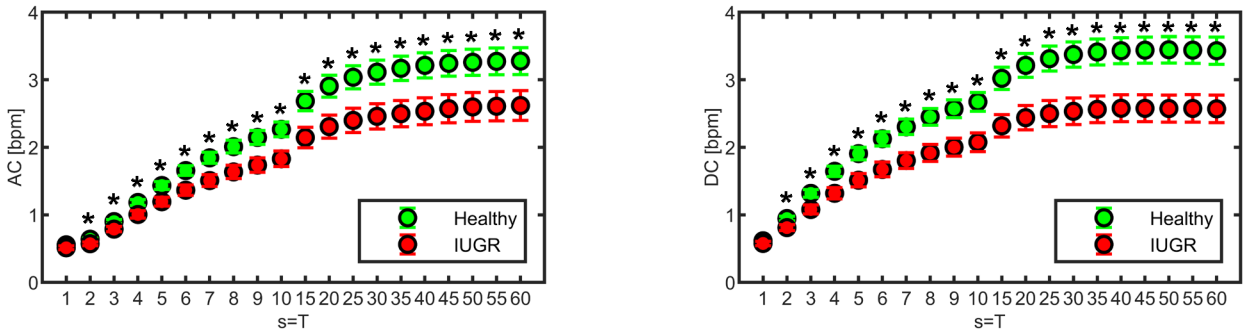


Figure A. AC and DC (mean and 95% CI) distributions over a range of pairs s and T, comparing healthy versus IUGR fetuses. * denotes $p < 0.05$.

sampling rate was employed). AC and DC were therefore able to capture differences in autonomic regulations as a consequence of nutrient restriction and intermittent hypoxemia, two conditions typical of IUGR pathology.

The study presents a few limitations. First, CTG recordings do not provide a true beat-to-beat information, thus possibly masking ANS contributions at very low T values. Moreover, as recently reported in [48] PRSA is linearly dependent on the standard deviation of the signal. Therefore, the differences reported in this work might be confounded by this signal intrinsic factor, rather than representing an absolute separation between groups. We leave this investigation as future work.

List of Publications

Journal Papers:

- Lucchini, M., **Pini, N.**, Fifer, W. P., Burtchen, N., & Signorini, M. G. (2017). Entropy information of cardiorespiratory dynamics in Neonates during sleep. *Entropy*, *19*(5), 225.

- Lucchini, M., **Pini, N.**, Fifer, W. P., Burtchen, N., & Signorini, M. G. (2018). Characterization of cardiorespiratory phase synchronization and directionality in late premature and full term infants. *Physiological measurement*, *39*(6), 064001.

- Signorini, M. G., **Pini, N.**, Malovini, A., Bellazzi, R., & Magenes, G. (2020). Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring. *Computer methods and programs in biomedicine*, *185*, 105015.

- Signorini, M. G., **Pini, N.**, Malovini, A., Bellazzi, R., & Magenes, G. (2020). Dataset on linear and non-linear indices for discriminating healthy and IUGR fetuses. *Data in brief*, *29*, 105164.

- Shuffrey, L. C., Myers, M. M., Isler, J. R., Lucchini, M., Sania, A., **Pini, N.**, ... & du Plessis, C. (2020). Association Between Prenatal Exposure to Alcohol and Tobacco and Neonatal Brain Activity: Results From the Safe Passage Study. *JAMA network open*, *3*(5), e204714-e204714.

- Sania, A., **Pini, N.**, Nelson, M. E., Myers, M. M., Shuffrey, L. C., Lucchini, M., ... & Fifer, W. P. (2020). The K nearest neighbor algorithm for imputation of missing longitudinal prenatal alcohol data. Under review in *BMC Medical Research Methodology.*

- **Pini, N.**, Lucchini, M., Burtchen, N., Signorini, M. G., & Fifer, W. P. (2020). Transfer Entropy Modeling of Newborn Cardiorespiratory Regulation. *Frontiers in Physiology*, *11*, 1095.

- **Pini, N.**, Lucchini, M., Esposito, G., Campanile, M., Magenes, G., & Signorini, M. G. A Machine Learning Approach to Monitor the Emergence of Late Intrauterine Growth Restriction. Under review in *Frontiers in Artificial Intelligence.*

Conference Proceedings:

- Lucchini M., **Pini N.**, Fifer W.P., Burtchen N., & Signorini M.G. (2018) Cardio-respiratory phase locking in newborn and one month infants as a function of sleep state. *In Eskola H., Väisänen O., Viik J., Hyttinen J. (eds) EMBEC & NBC 2017. EMBEC 2017, NBC 2017.* IFMBE Proceedings, vol 65. Springer, Singapore.

- **Pini, N.**, Lucchini, M., Fifer, W. P., Myers, M. M., & Signorini, M. G. (2018, July). Influence of prenatal alcohol and smoke exposure on neonatal vagal tone in response to head-up tilt. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5874-5877). IEEE.

- **Pini N.**, Lucchini M., Fifer W.P., Burtchen N., & Signorini M. G. (2019, February). Lagged transfer entropy analysis to investigate cardiorespiratory regulation in newborns during sleep. In *2019 12th International Conference on Bio-Inspired System and Signal Processing (BIOSIGNAL).*

- **Pini, N.**, Lucchini, M., Fifer, W. P., Signorini, M. G., & Barbieri, R. (2019, July). A Point Process Framework for the Characterization of Sleep States in Early Infancy. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3645-3648). IEEE.

- **Pini, N.**, Magenes, G., Fanelli, A., & Signorini, M. G. (2019, July). An Efficient Algorithm for the Extraction of Fetal ECG from Standard and Non-Standard Multi Abdominal Maternal Leads. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5717-5720). IEEE.

- **Pini, N.**, Shuffrey, L. C., Lucchini, M., Sania, A., Nelson, M. E., Nugent, J. D., ... & Elliott, A. J. (2019, July). Cluster Analysis of Alcohol Consumption during Pregnancy in the Safe Passage Study. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1338-1341). IEEE.

- Signorini, M. G., **Pini, N.**, Pani, D., & Magenes, G. (2019, September). ICT4MOMs: An ICT Integrated Approach to Monitor and Manage Pregnancy Development. In *Mediterranean Conference on Medical and Biological Engineering and Computing* (pp. 924-929). Springer, Cham.

- **Pini, N.**, Lucchini, M., Fifer, W. P., & Barbieri, R. (2020, July). A Point Process Framework for the Characterization of Fetal Sleep States. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* In press.

- **Pini, N.**, Lucchini M., Campanile, M., Rivolta, M. W., Magenes, G., & Signorini, M. G. (2020, July). Quantification of Acceleration and Deceleration Capacities in Late Fetal Growth Restriction. In *ESGCO 2020 International Conference.* In press.

- **Pini, N.**, Rivolta, M. W., Shair, M., Elliott, A. J., Fifer, W. P., & Lucchini, M. (2020, September). Phase Rectified Signal Averaging Technique Improves Characterization of Sleep State in Healthy Fetuses. In *2020 Computing in Cardiology (CinC).* In press.

References

[1] L. R. Blackmon *et al.*, "Age terminology during the perinatal period," *Pediatrics*, vol. 114, no. 5, pp. 1362–1364, Nov. 2004, doi: 10.1542/peds.2004-1915.

[2] P. D. Gluckman, M. A. Hanson, D. Phil, C. Cooper, and K. L. Thornburg, "Effect of In Utero and Early-Life Conditions on Adult Health and Disease," 2008, doi: 10.1056/NEJMra0708473.

[3] U. Simeoni, J. B. Armengaud, B. Siddeek, and J. F. Tolsa, "Perinatal Origins of Adult Disease," *Neonatology*, vol. 113, no. 4, pp. 393–399, May 2018, doi: 10.1159/000487618.

[4] D. J. P. Barker, "The Developmental Origins of Adult Disease," *J. Am. Coll. Nutr.*, vol. 23, no. sup6, pp. 588S-595S, Dec. 2004, doi: 10.1080/07315724.2004.10719428.

[5] B. E. Stephens and B. R. Vohr, "Neurodevelopmental Outcome of the Premature Infant," *Pediatric Clinics of North America*, vol. 56, no. 3. pp. 631–646, Jun. 2009, doi: 10.1016/j.pcl.2009.03.005.

[6] W. A. Engle, K. M. Tomashek, and C. Wallman, "'Late-Preterm' Infants: A Population at Risk," *Pediatrics*, vol. 120, no. 6, pp. 1390–1401, 2007.

[7] A. Delpisheh, L. Brabin, S. Drummond, and B. J. Brabin, "Prenatal smoking exposure and asymmetric fetal growth restriction," *Ann. Hum. Biol.*, vol. 35, no. 6, pp. 573–583, Jan. 2008, doi: 10.1080/03014460802375596.

[8] E. A. Huhn *et al.*, "New computerized fetal heart rate analysis for surveillance of intrauterine growth restriction," *Prenat. Diagn.*, vol. 31, no. 5, pp. 509–514, 2011, doi: 10.1002/pd.2728.

[9] M. G. Signorini, N. Pini, A. Malovini, R. Bellazzi, and G. Magenes, "Dataset on linear and non-linear indices for discriminating healthy and IUGR fetuses," *Data Br.*, vol. 29, p. 105164, Jan. 2020, doi: 10.1016/j.dib.2020.105164.

[10] K. B. Mikkelsen *et al.*, "Machine-learning-derived sleep–wake staging from around-the-ear electroencephalogram outperforms manual scoring and actigraphy," *J. Sleep Res.*, vol. 28, no. 2, Apr. 2019, doi: 10.1111/jsr.12786.

[11] M. Peker, "An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms," *Neurocomputing*, vol. 207, pp. 165–177, Sep. 2016, doi: 10.1016/j.neucom.2016.04.049.

[12] H. A. Guvenir, B. Acar, G. Demiroz, and A. Cekin, “Supervised machine learning algorithm for arrhythmia analysis,” in *Computers in Cardiology*, 1997, pp. 433–436, doi: 10.1109/cic.1997.647926.

[13] A. Özçift, “Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis,” *Comput. Biol. Med.*, vol. 41, no. 5, pp. 265–271, May 2011, doi: 10.1016/j.compbiomed.2011.03.001.

[14] J. Balayla and G. Shrem, “Use of artificial intelligence (AI) in the interpretation of intrapartum fetal heart rate (FHR) tracings: a systematic review and meta-analysis,” *Archives of Gynecology and Obstetrics*, vol. 300, no. 1. Springer Verlag, pp. 7–14, Jul. 01, 2019, doi: 10.1007/s00404-019-05151-7.

