

Inaugural-Dissertation

zur
Erlangung der Doktorwürde
der
Gesamtfakultät für Mathematik, Ingenieur- und
Naturwissenschaften
der
Ruprecht-Karls-Universität
Heidelberg

vorgelegt von
Elaine Serena Zaunseder
aus Trier

Tag der mündlichen Prüfung:

New approaches in mathematical and data-based modeling for newborn screening

Betreuer: Prof. Dr. Vincent Heuveline
Zweitbetreuer: Prof. Dr. Stefan Kölker

ABSTRACT

Newborn screening aims to detect rare, inherited metabolic diseases early in newborns. The pre-symptomatic diagnosis of these diseases, enabling effective therapies, can enhance the quality of life for affected children and their families. However, newborn screening for specific diseases faces challenges, such as false-positive screening results and a need for individualized disease management due to the substantial clinical variability of inherited metabolic diseases. This dissertation aims to develop new mathematical and data-based modeling approaches to support and improve newborn screening. Therefore, data-based machine learning models using data from over two million newborns are developed to enhance diagnostic accuracy in newborn screening for isovaleric aciduria and glutaric aciduria type 1. Furthermore, in the course of this dissertation a genome-based, mathematical whole-body model is developed to depict the metabolism of newborns and infants. This model enables personalized *in silico* simulations of newborns with an inherited metabolic disease, thereby facilitating individualized disease management. The proposed mathematical and data-based models hold considerable promise for application in personalized medicine and newborn screening, contributing to their improvement and support.

ZUSAMMENFASSUNG

Im Rahmen des Neugeborenen Screenings werden Neugeborene frühzeitig auf seltene Stoffwechselkrankheiten untersucht. Die präsymptomatische Diagnose dieser Krankheiten kann in Verbindung mit wirksamen Therapien die Lebensqualität der betroffenen Kinder und ihrer Familien verbessern. Das Neugeborenen Screening steht jedoch vor den Herausforderungen falsch-positiver Screening-Ergebnisse sowie der Notwendigkeit von individualisierten Behandlungsstrategien aufgrund der großen Variabilität der Stoffwechselkrankheiten. Ziel dieser Dissertation ist die Entwicklung neuer mathematischer und datenbasierter Modellierungsansätze zur Unterstützung und Verbesserung des Neugeborenen Screenings. Daher werden datenbasierte Modelle der künstlichen Intelligenz erarbeitet, die Daten von mehr als zwei Millionen Neugeborenen nutzen, um die diagnostische Genauigkeit beim Neugeborenen Screening auf Isovalerianazidurie und Glutarazidurie Typ 1 zu verbessern. Des Weiteren wird im Rahmen dieser Dissertation ein genombasiertes, mathematisches Ganzkörpermodell entwickelt, das den Stoffwechsel von Neugeborenen und Säuglingen abbildet. Dies ermöglicht personalisierte *in silico* Simulationen von betroffenen Neugeborenen und vereinfacht somit die gezielte Entwicklung von individualisierten Behandlungsstrategien. Die entwickelten mathematischen und datenbasierten Modelle zeigen vielversprechende Anwendungsmöglichkeiten im Bereich der personalisierten Medizin und des Neugeborenen Screenings auf, zu dessen Verbesserung und Unterstützung sie beitragen.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my gratitude to numerous people who have supported me over the past years.

First and foremost, I extend my heartfelt thanks to Prof. Dr. Vincent Heuveline for his exceptional support as my supervisor. Upon encountering him at the assessment center for the Helmholtz Information and Data Science School for Health (HIDSS4Health), I was thrilled to start my Ph.D. journey at the Engineering Mathematics and Computing Lab (EMCL) at Heidelberg University. He has been a consistent source of support throughout my research endeavors, equipping me with the skills and confidence to navigate my academic path. His scientific guidance and forward-thinking approach have been crucial in shaping my work on mathematical modeling in newborn screening.

I would also like to thank my second supervisor, Prof. Dr. Stefan Kölker, whose invaluable contributions made this interdisciplinary Ph.D. possible. He helped initiating my research project and opened the door to newborn screening and metabolic modeling. His willingness to engage in endless discussions and provide insightful knowledge, along with his valuable feedback, has been instrumental in shaping my research journey. His ability to ask the right questions has significantly enhanced the quality of my work.