[15] P. S. Zeskind and J. L. Gingras, “Maternal cigarette-smoking during pregnancy disrupts rhythms in fetal heart rate,” *J. Pediatr. Psychol.*, vol. 31, no. 1, pp. 5–14, Jan. 2006, doi: 10.1093/jpepsy/jsj031.

[16] L. Burd, D. Roberts, M. Olson, and H. Odendaal, “Ethanol and the placenta: A review,” *J. Matern. Neonatal Med.*, vol. 20, no. 5, pp. 361–375, 2007, doi: 10.1080/14767050701298365.

[17] W. P. Fifer, S. Ten Fingers, M. Youngman, E. Gomez-Gribben, and M. M. Myers, “Effects of alcohol and smoking during pregnancy on infant autonomic control,” *Dev. Psychobiol.*, vol. 51, no. 3, pp. 234–242, Apr. 2009, doi: 10.1002/dev.20366.

[18] K. Grewen *et al.*, “Prenatal cocaine effects on brain structure in early infancy,” *Neuroimage*, vol. 101, pp. 114–123, Nov. 2014, doi: 10.1016/j.neuroimage.2014.06.070.

[19] D. B. Rubin, “Inference and Missing Data,” *Biometrika*, vol. 63, no. 3, p. 581, Dec. 1976, doi: 10.2307/2335739.

[20] J. F. Williams and V. C. Smith, “Fetal alcohol spectrum disorders,” *Pediatrics*, vol. 136, no. 5. American Academy of Pediatrics, pp. e1395–e1406, Nov. 01, 2015, doi: 10.1542/peds.2015-3113.

[21] L. M. O’Keeffe *et al.*, “Prevalence and predictors of alcohol use during pregnancy: Findings from international multicentre cohort studies,” *BMJ Open*, vol. 5, no. 7, p. e006323, May 2015, doi: 10.1136/bmjopen-2014-006323.

[22] J. J. K. Jaakkola and M. Gissler, “Maternal Smoking in Pregnancy, Fetal Development, and Childhood Asthma,” *Am. J. Public Health*, vol. 94, no. 1, pp. 136–140, 2004, doi:

10.2105/AJPH.94.1.136.

[23] J. Cornman-Homonoff *et al.*, “Heavy prenatal alcohol exposure and risk of stillbirth and preterm delivery,” *J. Matern. Neonatal Med.*, vol. 25, no. 6, pp. 860–863, Jun. 2012, doi: 10.3109/14767058.2011.587559.

[24] K. Dukes *et al.*, “Drinking and smoking patterns during pregnancy: Development of group-based trajectories in the Safe Passage Study,” *Alcohol*, vol. 62, pp. 49–60, 2017, doi: 10.1016/j.alcohol.2017.03.001.

[25] T. Baker *et al.*, “A modified Timeline Followback assessment to capture alcohol exposure in pregnant women: Application in the Safe Passage Study,” *Alcohol*, vol. 62, no. Supplement C, pp. 17–27, 2017, doi: 10.1016/j.alcohol.2017.02.174.

[26] W. Dai, Q. Yang, G. R. Xue, and Y. Yu, “Boosting for transfer learning,” in *ACM International Conference Proceeding Series*, 2007, vol. 227, pp. 193–200, doi: 10.1145/1273496.1273521.

[27] R. E. Kass *et al.*, “Finite Mixture Models,” *Annu. Rev. Stat. Appl*, vol. 5, pp. 6.1-6.27, 2018, doi: 10.1146/annurev-statistics.

[28] C. Genolini, X. Alacoque, M. Sentenac, and C. Arnaud, “Kml and kml3d: R packages to cluster longitudinal data,” *J. Stat. Softw.*, vol. 65, no. 4, pp. 1–34, May 2015, doi: 10.18637/jss.v065.i04.

[29] C. Genolini, R. Ecochard, M. Benghezal, T. Driss, S. Andrieu, and F. Subtil, “kmlShape: An efficient method to cluster longitudinal data (Time-Series) according to their shapes,” *PLoS One*, vol. 11, no. 6, p. e0150738, Jun. 2016, doi: 10.1371/journal.pone.0150738.

[30] J. Bélair, L. Glass, U. An Der Heiden, and J. Milton, “Dynamical disease: Identification, temporal aspects and treatment strategies of human illness,” *Chaos*, vol. 5, no. 1, pp. 1–7, Mar. 1995, doi: 10.1063/1.166069.

[31] R. P. Bartsch, K. K. L. Liu, A. Bashan, and P. C. Ivanov, “Network physiology: How organ systems dynamically interact,” *PLoS One*, vol. 10, no. 11, 2015, doi: 10.1371/journal.pone.0142143.

[32] W. W. Stead, “Clinical implications and challenges of artificial intelligence and deep learning,” *JAMA - J. Am. Med. Assoc.*, vol. 320, no. 11, pp. 1107–1108, Sep. 2018, doi: 10.1001/jama.2018.11029.

[33] M. Ghassemi, T. Naumann, P. Schulam, A. L. Beam, I. Y. Chen, and R. Ranganath,

"Practical guidance on artificial intelligence for health-care data," *Lancet Digit. Heal.*, vol. 1, no. 4, pp. e157–e159, Aug. 2019, doi: 10.1016/s2589-7500(19)30084-6.

[34] A. Rosenberg, "The IUGR Newborn," *Semin. Perinatol.*, vol. 32, no. 3, pp. 219–224, 2008, doi: 10.1053/j.semperi.2007.11.003.

[35] M. Parra-Saavedra *et al.*, "Placental findings in late-onset SGA births without Doppler signs of placental insufficiency," *Placenta*, vol. 34, no. 12, pp. 1136–1141, Dec. 2013, doi: 10.1016/j.placenta.2013.09.018.

[36] F. Figueras and J. Gardosi, "Intrauterine growth restriction: New concepts in antenatal surveillance, diagnosis, and management," *Am. J. Obstet. Gynecol.*, vol. 204, no. 4, pp. 288–300, 2011, doi: 10.1016/j.ajog.2010.08.055.

[37] F. G. Esposito *et al.*, "Fetal heart rate monitoring and neonatal outcome in a population of early- and late-onset intrauterine growth restriction," *J. Obstet. Gynaecol. Res.*, vol. 45, no. 7, pp. 1343–1351, 2019, doi: 10.1111/jog.13981.

[38] J. Smith, M. Murphy, and Y. Kandasamy, "The IUGR infant: A case study and associated problems with IUGR infants," *J. Neonatal Nurs.*, vol. 19, no. 2, pp. 46–53, 2013, doi: 10.1016/j.jnn.2012.12.005.

[39] D. Sharma, S. Shastri, and P. Sharma, "Intrauterine Growth Restriction: Antenatal and Postnatal Aspects," *Clin. Med. Insights Pediatr.*, vol. 10, p. CMPed.S40070, 2016, doi: 10.4137/cmped.s40070.

[40] N. Giuliano *et al.*, "Computerised analysis of antepartum foetal heart parameters: New reference ranges," *J. Obstet. Gynaecol. (Lahore).*, vol. 37, no. 3, pp. 296–304, 2017, doi: 10.1080/01443615.2016.1239069.

[41] T. Todros, C. U. Preve, C. Plazzotta, M. Biolcati, and P. Lombardo, "Fetal heart rate tracings: Observers versus computer assessment," 1996. doi: 10.1016/0301-2115(96)02487-6.

[42] G. Magenes, M. G. Signorini, D. Arduini, and S. Cerutti, "Fetal Heart Rate Variability due to Vibroacoustic Stimulation: Linear and Nonlinear Contribution," *Methods Inf. Med.*, vol. 43, no. 1, pp. 47–51, 2004, doi: 10.1267/METH04010047.

[43] M. G. Signorini, G. Magenes, S. Cerutti, and D. Arduini, "Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 365–374, 2003, doi: 10.1109/TBME.2003.808824.

[44] M. Ferrario, M. G. Signorini, and G. Magenes, “Comparison between fetal heart rate standard parameters and complexity indexes for the identification of severe intrauterine growth restriction,” *Methods Inf. Med.*, vol. 46, no. 2, pp. 186–190, 2007, doi: 10.1055/s-0038-1625404.

[45] A. Lempel and J. Ziv, “On the complexity of finite sequences,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 75–81, 1976, doi: 10.1109/TIT.1976.1055501.

[46] A. Fanelli, G. Magenes, M. Campanile, and M. G. Signorini, “Quantitative assessment of fetal well-being through ctg recordings: A new parameter based on phase-rectified signal average,” *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 5, pp. 959–966, 2013, doi: 10.1109/JBHI.2013.2268423.

[47] T. Stampalija *et al.*, “Parameters influence on acceleration and deceleration capacity based on trans-abdominal ECG in early fetal growth restriction at different gestational age epochs,” *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 188, pp. 104–112, 2015, doi: 10.1016/j.ejogrb.2015.03.003.

[48] M. W. Rivolta, T. Stampalija, M. G. Frasch, and R. Sassi, “Theoretical Value of Deceleration Capacity Points to Deceleration Reserve of Fetal Heart Rate,” *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2019, doi: 10.1109/tbme.2019.2932808.

[49] M. G. Signorini, N. Pini, A. Malovini, R. Bellazzi, and G. Magenes, “Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring,” *Comput. Methods Programs Biomed.*, vol. 185, p. 105015, 2020, doi: 10.1016/j.cmpb.2019.105015.

[50] N. Pini *et al.*, “A Machine Learning Approach to Monitor the Emergence of Late Intrauterine Growth Restriction,” *Under Rev. Front. Artif. Intell.*, 2020.

[51] N. Pini, M. Lucchini, W. P. Fifer, and R. Barbieri, “A Point Process Framework for the Characterization of Fetal Sleep States,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Jul. 2020, vol. 2020-July, pp. 612–615, doi: 10.1109/EMBC44109.2020.9176169.

[52] N. Pini, G. Magenes, A. Fanelli, and M. G. Signorini, “An Efficient Algorithm for the Extraction of Fetal ECG from Standard and Non-Standard Multi Abdominal Maternal Leads,” *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2019, pp. 5717–5720, Jul. 2019, doi: 10.1109/EMBC.2019.8858152.

[53] N. Pini, M. Lucchini, M. Campanile, M. W. Rivolta, G. Magenes, and M. G. Signorini, "Quantification of Acceleration and Deceleration Capacities in Late Fetal Growth Restriction," in *2020 11th Conference of the European Study Group on Cardiovascular Oscillations: Computation and Modelling in Physiology: New Challenges and Opportunities, ESGCO 2020*, Jul. 2020, doi: 10.1109/ESGCO49734.2020.9158015.