Thanks to the encouragement and support of my two supervisors, I had the privilege of undertaking a research stay at the University of Galway, Ireland. During this time, I had the honor of working in the lab of Prof. Dr. Ines Thiele, who offered invaluable and continuous guidance for my research. I am deeply grateful to Ines and her exceptional team for their warm welcome in the wet west of Ireland. This opportunity was immensely enriching, and over the course of three months, I gained a lot of knowledge and experiences that have profoundly influenced both my personal growth and my research.

I would like to further thank my former and current colleagues at EMCL for contributing to the wonderful work environment of these past years. Countless special memories have been created through our lively discussions, ranging from the absurd to the profound, whether it be at the coffee corner, over lunch or at the hard-fought spikeball and badminton matches. A special thank you is reserved for Lydia Cyriax, whose support has been a constant source of strength. Regardless of the circumstances, seeking guidance from Lydia has proven to be immensely helpful; knocking on her door and having a brief chat with her not only resolves issues but also uplifts one's spirits significantly. The cornerstone of an interdisciplinary Ph.D. lies in receiving guidance from multiple disciplines. Thus, I express my heartfelt gratitude to Dr. Ulrike Mütze, Dr. Sven Garbade, Prof. Dr. Georg Hoffmann, and all their colleagues from the Division of Pediatric Neurology and Metabolic Medicine at the Medical Faculty of Heidelberg for their support throughout my research. Their invaluable advice, willingness to address all my questions, and facilitation of data sharing have been crucial in the success of my research work.

Moreover, I would like to thank HIDSS4Health along with its current and former coordinators, for their support throughout my research journey. Their assistance enabled my participation in numerous workshops and scientific courses, and their funding facilitated my travel to conferences and research stays. I am profoundly grateful for the opportunity to engage with such a diverse and inspiring community of doctoral researchers operating at the intersection of data science and healthcare. Thank you for all the good times at the retreats, regular tables, and networking

events, I had so much fun! I would also like to thank my colleagues at the Heidelberg Institute for Theoretical Studies (HITS), who welcomed me as a visiting scientist during my Ph.D and allowed me to broaden my horizon by attending seminars, workshops, and talks from different disciplines. Furthermore, I would like to thank three exceptional women who accompanied me during these four years as part of an E-Team (The 'E' stands for 'Erfolg!'). Thank you Julia, Hannah, and Frauke for always being there, for the intriguing discussions and the emotional support, especially at the start of my Ph.D.

And last but not least, I would like to thank all my friends and family for their continuous support and love. Old and new friends who made the past years to such a great time, from playing volleyball, hiking, going to the Sauna or traveling the world. I would like to thank my parents, siblings, and extended family for shaping the person I have become. Saving the most heartfelt thanks for the end, I want to extend my deepest appreciation to Hendrik. You are, without a doubt, the most incredible husband in the world, and none of this journey would have been possible without you by my side. Thank you for your advice and guidance, thank you for always being there for me, for believing and encouraging me, for lending me your ear and being my rock through it all.

PREFACE

This Ph.D. work has led to the following articles that were published throughout the Ph.D. period in scientific journals.

Peer-reviewed publications

Elaine Zaunseder, Ulrike Mütze, Jürgen G. Okun, Georg F. Hoffmann, Stefan Kölker, Vincent Heuveline, Ines Thiele, Personalised metabolic whole-body models for newborns and infants predict growth and biomarkers of inherited metabolic diseases, *Cell Metabolism*, 2024, <https://doi.org/10.1016/j.cmet.2024.05.006>.

Elaine Zaunseder, Ulrike Mütze, Sven F. Garbade, Saskia Haupt, Stefan Kölker, Vincent Heuveline, Deep learning and explainable artificial intelligence for improving specificity and detecting metabolic patterns in newborn screening. *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, Mexico City, Mexico, 2023, pp. 1566-1571. <https://doi.org/10.1109/SSCI52147.2023.10371991>.

Elaine Zaunseder, Ulrike Mütze, Sven F. Garbade, Saskia Haupt, Patrik Feyh, Georg F. Hoffmann, Vincent Heuveline, Stefan Kölker, Machine Learning Methods Improve Specificity in Newborn Screening for Isovaleric Aciduria. *Metabolites* 2023, 13, 304. <https://doi.org/10.3390/metabo13020304>.

Elaine Zaunseder, Saskia Haupt, Ulrike Mütze, Sven F. Garbade, Stefan Kölker, Vincent Heuveline: Opportunities and Challenges in Machine Learning-based Newborn Screening - A systematic literature review. *JIMD Reports*, 2022. <https://doi.org/10.1002/jmd2.12285>.

Preprints

Elaine Zaunseder, Ulrike Mütze, Jürgen G. Okun, Georg F. Hoffmann, Stefan Kölker, Vincent Heuveline, Ines Thiele, Personalised metabolic whole-body models for newborns and infants predict growth and biomarkers of inherited metabolic diseases. *bioRxiv*, 2023, <https://doi.org/10.1101/2023.10.20.563364>.

This dissertation reuses some of these articles verbatim, indicating such use at the beginning of the corresponding sections.

Contents

1	Introduction	1
1.1	Background	2
1.2	Current research challenges	4
1.3	Main goals of this work	6
1.4	Main contributions of this work	8
1.5	Motivation and research impact	8
1.6	Thesis structure	9
2	State-of-the-art	11
2.1	Medical background on newborn screening	11
2.1.1	Newborn screening	11
2.1.2	Human metabolism	12
2.1.3	Inherited metabolic diseases	13
2.1.4	Isovaleric aciduria (IVA)	15
2.1.5	Glutaric aciduria type 1 (GA1)	16
2.2	Data-based modeling in newborn screening	18
2.2.1	Methodology of systematic literature review	18
2.2.2	Literature review for data-based models in newborn screening	20
2.3	Metabolic whole-body modeling	26
2.3.1	Metabolic modeling for human metabolism	26
2.3.2	Constrained-based modeling	28
2.3.3	Modeling of inherited metabolic diseases	29
2.3.4	Time-dependent metabolic modeling	31
3	Data-based modeling in newborn screening	33
3.1	Methodological background	34
3.1.1	Data analysis methods	35
3.1.2	Machine learning classification methods	35
3.1.3	Experimental setup	40

3.2	Data-based newborn screening for IVA	42
3.2.1	Newborn screening data set – IVA	42
3.2.2	Data analysis	44
3.2.3	Machine learning for IVA classification	47
3.2.4	Machine learning for IVA severity	50
3.2.5	Proof of concept on IVA test data set	51
3.3	Data-based newborn screening for GA1	55
3.3.1	Newborn screening data set – GA1	55
3.3.2	Data analysis	56
3.3.3	Machine learning for GA1 classification	58
3.3.4	Proof of concept on GA1 test data set	61
3.4	Summary and outcome	63
3.4.1	Digital-tier strategy in newborn screening	63
3.4.2	Method validation	64
4	Explainable AI (XAI) in newborn screening	65
4.1	Methodological background	66
4.1.1	Ensemble and deep learning methods	66
4.1.2	Explainable artificial intelligence methods	69
4.2	XAI in newborn screening	72
4.2.1	Data set and experimental setup	72
4.2.2	Ensemble and deep learning for IVA classification	73
4.2.3	LIME method for pattern recognition	74
4.2.4	SHAP method for interpretability	76
4.3	Summary and outcome	79
4.3.1	Ensemble and deep learning in newborn screening for IVA	79
4.3.2	XAI in newborn screening	79
5	Whole-body modeling for infant metabolism	83
5.1	Methodological background	83
5.1.1	Whole-body models	84
5.1.2	Flux balance analysis	84
5.1.3	Quadratic flux balance analysis	86
5.1.4	Biomass objective function	87