[54] M. B. Landon and S. G. Gabbe, "Antepartum fetal surveillance in gestational diabetes mellitus," *Diabetes*, vol. 34, no. SUPPL. 2, pp. 50–54, 1985, Accessed: Feb. 26, 2019. [Online]. Available: http://diabetes.diabetesjournals.org/content/34/Supplement_2/50.full-text.pdf.

[55] A. A. Baschat, R. M. Viscardi, B. Hussey-Gardner, N. Hashmi, and C. Harman, "Infant neurodevelopment following fetal growth restriction: Relationship with antepartum surveillance parameters," *Ultrasound Obstet. Gynecol.*, vol. 33, no. 1, pp. 44–50, Jan. 2009, doi: 10.1002/uog.6286.

[56] J. de Haan *et al.*, "Quantitative evaluation of fetal heart rate patterns: I. Processing methods," *Eur. J. Obstet. Gynecol.*, vol. 1, no. 3, pp. 95–102, Jan. 1971, doi: 10.1016/0028-2243(71)90056-6.

[57] D. Hoyer *et al.*, "Monitoring fetal maturation - Objectives, techniques and indices of autonomic function," *Physiol. Meas.*, vol. 38, no. 5, pp. R61–R88, 2017, doi: 10.1088/1361-6579/aa5fca.

[58] D. Arduini, G. Rizzo, A. Piana, P. Bonalumi, P. Brambilla, and C. Romanini, "Computerized analysis of fetal heart rate—Part I: description of the sys- tem (2CTG)," *J Matern Fetal Invest*, vol. 3, pp. 159–164, 1993.

[59] M. G. Signorini, A. Fanelli, and G. Magenes, "Monitoring fetal heart rate during pregnancy: Contributions from advanced signal processing and wearable technology," *Comput. Math. Methods Med.*, vol. 2014, 2014, doi: 10.1155/2014/707581.

[60] S. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos*, vol. 5, no. 1, pp. 110–117, Mar. 1995, doi: 10.1063/1.166092.

[61] S. M. Pincus and R. R. Viscarello, "Approximate entropy: a regularity measure for fetal heart rate analysis.," *Obstet. Gynecol.*, vol. 79, no. 2, pp. 249–55, 1992.

[62] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am. J. Physiol. Circ. Physiol.*, vol. 278, no. 6, pp. H2039–

H2049, 2000, doi: 10.1152/ajpheart.2000.278.6.H2039.

[63] G. Hinton, “Deep Learning—A Technology With the Potential to Transform Health Care,” *JAMA*, vol. 320, no. 11, p. 1101, Sep. 2018, doi: 10.1001/jama.2018.11100.

[64] C. D. Naylor, “On the prospects for a (Deep) learning health care system,” *JAMA - J. Am. Med. Assoc.*, vol. 320, no. 11, pp. 1099–1100, Sep. 2018, doi: 10.1001/jama.2018.11103.

[65] Hewlett-Packard, “Fetal Monitor Test - A Brief Summary,” *Hewlett-Packard Ger.*, pp. 1–6, 1995.

[66] G. Magenes, M. G. Signorini, M. Ferrario, and F. Lunghi, “2CTG2: A new system for the antepartum analysis of fetal heart rate,” in *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 781–784.

[67] H. Gonçalves, A. Costa, D. Ayres-de-Campos, C. Costa-Santos, A. P. Rocha, and J. Bernardes, “Comparison of real beat-to-beat signals with commercially available 4 Hz sampling on the evaluation of foetal heart rate variability,” *Med. Biol. Eng. Comput.*, vol. 51, no. 6, pp. 665–676, Jun. 2013, doi: 10.1007/s11517-013-1036-7.

[68] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, “Heart rate variability. Standard of measurement, physiological interpretation and clinical use,” *Circulation*, vol. 93, pp. 1043–1065, 1996.

[69] A. Bauer *et al.*, “Phase-rectified signal averaging detects quasi-periodicities in non-stationary data,” *Phys. A Stat. Mech. its Appl.*, vol. 364, pp. 423–434, 2006, doi: 10.1016/j.physa.2005.08.080.

[70] S. M. Lobmaier *et al.*, “Phase-rectified signal averaging as a new method for surveillance of growth restricted fetuses,” *J. Matern. Neonatal Med.*, vol. 25, no. 12, pp. 2523–2528, Dec. 2012, doi: 10.3109/14767058.2012.696163.

[71] R. C. and others Team, “R: A language and environment for statistical computing,” 2013.

[72] D. J. Stekhoven and P. Bühlmann, “Missforest—non-parametric missing value imputation for mixed-type data,” *Bioinformatics*, vol. 28, no. 1, pp. 112–118, Jan. 2012, doi: 10.1093/bioinformatics/btr597.

[73] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. 2013.

[74] H. Gonçalves, C. Amorim-Costa, D. Ayres-de-Campos, and J. Bernardes, “Evolution of linear and nonlinear fetal heart rate indices throughout pregnancy in appropriate, small for

gestational age and preterm fetuses: A cohort study," *Comput. Methods Programs Biomed.*, vol. 153, pp. 191–199, 2018, doi: 10.1016/j.cmpb.2017.10.015.

[75] G. Magenes, R. Bellazzi, A. Malovini, and M. G. Signorini, "Comparison of data mining techniques applied to fetal heart rate parameters for the early identification of IUGR fetuses," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2016, pp. 916–919, doi: 10.1109/EMBC.2016.7590850.

[76] C. E. McCulloch, "Generalized linear models," in *Statistics in the 21st Century*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 387–396.

[77] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," 1970. doi: 10.1080/00401706.1970.10488634.

[78] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," 2005. doi: 10.1111/j.1467-9868.2005.00503.x.

[79] R. Tibshirani, "Regression Shriknage and Selectino via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.2307/2346178.

[80] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Chapman & Hall/CRC, 1984.

[81] M. Krzywinski and N. Altman, "Classification and regression trees," *Nature Methods*, vol. 14, no. 8. Nature Publishing Group, pp. 757–758, Jul. 28, 2017, doi: 10.1038/nmeth.4370.

[82] Y. Zhang, "Support vector machine classification algorithm and its application," in *Communications in Computer and Information Science*, 2012, vol. 308 CCIS, no. PART 2, pp. 179–186, doi: 10.1007/978-3-642-34041-3_27.

[83] M. A. Hall, "Correlation-based Feature Selection for Machine Learning," 1999.

[84] M. Ferrario, M. G. Signorini, and G. Magenes, "Complexity analysis of the fetal heart rate for the identification of pathology in fetuses," *Comput. Cardiol.*, vol. 32, pp. 989–992, 2005, doi: 10.1109/CIC.2005.1588275.

[85] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[86] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.

[87] H. Y. Chen, S. P. Chauhan, C. V. Ananth, A. M. Vintzileos, and A. Z. Abuhamad,

"Electronic fetal heart rate monitoring and its relationship to neonatal and infant mortality in the United States," *Am. J. Obstet. Gynecol.*, vol. 204, no. 6, pp. 491.e1-491.e10, 2011, doi: 10.1016/j.ajog.2011.04.024.

[88] D. Mureşan, I. C. Rotar, and F. Stamatian, "The usefulness of fetal Doppler evaluation in early versus late onset intrauterine growth restriction. Review of the literature," *Medical Ultrasonography*, vol. 18, no. 1. Societatea Romana de Ultrasonografie in Medicina si Biologie, pp. 103–109, 2016, doi: 10.11152/mu.2013.2066.181.dop.

[89] J. Warland and E. A. Mitchell, "A triple risk model for unexplained late stillbirth," *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 142, 2014, doi: 10.1186/1471-2393-14-142.

[90] FIGO, "Guidelines for the Use of Fetal Monitoring," *Int. J. Gynecol. Obstet.*, vol. 25, pp. 159–167, 1986.

[91] J. Spilka, J. Frecon, R. Leonarduzzi, N. Pustelnik, P. Abry, and M. Doret, "Sparse Support Vector Machine for Intrapartum Fetal Heart Rate Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 3, pp. 664–671, 2017, doi: 10.1109/JBHI.2016.2546312.

[92] S. J. Gordijn *et al.*, "Consensus definition of fetal growth restriction: a Delphi procedure," *Ultrasound Obstet. Gynecol.*, vol. 48, no. 3, pp. 333–339, Sep. 2016, doi: 10.1002/uog.15884.

[93] R. Mantel, H. P. van Geijn, F. J. Caron, J. M. Swartjes, E. E. van Woerden, and H. W. Jongsma, "Computer analysis of antepartum fetal heart rate: 1. Baseline determination.," *Int. J. Biomed. Comput.*, vol. 25, no. 4, pp. 261–272, May 1990, Accessed: Mar. 26, 2018. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2194979.

[94] R. Rabinowitz, E. Persitz, and E. Sadovsky, "The relation between fetal heart rate accelerations and fetal movements," *Obstet. Gynecol.*, vol. 61, no. 1, pp. 16–18, 1983.

[95] J. Spilka *et al.*, "Using nonlinear features for fetal heart rate classification," *Biomed. Signal Process. Control*, vol. 7, no. 4, pp. 350–357, Jul. 2012, doi: 10.1016/j.bspc.2011.06.008.

[96] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," *Am. J. Physiol. - Regul. Integr. Comp. Physiol.*, vol. 283, no. 3, pp. R789–R797, 2002, doi: 10.1152/ajpregu.00069.2002.

[97] H. Gonçalves, C. Amorim-Costa, D. Ayres-de-Campos, and J. Bernardes, "Gender-specific evolution of fetal heart rate variability throughout gestation: A study of 8823 cases," *Early Hum. Dev.*, vol. 115, pp. 38–45, 2017, doi: 10.1016/j.earlhumdev.2017.09.002.

[98] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, “A practical guide to support vector classification,” 2003, doi: 10.1007/s11119-014-9370-9.

[99] Q. Liu, C. Chen, Y. Zhang, and Z. Hu, “Feature selection for support vector machines with RBF kernel,” *Artif. Intell. Rev.*, vol. 36, no. 2, pp. 99–115, 2011, doi: 10.1007/s10462-011-9205-2.

[100] H. Lin and C. Lin, “A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods,” 2003. doi: 10.1.1.14.6709.

[101] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997, doi: 10.1016/s0004-3702(97)00043-x.

[102] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Mach. Learn.*, vol. 46, no. 1–3, pp. 389–422, 2002, doi: 10.1023/A:1012487302797.

[103] Y. Sun, “Iterative RELIEF for feature weighting: Algorithms, theories, and applications,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1035–1051, Jun. 2007, doi: 10.1109/TPAMI.2007.1093.

[104] R. Fluss, D. Faraggi, and B. Reiser, “Estimation of the Youden Index and its associated cutoff point,” *Biometrical J.*, vol. 47, no. 4, pp. 458–472, 2005, doi: 10.1002/bimj.200410135.

[105] V. Van Belle and P. Lisboa, “Automated selection of interaction effects in sparse kernel methods to predict pregnancy viability,” in *Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013*, 2013, pp. 26–31, doi: 10.1109/CIDM.2013.6597213.

[106] V. Van Belle and P. Lisboa, “White box radial basis function classifiers with component selection for clinical prediction models,” *Artif. Intell. Med.*, vol. 60, no. 1, pp. 53–64, Jan. 2014, doi: 10.1016/j.artmed.2013.10.001.

[107] R. Poudel, I. C. McMillen, S. L. Dunn, S. Zhang, and J. L. Morrison, “Impact of chronic hypoxemia on blood flow to the brain, heart, and adrenal gland in the late-gestation IUGR sheep fetus,” *Am. J. Physiol. - Regul. Integr. Comp. Physiol.*, vol. 308, no. 3, pp. R151–R162, Feb. 2015, doi: 10.1152/ajpregu.00036.2014.

[108] M. Sanz-Cortés, R. J. Carbajo, F. Crispi, F. Figueras, A. Pineda-Lucena, and E. Gratacós,

"Metabolomic profile of umbilical cord blood plasma from early and late intrauterine growth restricted (IUGR) neonates with and without signs of brain vasodilation," *PLoS One*, vol. 8, no. 12, p. e80121, Dec. 2013, doi: 10.1371/journal.pone.0080121.

[109] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7.

[110] S. Popova *et al.*, "Comorbidity of fetal alcohol spectrum disorder: A systematic review and meta-analysis," *Lancet*, vol. 387, no. 10022, pp. 978–987, Mar. 2016, doi: 10.1016/S0140-6736(15)01345-8.

[111] S. Iyasu *et al.*, "Risk factors for sudden infant death syndrome among Northern Plains Indians," *J. Am. Med. Assoc.*, vol. 288, no. 21, pp. 2717–2723, Dec. 2002, doi: 10.1001/jama.288.21.2717.

[112] E. A. Jacobs *et al.*, "Fetal alcohol syndrome and alcohol-related neurodevelopmental disorders," *Pediatrics*, vol. 106, no. 2 I. American Academy of Pediatrics, pp. 358–361, Aug. 01, 2000, doi: 10.1542/peds.106.2.358.

[113] C. H. Tan, C. H. Denny, N. E. Cheal, J. E. Sniezek, and D. Kanny, "Alcohol use and binge drinking among women of childbearing age - United States, 2011-2013," Department of Health and Human Services, Sep. 2015. doi: 10.15585/mmwr.mm6437a3.

[114] L. B. Finer and M. R. Zolna, "Unintended pregnancy in the United States: Incidence and disparities, 2006," *Contraception*, vol. 84, no. 5, pp. 478–485, Nov. 2011, doi: 10.1016/j.contraception.2011.07.013.

[115] B. L. Thompson, P. Levitt, and G. D. Stanwood, "Prenatal exposure to drugs: Effects on brain development and implications for policy and education," *Nature Reviews Neuroscience*, vol. 10, no. 4. Nature Publishing Group, pp. 303–312, Apr. 11, 2009, doi: 10.1038/nrn2598.

[116] B. Silvestre de Ferron *et al.*, "Increase of KCC2 in hippocampal synaptic plasticity disturbances after perinatal ethanol exposure," *Addict. Biol.*, vol. 22, no. 6, pp. 1870–1882, Nov. 2017, doi: 10.1111/adb.12465.

[117] L. W. Role and D. K. Berg, "Nicotinic receptors in the development and modulation of CNS synapses," Sargent, 1996. doi: 10.1016/S0896-6273(00)80134-8.

[118] J. R. Duncan *et al.*, "Prenatal nicotine-exposure alters fetal autonomic activity and

medullary neurotransmitter receptors: implications for sudden infant death syndrome," *J. Appl. Physiol.*, vol. 107, no. 5, pp. 1579–1590, Nov. 2009, doi: 10.1152/japplphysiol.91629.2008.

[119] A. Delpisheh, E. Attia, S. Drammond, and B. J. Brabin, "Adolescent smoking in pregnancy and birth outcomes," *Eur. J. Public Health*, vol. 16, no. 2, pp. 168–172, 2006, doi: 10.1093/eurpub/cki219.

[120] X. Wang *et al.*, "Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight," *J. Am. Med. Assoc.*, vol. 287, no. 2, pp. 195–202, Jan. 2002, doi: 10.1001/jama.287.2.195.

[121] C. Salafia and K. Shiverick, "Cigarette smoking and pregnancy. II: Vascular effects," *Placenta*, vol. 20, no. 4, pp. 273–279, May 1999, doi: 10.1053/plac.1998.0378.

[122] P. J. H. Jones, J. Leichter, and M. Lee, "Placental blood flow in rats fed alcohol before and during gestation," *Life Sci.*, vol. 29, no. 11, pp. 1153–1159, Sep. 1981, doi: 10.1016/0024-3205(81)90204-6.

[123] A. Pastrakuljic, R. Schwartz, C. Simone, L. O. Derewlany, B. Knie, and G. Koren, "Transplacental transfer and biotransformation studies of nicotine in the human placental cotyledon perfused in vitro," *Life Sci.*, vol. 63, no. 26, pp. 2333–2342, Nov. 1998, doi: 10.1016/S0024-3205(98)00522-0.

[124] L. C. Castro, R. Allen, D. Ogunyemi, K. Roll, and L. D. Platt, "Cigarette smoking during pregnancy: Acute effects on uterine flow velocity waveforms," *Obstet. Gynecol.*, vol. 81, no. 4, pp. 551–555, 1993.

[125] W. J. Van der Velde, J. H. J. Copius Peereboom-Stegeman, P. E. Treffers, and J. James, "Structural changes in the placenta of smoking mothers: A quantitative study," *Placenta*, vol. 4, no. 3, pp. 231–240, Jul. 1983, doi: 10.1016/S0143-4004(83)80002-2.

[126] K. T. Shiverick and C. Salafia, "Cigarette smoking and pregnancy. I: Ovarian, uterine and placental effects," *Placenta*, vol. 20, no. 4, pp. 265–272, 1999, doi: 10.1053/plac.1998.0377.

[127] A. Sania *et al.*, "The K nearest neighbor algorithm for imputation of missing longitudinal prenatal alcohol data," *Under Rev. BMC Med. Res. Methodol.*, 2020.

[128] N. Pini *et al.*, "Cluster Analysis of Alcohol Consumption during Pregnancy in the Safe Passage Study," in *Proceedings of the Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society, EMBS, Jul. 2019, pp. 1338–1341, doi: 10.1109/EMBC.2019.8857428.

[129] D. A. Dawson, "Methodological Issues in Measuring Alcohol Use," *Alcohol Res. Heal.*, vol. 27, no. 1, pp. 18–29, 2003.

[130] G. I. J. Feunekes, P. Van 't Veer, W. A. Van Staveren, and F. J. Kok, "Alcohol intake assessment: The sober facts," *Am. J. Epidemiol.*, vol. 150, no. 1, pp. 105–112, Jul. 1999, doi: 10.1093/oxfordjournals.aje.a009909.

[131] A. Buu, S. Yang, R. Li, M. A. Zimmerman, R. M. Cunningham, and M. A. Walton, "Examining measurement reactivity in daily diary data on substance use: Results from a randomized experiment," *Addict. Behav.*, vol. 102, Mar. 2020, doi: 10.1016/j.addbeh.2019.106198.

[132] C. McQuire, S. Paranjothy, L. Hurt, M. Mann, D. Farewell, and A. Kemp, "Objective measures of prenatal alcohol exposure: A systematic review," *Pediatrics*, vol. 138, no. 3. American Academy of Pediatrics, Sep. 01, 2016, doi: 10.1542/peds.2016-0517.

[133] B. Simkhada, E. R. Van Teijlingen, M. Porter, and P. Simkhada, "Factors affecting the utilization of antenatal care in developing countries: Systematic review of the literature," *J. Adv. Nurs.*, vol. 61, no. 3, pp. 244–260, Feb. 2008, doi: 10.1111/j.1365-2648.2007.04532.x.

[134] K. A. Dukes *et al.*, "The safe passage study: Design, methods, recruitment, and follow-up approach," *Paediatr. Perinat. Epidemiol.*, vol. 28, no. 5, pp. 455–465, 2014, doi: 10.1111/ppe.12136.

[135] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.

[136] P. Elliott and G. Hawthorne, "Imputing Missing Repeated Measures Data: How Should We Proceed?," *Aust. New Zeal. J. Psychiatry*, vol. 39, no. 7, pp. 575–582, Jul. 2005, doi: 10.1080/j.1440-1614.2005.01629.x.

[137] A. K. Waljee *et al.*, "Comparison of imputation methods for missing laboratory data in medicine," *BMJ Open*, vol. 3, no. 8, p. e002847, Aug. 2013, doi: 10.1136/bmjopen-2013-002847.

[138] J. Brick, "Standardization of alcohol calculations in research," *Alcoholism: Clinical and Experimental Research*, vol. 30, no. 8. Alcohol Clin Exp Res, pp. 1276–1287, Aug. 2006, doi: 10.1111/j.1530-0277.2006.00155.x.

[139] R. Room *et al.*, “Times to drink: cross-cultural variations in drinking in the rhythm of the week,” *Int J Public Heal.*, vol. 57, no. 1, pp. 107–117, 2012, doi: 10.1007/s00038-011-0259-3.

[140] T. J. Grigsby and J. McLawhorn, “Missing Data Techniques and the Statistical Conclusion Validity of Survey-Based Alcohol and Drug Use Research Studies: A Review and Comment on Reproducibility,” *J. Drug Issues*, vol. 49, no. 1, pp. 44–56, Jan. 2019, doi: 10.1177/0022042618795878.

[141] K. A. Hallgren *et al.*, “Missing Data in Alcohol Clinical Trials with Binary Outcomes,” *Alcohol. Clin. Exp. Res.*, vol. 40, no. 7, pp. 1548–1557, Jul. 2016, doi: 10.1111/acer.13106.

[142] K. A. Hallgren and K. Witkiewitz, “Missing Data in Alcohol Clinical Trials: A Comparison of Methods,” *Alcohol. Clin. Exp. Res.*, vol. 37, no. 12, pp. 2152–2160, 2013, doi: 10.1111/acer.12205.

[143] U. Grittner, G. Gmel, S. Ripatti, K. Bloomfield, and M. Wicki, “Missing value imputation in longitudinal measures of alcohol consumption,” *Int. J. Methods Psychiatr. Res.*, vol. 20, no. 1, pp. 50–61, Mar. 2011, doi: 10.1002/mpr.330.

[144] L. C. R. Pessenda, C. S. Lisi, and S. E. M. Gouveia, *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1987.

[145] M. H. Huque, J. B. Carlin, J. A. Simpson, and K. J. Lee, “A comparison of multiple imputation methods for missing data in longitudinal studies 01 Mathematical Sciences,” *BMC Med. Res. Methodol.*, vol. 18, no. 1, p. 168, Dec. 2018, doi: 10.1186/s12874-018-0615-6.

[146] S. G. Liao *et al.*, “Missing value imputation in high-dimensional phenomic data: Imputable or not, and how?,” *BMC Bioinformatics*, vol. 15, no. 1, Nov. 2014, doi: 10.1186/s12859-014-0346-6.

[147] J. S. Shah, S. N. Rai, A. P. DeFilippis, B. G. Hill, A. Bhatnagar, and G. N. Brock, “Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies,” *BMC Bioinformatics*, vol. 18, no. 1, Feb. 2017, doi: 10.1186/s12859-017-1547-6.

[148] A. Jadhav, D. Pramod, and K. Ramanathan, “Comparison of Performance of Data Imputation Methods for Numeric Dataset,” *Appl. Artif. Intell.*, vol. 33, no. 10, pp. 913–933, Aug. 2019, doi: 10.1080/08839514.2019.1637138.

[149] T. Mahboob, A. Ijaz, A. Shahzad, and M. Kalsoom, “Handling Missing Values in Chronic Kidney Disease Datasets Using KNN, K-Means and K-Medoids Algorithms,” in *ICOSST 2018 - 2018 International Conference on Open Source Systems and Technologies, Proceedings*, Jan. 2019, pp. 76–81, doi: 10.1109/ICOSST.2018.8632179.

[150] M. C. P. D. Souto, P. A. Jaskowiak, and I. G. Costa, “Impact of missing data imputation methods on gene expression clustering and classification,” *BMC Bioinformatics*, vol. 16, no. 1, Feb. 2015, doi: 10.1186/s12859-015-0494-3.

[151] M. Kokla, J. Virtanen, M. Kolehmainen, J. Paananen, and K. Hanhineva, “Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: A comparative study,” *BMC Bioinformatics*, vol. 20, no. 1, p. 492, Oct. 2019, doi: 10.1186/s12859-019-3110-0.

[152] H. Schwender, “Imputing missing genotypes with weighted k nearest neighbors,” *J. Toxicol. Environ. Heal. - Part A Curr. Issues*, vol. 75, no. 8–10, pp. 438–446, Apr. 2012, doi: 10.1080/15287394.2012.674910.

[153] B. A. Bailey and R. J. Sokol, “Prenatal alcohol exposure and miscarriage, stillbirth, preterm delivery, and sudden infant death syndrome,” vol. 34, no. 1, 2011, Accessed: Aug. 01, 2020. [Online]. Available: /pmc/articles/PMC3860553/?report=abstract.

[154] N. L. Day, D. K. Wagener, and P. M. Taylor, “Measurement of substance use during pregnancy: methodologic issues,” *NIDA Res. Monogr.*, vol. 59, pp. 36–47, 1985, Accessed: Feb. 14, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/3929131.

[155] L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, “mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.,” *R J.*, vol. 8, no. 1, pp. 289–317, Aug. 2016, Accessed: Feb. 12, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/27818791.

[156] J. L. Mills, B. I. Graubard, E. E. Harley, G. G. Rhoads, and H. W. Berendes, “Maternal Alcohol Consumption and Birth Weight. How Much Drinking During Pregnancy Is Safe?,” *JAMA - J. Am. Med. Assoc.*, vol. 252, no. 14, pp. 1875–1879, 1984.

[157] J. Henderson, U. Kesmodel, and R. Gray, “Systematic review of the fetal effects of prenatal binge-drinking,” *J. Epidemiol. Community Health*, vol. 61, no. 12, pp. 1069–1073, 2007, doi: 10.1136/jech.2006.054213.

[158] P. S. Eriksen, G. Gennser, R. Lindvall, and K. Nilsson, “Acute Effects of Maternal Smoking

on Fetal Heart Beat Intervals," *Acta Obstet. Gynecol. Scand.*, vol. 63, no. 5, pp. 385–390, 1984, doi: 10.3109/00016348409156689.

[159] J. M. Barrett, J. E. Vanhooydonk, and F. H. Boehm, "Acute effect of cigarette smoking on the fetal heart rate nonstress test," *Obstet. Gynecol.*, vol. 57, no. 4, pp. 422–425, Apr. 1981.

[160] P. S. Eriksen and K. Maršál, "Circulatory changes in the fetal aorta after maternal smoking," *BJOG An Int. J. Obstet. Gynaecol.*, vol. 94, no. 4, pp. 301–305, 1987, doi: 10.1111/j.1471-0528.1987.tb03095.x.

[161] K. L. Law, L. R. Stroud, L. L. LaGasse, R. Niaura, J. Liu, and B. M. Lester, "Smoking during pregnancy and newborn neurobehavior," *Pediatrics*, vol. 111, no. 6 I, pp. 1318–1323, Jun. 2003, doi: 10.1542/peds.111.6.1318.

[162] E. D. Shenassa and M. J. Brown, "Maternal smoking and infantile gastrointestinal dysregulation: The case of colic," *Pediatrics*, vol. 114, no. 4. Pediatrics, Oct. 2004, doi: 10.1542/peds.2004-1036.

[163] L. R. Stroud *et al.*, "Maternal smoking during pregnancy and neonatal behavior: A large-scale community study," *Pediatrics*, vol. 123, no. 5, p. e842, May 2009, doi: 10.1542/peds.2008-2084.

[164] B. Cowperthwaite, S. M. J. Hains, and B. S. Kisilevsky, "Fetal behavior in smoking compared to non-smoking pregnant women," *Infant Behav. Dev.*, vol. 30, no. 3, pp. 422–430, Sep. 2007, doi: 10.1016/j.infbeh.2006.12.004.

[165] S. A. Reijneveld, C. I. Lanting, M. R. Crone, and J. P. Wouwe, "Exposure to tobacco smoke and infant crying," *Acta Paediatr.*, vol. 94, no. 2, pp. 217–221, Jan. 2007, doi: 10.1111/j.1651-2227.2005.tb01894.x.

[166] S. Lange, C. Probst, J. Rehm, and S. Popova, "National, regional, and global prevalence of smoking during pregnancy in the general population: a systematic review and meta-analysis," *Lancet Glob. Heal.*, vol. 6, no. 7, pp. e769–e776, Jul. 2018, doi: 10.1016/S2214-109X(18)30223-7.

[167] S. Beck *et al.*, "The worldwide incidence of preterm birth: A systematic review of maternal mortality and morbidity," *Bull. World Health Organ.*, vol. 88, no. 1, pp. 31–38, Jan. 2010, doi: 10.2471/BLT.08.062554.

[168] L. V. E. Simmons, C. E. Rubens, G. L. Darmstadt, and M. G. Gravett, "Preventing Preterm Birth and Neonatal Mortality: Exploring the Epidemiology, Causes, and Interventions,"

Seminars in Perinatology, vol. 34, no. 6. pp. 408–415, Dec. 2010, doi: 10.1053/j.semperi.2010.09.005.

[169] R. L. Goldenberg, J. F. Culhane, J. D. Iams, and R. Romero, "Epidemiology and causes of preterm birth," *Lancet*, vol. 371, no. 9606, pp. 75–84, Jan. 2008, doi: 10.1016/S0140-6736(08)60074-4.

[170] T. Xiong, F. Gonzalez, and D. Z. Mu, "An overview of risk factors for poor neurodevelopmental outcome associated with prematurity," *World Journal of Pediatrics*, vol. 8, no. 4. Springer, pp. 293–300, Nov. 15, 2012, doi: 10.1007/s12519-012-0372-2.

[171] R. W. Loftin, M. Habli, C. C. Snyder, C. M. Cormier, D. F. Lewis, and E. A. DeFranco, "Late Preterm Birth," *Rev. Obstet. Gynecol.*, vol. 3, no. 1, pp. 10–19, 2010, doi: 10.3909/riog0098.

[172] N. Burtchen, M. M. Myers, M. Lucchini, M. Ordonez Retamar, D. Rodriguez, and W. P. Fifer, "Autonomic signatures of late preterm, early term, and full term neonates during early postnatal life," *Early Hum. Dev.*, vol. 137, p. 104817, Oct. 2019, doi: 10.1016/j.earlhumdev.2019.06.012.

[173] J. H. Zavala, L. Ecklund-Flores, M. M. Myers, and W. P. Fifer, "Assessment of autonomic function in the late term fetus: The effects of sex and state," *Dev. Psychobiol.*, vol. 62, no. 2, pp. 224–231, Mar. 2020, doi: 10.1002/dev.21865.

[174] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J. H. Peter, and K. Voigt, "Correlated and uncorrelated regions in heart-rate fluctuations during sleep," 2000. doi: 10.1103/PhysRevLett.85.3736.

[175] P. C. Ivanov *et al.*, "Scaling behaviour of heartbeat intervals obtained by wavelet-based time- series analysis," *Nature*, vol. 383, no. 6598, pp. 323–327, 1996, doi: 10.1038/383323a0.

[176] M. Lucchini, N. Pini, W. P. Fifer, N. Burtchen, and M. G. Signorini, "Characterization of cardiorespiratory phase synchronization and directionality in late premature and full term infants," *Physiol. Meas.*, vol. 39, no. 6, p. 064001, 2018, doi: 10.1088/1361-6579/aac553.

[177] M. Lucchini, N. Pini, N. Burtchen, M. G. Signorini, and W. P. Fifer, "Transfer Entropy Modeling of Newborn Cardiorespiratory Regulation," *Front. Physiol.*, vol. 11, p. 1095, Aug. 2020, doi: 10.3389/fphys.2020.01095.

[178] N. Pini, M. Lucchini, W. P. Fifer, N. Burtchen, and M. G. Signorini, "Lagged transfer

entropy analysis to investigate cardiorespiratory regulation in newborns during sleep," in *BIOSIGNALS 2019 - 12th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 12th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2019*, 2019, pp. 139–146, doi: 10.5220/0007363301390146.

[179] N. O. Pini, M. Lucchini, W. P. Fifer, M. M. Myers, and M. G. Signorini, "Influence of prenatal alcohol and smoke exposure on neonatal vagal tone in response to head-up tilt," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, Jul. 2018, vol. 2018, pp. 5874–5877, doi: 10.1109/EMBC.2018.8513652.

[180] L. C. Shuffrey *et al.*, "Association Between Prenatal Exposure to Alcohol and Tobacco and Neonatal Brain Activity: Results From the Safe Passage Study," *JAMA Netw. open*, vol. 3, no. 5, p. e204714, May 2020, doi: 10.1001/jamanetworkopen.2020.4714.

[181] N. Pini, M. Lucchini, W. P. Fifer, M. G. Signorini, and R. Barbieri, "A Point Process Framework for the Characterization of Sleep States in Early Infancy," Oct. 2019, pp. 3645–3648, doi: 10.1109/embc.2019.8857555.

[182] M. G. Rosenblum, J. Kurths, A. Pikovsky, C. Schäfer, P. Tass, and H. H. Abel, "Synchronization in noisy systems and cardiorespiratory interaction," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 6. pp. 46–53, 1998, doi: 10.1109/51.731320.

[183] A. D. Loewy and K. M. Spyer, "Central Regulation of Autonomic Functions," *Oxford Univ. Press*, 1990, [Online]. Available: http://books.google.com.tw/books?id=Oow8w12CHwgC.

[184] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, and R. Mrowka, "Identification of coupling direction: Application to cardiorespiratory interaction," *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 65, no. 4, p. 11, 2002, doi: 10.1103/PhysRevE.65.041909.

[185] G. G. Berntson, J. T. Cacioppo, and K. S. Quigley, "Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications," *Psychophysiology*, vol. 30, no. 2, pp. 183–196, Mar. 1993, doi: 10.1111/j.1469-8986.1993.tb01731.x.

[186] S. W. Porges and E. A. Byrne, "Research methods for measurement of heart rate and

respiration," *Biol. Psychol.*, vol. 34, no. 2–3, pp. 93–130, 1992, doi: 10.1016/0301-0511(92)90012-J.

[187] T. Penzel *et al.*, "Cardiovascular and respiratory dynamics during normal and pathological sleep," *Chaos*, vol. 17, no. 1, 2007, doi: 10.1063/1.2711282.

[188] T. E. Dick *et al.*, "Cardiorespiratory coupling: Common rhythms in cardiac, sympathetic, and respiratory activities," *Prog. Brain Res.*, vol. 209, pp. 191–205, 2014, doi: 10.1016/B978-0-444-63274-6.00010-2.

[189] S. Schulz *et al.*, "Cardiovascular and cardiorespiratory coupling analyses: a review.," *Philos. Trans. A. Math. Phys. Eng. Sci.*, vol. 371, no. 1997, p. 20120191, 2013, doi: 10.1098/rsta.2012.0191.

[190] F. Yasuma and J. I. Hayano, "Respiratory Sinus Arrhythmia: Why Does the Heartbeat Synchronize with Respiratory Rhythm?," *Chest*, vol. 125, no. 2, pp. 683–690, 2004, doi: 10.1378/chest.125.2.683.

[191] A. Angelone and N. a. Coulter, "Respiratory sinus arrhythmia: a frequency dependent phenomenon," *J. Appl. Physiol.*, vol. 19, no. 3, pp. 479–82, 1964, Accessed: Mar. 28, 2017. [Online]. Available: http://jap.physiology.org/content/19/3/479.short%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/14173545.

[192] C. Schäfer, M. G. Rosenblum, H. H. Abel, and J. Kurths, "Synchronization in the human cardiorespiratory system," *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 60, no. 1, pp. 857–870, 1999, doi: 10.1103/PhysRevE.60.857.

[193] M. B. Lotrič and A. Stefanovska, "Synchronization and modulation in the human cardiorespiratory system," *Phys. A Stat. Mech. its Appl.*, vol. 283, no. 3, pp. 451–461, 2000, doi: 10.1016/S0378-4371(00)00204-1.

[194] D. Hoyer, M. G. Frasch, M. Eiselt, O. Hoyer, and U. Zwiener, "Validating phase relations between cardiac and breathing cycles during sleep: A systems theoretical approach to cardiorespiratory phase synchronizations," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 2, pp. 101–106, 2001, doi: 10.1109/51.917730.

[195] M. M. Kabir *et al.*, "Cardiorespiratory phase-coupling is reduced in patients with obstructive sleep apnea," *PLoS One*, vol. 5, no. 5, p. e10602, 2010, doi: 10.1371/journal.pone.0010602.

[196] A. Pikovsky, M. Rosenblum, and J. Kurths, “Phase synchronization in regular and chaotic systems,” *Int. J. Bifurcat. Chaos*, vol. 10, no. 10, pp. 2291–2305, 2000, doi: 10.1142/S0218127400001481.

[197] R. P. Bartsch, K. K. L. Liu, Q. D. Y. Ma, and P. C. Ivanov, “Three independent forms of cardio-respiratory coupling: Transitions across sleep stages,” *Comput. Cardiol. (2010).*, vol. 41, no. January, pp. 781–784, 2014.

[198] R. P. Bartsch, A. Y. Schumann, J. W. Kantelhardt, T. Penzel, and P. C. Ivanov, “Phase transitions in physiologic coupling,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 26, pp. 10181–10186, 2012, doi: 10.1073/pnas.1204568109.

[199] G. Schäfer, M. G. Rosenblum, J. Kurths, and H. H. Abel, “Heartbeat synchronized with ventilation [7],” *Nature*, vol. 392, no. 6673, pp. 239–240, 1998, doi: 10.1038/32567.

[200] P. Langhorst, B. Schulz, G. Schulz, M. Lambertz, and B. Krienke, “Reticular formation of the lower brainstem. A common system for cardiorespiratory and somatomotor functions: discharge patterns of neighboring neurons influenced by cardiovascular and respiratory afferents,” *J. Auton. Nerv. Syst.*, vol. 9, no. 2–3, pp. 411–432, 1983, doi: 10.1016/0165-1838(83)90005-X.

[201] M. K. Hathorn, “Respiratory sinus arrhythmia in new-born infants.,” *J. Physiol.*, vol. 385, no. 1, pp. 1–12, 1987, doi: 10.1113/jphysiol.1987.sp016480.

[202] C. Van Ravenswaaij-Arts, J. Hopman, L. Kollee, G. Stoelinga, and H. Van Geijn, “Spectral analysis of heart rate variability in spontaneously breathing very preterm infants,” *Acta Paediatr.*, vol. 83, no. 5, pp. 473–480, 1994.

[203] M. Lucchini, N. Burtchen, W. P. Fifer, and M. G. . Signorini, “Cardiorespiratory analysis in late preterm, early term, and full term infants at birth,” *under Revis.*

[204] E. Longin, T. Gerstner, T. Schaible, T. Lenz, and S. König, “Maturation of the autonomic nervous system: Differences in heart rate variability in premature vs. term infants,” *J. Perinat. Med.*, vol. 34, no. 4, pp. 303–308, 2006, doi: 10.1515/JPM.2006.058.

[205] A. Lin, K. K. L. Liu, R. P. Bartsch, and P. C. Ivanov, “Delay-correlation landscape reveals characteristic time delays of brain rhythms and heart interactions,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2067, p. 20150182, 2016, doi: 10.1098/rsta.2015.0182.

[206] T. Penzel *et al.*, “Modulations of heart rate, ECG, and cardio-respiratory coupling observed in polysomnography,” *Front. Physiol.*, vol. 7, no. OCT, p. 460, 2016, doi:

10.3389/fphys.2016.00460.

[207] M. G. Rosenblum and A. S. Pikovsky, “Detecting direction of coupling in interacting oscillators,” *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 64, no. 4, p. 4, 2001, doi: 10.1103/PhysRevE.64.045202.

[208] A. Kugelman and A. A. Colin, “Late Preterm Infants: Near Term But Still in a Critical Developmental Time Period,” *Pediatrics*, vol. 132, no. 4, pp. 31–4005, 2013.

[209] A. J. Garcia, J. E. Koschnitzky, T. Dashevskiy, and J. M. Ramirez, “Cardiorespiratory coupling in health and disease,” *Auton. Neurosci. Basic Clin.*, vol. 175, no. 1–2, pp. 26–37, 2013, doi: 10.1016/j.autneu.2013.02.006.

[210] A. J. Garcia, J. E. Koschnitzky, and J.-M. Ramirez, “The physiological determinants of Sudden Infant Death Syndrome,” *Respir. Physiol. Neurobiol.*, vol. 189, no. 2, pp. 288–300, 2013, doi: 10.1016/j.resp.2013.05.032.

[211] J. R. Isler, T. Thai, M. M. Myers, and W. P. Fifer, “An automated method for coding sleep states in human infants based on respiratory rate variability,” *Dev. Psychobiol.*, vol. 58, no. 8, pp. 1108–1115, 2016, doi: 10.1002/dev.21482.

[212] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985, doi: 10.1109/TBME.1985.325532.

[213] B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, and R. Mrowka, “Phase dynamics of coupled oscillators reconstructed from data,” *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 77, no. 6, pp. 1–16, 2008, doi: 10.1103/PhysRevE.77.066205.

[214] B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, and R. Mrowka, “Uncovering interaction of coupled oscillators from data,” *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 76, no. 5, pp. 1–4, 2007, doi: 10.1103/PhysRevE.76.055201.

[215] B. Kralemann, A. Pikovsky, and M. Rosenblum, “Reconstructing phase dynamics of oscillator networks,” *Chaos*, vol. 21, no. 2, 2011, doi: 10.1063/1.3597647.

[216] D. Cysarz, H. Bettermann, S. Lange, D. Geue, and P. van Leeuwen, “A quantitative comparison of different methods to detect cardiorespiratory coordination during night-time sleep,” *Biomed. Eng. Online*, vol. 3, no. 1, p. 44, 2004, doi: 10.1186/1475-925X-3-44.

[217] R. MROWKA, A. PATZAK, and M. ROSENBLUM, “Quantitative Analysis of Cardiorespiratory Synchronization in Infants,” *Int. J. Bifurc. Chaos*, vol. 10, no. 11, pp. 2479–2488, 2000, doi: 10.1142/S0218127400001754.

[218] R. Mrowka, L. Cimponeriu, A. Patzak, and M. G. Rosenblum, “Directionality of coupling of physiological subsystems: age-related changes of cardiorespiratory interaction during different sleep stages in babies.,” *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 285, no. 6, pp. R1395-401, 2003, doi: 10.1152/ajpregu.00373.2003.

[219] M. Lucchini, N. Burtchen, W. P. Fifer, and M. G. Signorini, “Multi-parametric cardiorespiratory analysis in late-preterm, early-term, and full-term infants at birth,” *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 99–106, Jan. 2019, doi: 10.1007/s11517-018-1866-4.

[220] M. Lucchini, N. Pini, W. P. Fifer, N. Burtchen, and M. G. Signorini, “Entropy information of cardiorespiratory dynamics in Neonates during sleep,” *Entropy*, vol. 19, no. 5, p. 225, 2017, doi: 10.3390/e19050225.

[221] M. Lucchini, N. Pini, W. P. Fifer, N. Burtchen, and M. G. Signorini, “Cardio-respiratory phase locking in newborn and one month infants as a function of sleep state,” in *IFMBE Proceedings*, 2017, vol. 65, pp. 791–794, doi: 10.1007/978-981-10-5122-7_198.

[222] R. Bartsch, J. W. Kantelhardt, T. Penzel, and S. Havlin, “Experimental evidence for phase synchronization transitions in the human cardiorespiratory system,” *Phys. Rev. Lett.*, vol. 98, no. 5, pp. 1–4, 2007, doi: 10.1103/PhysRevLett.98.054102.

[223] P. N. Sachis, D. L. Armstrong, L. E. Becker, and A. C. Bryan, “Myelination of the human vagus nerve from 24 weeks postconceptional age to adolescence,” *J. Neuropathol. Exp. Neurol.*, vol. 41, no. 4, pp. 466–472, 1982, doi: 10.1097/00005072-198207000-00009.

[224] M. T. Clark *et al.*, “Breath-by-breath analysis of cardiorespiratory interaction for quantifying developmental maturity in premature infants,” *J. Appl. Physiol.*, vol. 112, no. 5, pp. 859–867, 2012, doi: 10.1152/japplphysiol.01152.2011.

[225] C. K. Shapiro-Mendoza *et al.*, “Effect of late-preterm birth and maternal medical conditions on newborn morbidity risk,” *Pediatrics*, vol. 121, no. 2, pp. e223–e232, 2008, doi: 10.1542/peds.2006-3629.

[226] P. C. H. Ivanov, K. K. L. Liu, and R. P. Bartsch, “Focus on the emerging new fields of network physiology and network medicine,” *New J. Phys.*, vol. 18, no. 10, p. 100201, 2016, doi: 10.1088/1367-2630/18/10/100201.

[227] L. Liu *et al.*, “Global , regional , and national causes of under-5 mortality in 2000 – 15 : an updated systematic analysis with implications for the Sustainable Development Goals,”

Lancet, vol. 388, pp. 3027–3035, 2016, doi: 10.1016/S0140-6736(16)31593-8.

[228] March of Dimes, "2018 Premature Birth Report Card United States," Arlington, VA 22202, USA, 2019.

[229] M. Wang, D. Dorer, M. Fleming, and E. Catlin, "Clinical outcomes of near-term infants.," *Pediatrics*, vol. 114, no. 2 PG-372–6, pp. 372–376, 2004, doi: 114/2/372 [pii].

[230] J. L. Richards *et al.*, "Temporal Trends in Late Preterm and Early Term Birth Rates in 6 High-Income Countries in North America and Europe and Association With Clinician-Initiated Obstetric Interventions," *Jama*, vol. 316, no. 4, pp. 410–419, 2016, doi: 10.1001/jama.2016.9635.

[231] J. M. D. Thompson and E. A. Mitchell, "Are the risk factors for SIDS different for preterm and term infants?," *Arch. Dis. Child.*, vol. 91, no. 2, pp. 107–111, 2006, doi: 10.1136/adc.2004.071167.

[232] D. H. Adamkin, "Postnatal Glucose Homeostasis in Late-Preterm and Term Infants Clinical Report—Postnatal Glucose Homeostasis in Late-Preterm and Term Infants," *Pediatrics*, vol. 127, no. 3, pp. 575–579, 2013, doi: 10.1542/peds.2010-3851.

[233] C. McEvoy, S. Venigalla, D. Schilling, N. Clay, P. Spitale, and T. Nguyen, "Respiratory function in healthy late preterm infants delivered at 33-36 weeks of gestation," *J. Pediatr.*, vol. 162, no. 3, pp. 464–469, 2013, doi: 10.1016/j.jpeds.2012.09.042.

[234] C. E. Hunt, "Ontogeny of Autonomic Regulation in Late Preterm Infants Born at 34-37 Weeks Postmenstrual Age," *Semin. Perinatol.*, vol. 30, no. 2, pp. 73–76, 2006, doi: 10.1053/j.semperi.2006.02.005.

[235] M. S. Scher *et al.*, "Physiologic Brain Dysmaturity in Late Preterm Infants," *Pediatr. Res.*, vol. 70, no. 5, pp. 524–528, 2011.

[236] P. C. Ivanov and R. P. Bartsch, "Network Physiology: Mapping Interactions Between Networks of Physiologic Networks," in *Networks of Networks: the last Frontier of Complexity.*, no. February, Springer, 2014, pp. 203–222.

[237] R. M. Harper, R. C. Frysinger, J. Zhang, R. B. Trelease, and R. R. Terreberry, "Cardiac and Respiratory Interactions Maintaining Homeostasis During Sleep," in *Clinical Physiology of Sleep*, New York, NY: Springer New York, 1988, pp. 67–78.

[238] R. M. Harper, R. Kumar, P. M. Macey, J. A. Ogren, and H. L. Richardson, "Functional Neuroanatomy and Sleep-Disordered Breathing: Implications for Autonomic Regulation,"

Anat. Rec., vol. 295, no. 9, pp. 1385–1395, 2012, doi: 10.1002/ar.22514.

[239] V. L. Schechtman, R. M. Harper, A. J. Wilson, and D. P. Shouthall, "Sleep-Apnea in Infants Who Succumb To the Sudden-Infant-Death-Syndrome," *Pediatrics*, vol. 87, no. 6, pp. 841–846, 1991.

[240] R. S. C. Horne, "Cardio-respiratory control during sleep in infancy," *Paediatr. Respir. Rev.*, vol. 15, no. 2, pp. 163–169, 2014, doi: 10.1016/j.prrv.2013.02.012.

[241] M. G. Frasch, U. Zwiener, D. Hoyer, and M. Eiselt, "Autonomic organization of respirocardial function in healthy human neonates in quiet and active sleep," *Early Hum. Dev.*, vol. 83, no. 4, pp. 269–277, 2007, doi: 10.1016/j.earlhumdev.2006.05.023.

[242] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000, doi: 10.1103/PhysRevLett.85.461.

[243] L. Faes, D. Marinazzo, A. Montalto, and G. Nollo, "Lag-specific transfer entropy as a tool to assess cardiovascular and cardiorespiratory information transfer," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2556–2568, 2014, doi: 10.1109/TBME.2014.2323131.

[244] D. Kugiumtzis, "Direct-coupling information measure from nonuniform embedding," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 87, no. 6, 2013, doi: 10.1103/PhysRevE.87.062918.

[245] M. Stefanski *et al.*, "A scoring system for states of sleep and wakefulness in term and preterm infants," *Pediatric research*, vol. 18, no. 1. pp. 58–62, 1984, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/6701035.

[246] W. Xiong, L. Faes, and P. C. Ivanov, "Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations," *Phys. Rev. E*, vol. 95, no. 6, p. 062114, Jun. 2017, doi: 10.1103/PhysRevE.95.062114.

[247] A. Montalto, L. Faes, and D. Marinazzo, "MuTE: A MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy," *PLoS One*, vol. 9, no. 10, 2014, doi: 10.1371/journal.pone.0109462.

[248] D. Hoyer, B. Pompe, K. H. Chon, H. Hardraht, C. Wicher, and U. Zwiener, "Mutual information function assesses autonomic information flow of heart rate dynamics at different time scales," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 584–592, 2005, doi: 10.1109/TBME.2005.844023.

[249] R. D. Brook and S. Julius, “Autonomic Imbalance, Hypertension, and Cardiovascular Risk,” *Am. J. Hypertens.*, vol. 13, no. S4, pp. 112s-122s, 2000.

[250] T. Nakamura, K. Kiyono, H. Wendt, P. Abry, and Y. Yamamoto, “Multiscale Analysis of Intensive Longitudinal Biomedical Signals and Its Clinical Applications,” *Proc. IEEE*, vol. 104, no. 2, pp. 242–261, 2016.

[251] M. Lucchini *et al.*, “Feasibility study for the assessment of cardio-respiratory coupling in newborn infants,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016, vol. 2016-Octob, pp. 5509–5512, doi: 10.1109/EMBC.2016.7591974.

[252] X. Xiao, T. J. Mullen, and R. Mukkamala, “System identification: A multi-signal approach for probing neural cardiovascular regulation,” *Physiol. Meas.*, vol. 26, no. 3, pp. R41–R71, Jun. 2005, doi: 10.1088/0967-3334/26/3/R01.

[253] R. F. Carlin and R. Y. Moon, “Risk factors, protective factors, and current recommendations to reduce sudden infant death syndrome a review,” *JAMA Pediatr.*, vol. 171, no. 2, pp. 175–180, Feb. 2017, doi: 10.1001/jamapediatrics.2016.3345.

[254] R. M. Harper and H. C. Kinney, “Potential Mechanisms of Failure in the Sudden Infant Death Syndrome.,” *Curr. Pediatr. Rev.*, vol. 6, no. 1, pp. 39–47, 2010, doi: 10.2174/157339610791317214.

[255] N. Montano, T. G. Ruscone, A. Porta, F. Lombardi, M. Pagani, and A. Malliani, “Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt.,” *Circulation*, vol. 90, no. 4, pp. 1826–31, Oct. 1994, doi: 10.1161/01.CIR.90.4.1826.

[256] P. Graziano and K. Derefinko, “Cardiac vagal control and children’s adaptive functioning: a meta-analysis.,” *Biol. Psychol.*, vol. 94, no. 1, pp. 22–37, Sep. 2013, doi: 10.1016/j.biopsycho.2013.04.011.

[257] M. M. Myers *et al.*, “Cardiorespiratory physiology in the safe passage study: protocol, methods and normative values in unexposed infants,” *Acta Paediatr. Int. J. Paediatr.*, vol. 106, no. 8, pp. 1260–1272, 2017, doi: 10.1111/apa.13873.

[258] W. G. Guntheroth and P. S. Spiers, “The Triple Risk Hypotheses in Sudden Infant Death Syndrome,” *Pediatrics*, vol. 110, no. 5, pp. e64–e64, 2002, doi: 10.1542/peds.110.5.e64.

[259] D. E. Lake, “Renyi entropy measures of heart rate Gaussianity,” *IEEE Trans. Biomed. Eng.*,

vol. 53, no. 1, pp. 21–27, 2006, doi: 10.1109/TBME.2005.859782.

[260] J. Penttilä *et al.*, “Effect of cardiac vagal outflow on complexity and fractal correlation properties of heart rate dynamics,” *Auton. Autacoid Pharmacol.*, vol. 23, no. 3, pp. 173–179, 2003, doi: 10.1046/j.1474-8673.2003.00293.x.

[261] S. W. Porges, J. A. Doussard-Roosevelt, A. L. Portales, and S. I. Greenspan, “Infant regulation of the vagal ‘brake’ predicts child behavior problems: A psychobiological model of social behavior,” *Dev. Psychobiol.*, vol. 29, no. 8, pp. 697–712, Dec. 1996, doi: 10.1002/(SICI)1098-2302(199612)29:8<697::AID-DEV5>3.0.CO;2-O.

[262] A. Edner, M. Katz-Salamon, H. Lagercrantz, and J. Milerad, “Heart rate response profiles during head upright tilt test in infants with apparent life threatening events,” *Arch Dis Child*, vol. 76, pp. 27–30, 1997, Accessed: Jan. 31, 2018. [Online]. Available: http://adc.bmj.com/content/archdischild/76/1/27.full.pdf.

[263] B. R. H. Van den Bergh *et al.*, “Prenatal developmental origins of behavior and mental health: The influence of maternal stress in pregnancy,” *Neurosci. Biobehav. Rev.*, Jul. 2017, doi: 10.1016/j.neubiorev.2017.07.003.

[264] C. Buss, S. Entringer, and P. D. Wadhwa, “Fetal programming of brain development: Intrauterine stress and susceptibility to psychopathology,” in *Science Signaling*, Oct. 2012, vol. 5, no. 245, pp. pt7–pt7, doi: 10.1126/scisignal.2003406.

[265] G. Z. Tau and B. S. Peterson, “Normal development of brain circuits,” *Neuropsychopharmacology*, vol. 35, no. 1. Nature Publishing Group, pp. 147–168, Jan. 30, 2010, doi: 10.1038/npp.2009.115.

[266] C. Monk, C. Lugo-Candelas, and C. Trumpff, “Prenatal Developmental Origins of Future Psychopathology: Mechanisms and Pathways,” *Annu. Rev. Clin. Psychol.*, vol. 15, no. 1, pp. 317–344, May 2019, doi: 10.1146/annurev-clinpsy-050718-095539.

[267] K. Walsh *et al.*, “Maternal prenatal stress phenotypes associate with fetal neurodevelopment and birth outcomes,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 48, pp. 23996–24005, Nov. 2019, doi: 10.1073/pnas.1905890116.

[268] N. K. Moog *et al.*, “Intergenerational Effect of Maternal Exposure to Childhood Maltreatment on Newborn Brain Anatomy,” *Biol. Psychiatry*, vol. 83, no. 2, pp. 120–127, Jan. 2018, doi: 10.1016/j.biopsych.2017.07.009.

[269] M. Matoušek and I. Petersén, “Automatic evaluation of EEG background activity by means

of age-dependent EEG quotients," *Electroencephalogr. Clin. Neurophysiol.*, vol. 35, no. 6, pp. 603–612, Dec. 1973, doi: 10.1016/0013-4694(73)90213-7.

[270] A. R. Clarke, R. J. Barry, R. McCarthy, and M. Selikowitz, "Age and sex effects in the EEG: Development of the normal child," *Clin. Neurophysiol.*, vol. 112, no. 5, pp. 806–814, May 2001, doi: 10.1016/S1388-2457(01)00488-6.

[271] R. J. M. Somsen, B. J. Van't Klooster, M. W. Van Der Molen, H. M. P. Van Leeuwen, and R. Licht, "Growth spurts in brain maturation during middle childhood as indexed by EEG power spectra," *Biol. Psychol.*, vol. 44, no. 3, pp. 187–209, Jan. 1997, doi: 10.1016/S0301-0511(96)05218-0.

[272] R. E. Dustman, D. E. Shearer, and R. Y. Emmerson, "Life-span changes in EEG spectral amplitude, amplitude variability and mean frequency," *Clin. Neurophysiol.*, vol. 110, no. 8, pp. 1399–1409, Aug. 1999, doi: 10.1016/S1388-2457(99)00102-9.

[273] P. J. Marshall, Y. Bar-Haim, and N. A. Fox, "Development of the EEG from 5 months to 4 years of age," *Clin. Neurophysiol.*, vol. 113, no. 8, pp. 1199–1208, Aug. 2002, doi: 10.1016/S1388-2457(02)00163-3.

[274] L. Tarokh, M. A. Carskadon, and P. Achermann, "Developmental changes in brain connectivity assessed using the sleep EEG," *Neuroscience*, vol. 171, no. 2, pp. 622–634, Dec. 2010, doi: 10.1016/j.neuroscience.2010.08.071.

[275] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision trees: An overview and their use in medicine," *Journal of Medical Systems*, vol. 26, no. 5. Springer, pp. 445–463, Oct. 2002, doi: 10.1023/A:1016409317640.

[276] J. H. Friedman, "Stochastic gradient boosting," *Comput. Stat. Data Anal.*, vol. 38, no. 4, pp. 367–378, Feb. 2002, doi: 10.1016/S0167-9473(01)00065-2.

[277] J. L. Cox, G. Chapman, D. Murray, and P. Jones, "Validation of the Edinburgh postnatal depression scale (EPDS) in non- postnatal women," *J. Affect. Disord.*, vol. 39, no. 3, pp. 185–189, Jul. 1996, doi: 10.1016/0165-0327(96)00008-0.

[278] R. Alvarado, E. Jadresic, V. Guajardo, and G. Rojas, "First validation of a Spanish-translated version of the Edinburgh postnatal depression scale (EPDS) for use in pregnant women. A Chilean study," *Arch. Womens. Ment. Health*, vol. 18, no. 4, pp. 607–612, Aug. 2015, doi: 10.1007/s00737-014-0466-z.

[279] C. P. Press, *Manual for the State-Trait Anxiety Inventory: STAI*. Palo Alto, CA, USA, 1983.

[280] N. Bayley, *Bayley scales of infant and toddler development*. San Antonio, Texas, USA, 2006.

[281] B. Gowan, M. J. Carter, and A. S. Carter, *ITSEA/BITSEA: Infant-Toddler and Brief Infant-Toddler Social and Emotional Assessment*. The Psychological Corporation, 2006.

[282] I. Kruizinga, J. C. Visser, T. van Batenburg-Eddes, A. S. Carter, W. Jansen, and H. Raat, "Screening for Autism Spectrum Disorders with the Brief Infant-Toddler Social and Emotional Assessment," *PLoS One*, vol. 9, no. 5, p. e97630, May 2014, doi: 10.1371/journal.pone.0097630.

[283] D. L. Robins, K. Casagrande, M. Barton, C. M. A. Chen, T. Dumont-Mathieu, and D. Fein, "Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F)," *Pediatrics*, vol. 133, no. 1, pp. 37–45, Jan. 2014, doi: 10.1542/peds.2013-1813.

[284] D. L. Robins and T. M. Dumont-Mathieu, "Early screening for autism spectrum disorders: Update on the modified checklist for autism in toddlers and other measures," *J. Dev. Behav. Pediatr.*, vol. 27, no. 2 SUPPL. 2, pp. S111–S119, 2006, doi: 10.1097/00004703-200604002-00009.

[285] D. L. Robins, D. Fein, M. L. Barton, and J. A. Green, "The Modified Checklist for Autism in Toddlers: An Initial Study Investigating the Early Detection of Autism and Pervasive Developmental Disorders," *J. Autism Dev. Disord.*, vol. 31, no. 2, pp. 131–144, 2001, doi: 10.1023/A:1010738829569.

[286] M. Poletti, E. Carretta, L. Bonvicini, and P. Giorgi-Rossi, "Cognitive Clusters in Specific Learning Disorder," *J. Learn. Disabil.*, vol. 51, no. 1, pp. 32–42, Jan. 2018, doi: 10.1177/0022219416678407.

[287] G. Brock, V. Pihur, S. Datta, and S. Datta, "ClValid: An R package for cluster validation," *J. Stat. Softw.*, vol. 25, no. 4, pp. 1–22, Mar. 2008, doi: 10.18637/jss.v025.i04.

[288] C. Strobl, J. Malley, and G. Tutz, "An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests," *Psychol. Methods*, vol. 14, no. 4, pp. 323–348, 2009, doi: 10.1037/a0016973.

[289] A. R. Levin, K. J. Varcin, H. M. O'Leary, H. Tager-Flusberg, and C. A. Nelson, "EEG power at 3 months in infants at high familial risk for autism," *J. Neurodev. Disord.*, vol. 9,

no. 1, p. 34, Sep. 2017, doi: 10.1186/s11689-017-9214-9.

[290] A. L. Tierney, L. Gabard-Durnam, V. Vogel-Farley, H. Tager-Flusberg, and C. A. Nelson, "Developmental trajectories of resting eeg power: An endophenotype of autism spectrum disorder," *PLoS One*, vol. 7, no. 6, p. e39127, Jun. 2012, doi: 10.1371/journal.pone.0039127.

[291] E. V. Orekhova *et al.*, "EEG hyper-connectivity in high-risk infants is associated with later autism," *J. Neurodev. Disord.*, vol. 6, no. 1, p. 40, Nov. 2014, doi: 10.1186/1866-1955-6-40.

[292] L. J. Gabard-Durnam *et al.*, "Longitudinal EEG power in the first postnatal year differentiates autism outcomes," *Nat. Commun.*, vol. 10, no. 1, p. 4188, Dec. 2019, doi: 10.1038/s41467-019-12202-9.

[293] N. H. Brito *et al.*, "Neonatal EEG linked to individual differences in socioemotional outcomes and autism risk in toddlers," *Dev. Psychobiol.*, vol. 61, no. 8, pp. 1110–1119, Dec. 2019, doi: 10.1002/dev.21870.

[294] S. Madigan *et al.*, "A Meta-Analysis of Maternal Prenatal Depression and Anxiety on Child Socioemotional Development," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 57, no. 9. Elsevier Inc., pp. 645-657.e8, Sep. 01, 2018, doi: 10.1016/j.jaac.2018.06.012.