5.1.5	Inherited metabolic diseases analysis	87
5.2	Development of whole-body models for infants (infant-WBM)	89
5.2.1	Organ-specific biomass growth reactions	90
5.2.2	Whole-body biomass reaction, organ weights, and blood flow	90
5.2.3	Physiological parameters	91
5.2.4	Water balance in infants	92
5.2.5	Energy balance in infants	93
5.2.6	Integration of newborn screening data	98
5.2.7	Model assembly	99
5.3	Infant-WBM - Model validation	101
5.3.1	Model growth prediction	101
5.3.2	Water excretion flux prediction	103
5.3.3	ATP synthase flux prediction	104
5.4	Infant-WBM - Model application	106
5.4.1	Personalization with newborn screening data	106
5.4.2	Biomarker prediction for inherited metabolic diseases	108
5.4.3	Dietary treatment effect prediction	110
5.5	Summary and outcome	113
5.5.1	Modeling choices	113
5.5.2	Model limitations and future work	114
6	Time-dependent modeling for infant-WBMs	117
6.1	Methodological background	117
6.1.1	Pharmacokinetic modeling	118
6.1.2	Physiological-based pharmacokinetic modeling (PBPK)	118
6.1.3	Uncertainty quantification	123
6.2	Development of time-dependent infant-WBMs	125
6.2.1	Coupling of PBPK modeling and infant-WBMs	125
6.2.2	Assembly of time-dependent infant-WBMs	129
6.2.3	Biomarker prediction with time-dependent infant-WBMs	131
6.3	Uncertainty quantification for time-dependent infant-WBMs	135
6.3.1	Monte Carlo simulations for uncertain model parameters	135
6.3.2	Evaluation of uncertain parameters	139

6.4	Summary and outcome	140
6.4.1	Modeling choices	140
6.4.2	Model limitations and future work	141
7	Singular value decomposition for infant-WBMs	143
7.1	Methodological background	143
7.1.1	Singular value decomposition (SVD)	143
7.1.2	SVD matrix approximation	145
7.1.3	Computation of SVD	146
7.2	SVD of whole-body models	147
7.2.1	SVD of infant and adult whole-body models	147
7.2.2	Mode-by-mode analysis	148
7.2.3	Preliminary work on model reduction with SVD	151
7.3	Summary and outcome	154
8	Discussion and future works	157
8.1	Main findings and limitations	158
8.2	Implications and future work	163
8.2.1	Data-based modeling in newborn screening	163
8.2.2	Metabolic whole-body modeling for infants	164
8.3	Conclusion	165
A	List of Abbreviations	167
B	Supplementary tables	169
	Bibliography	173

1

Introduction

”The essence of math is not to make simple things complicated, but to make complicated things simple.”

Stanley Gudder

In a chaotic and dynamic world, the ability to break down complex systems into simpler ones is key to comprehending reality. As Stanley Gudder states, mathematical models can enable this process by simplifying complex elements in the world, thus generating sufficient value for a given application. In the medical domain, mathematical models have been helpful in our understanding of the human body’s biochemical and dynamic processes related to diseases and drugs. However, in pediatric metabolic research which investigates the complex system of metabolism, there are specific challenges in the context of metabolic diseases. Mathematically modeling such a complex system would enable a detailed causal analysis of different processes, particularly regarding the metabolic diseases examined in newborn screening. Most of these diseases are classified as rare inherited metabolic diseases (IMDs) known to be severe and life-threatening without treatment. Fortunately, newborn screening enables an early diagnosis of some of these diseases, while the subsequent treatment can potentially decrease morbidity and mortality. However, there are still many open questions about the impact of different metabolites on infant metabolism and associated disease diagnosis. This research work aims at developing new mathematical and data-based models to support newborn screening by improving diagnostic accuracy and enabling a systematic analysis of the infant metabolism and its disease processes. This chapter introduces the research by contextualizing its relevance and motivation, specifying research goals, and discussing current research challenges while providing an overview of the different sections of the dissertation as a whole.

1.1 Background

More than 50 years ago, newborn screening programs started worldwide. These programs aim to identify newborns with treatable severe rare diseases early, ideally pre-symptomatic. Therefore, blood samples from newborns are collected on the first days of life (i.e., in Germany at 36-72 h of life) and sent to a newborn screening center for analysis [1]. Early identification allows the early start of treatment of affected newborns to reduce morbidity and mortality, making them highly successful instruments of secondary prevention [2]. In the last 50 years, the panel of screened conditions has been growing, including 19 screened conditions in Germany in 2022 [3]. Two of these metabolic diseases are isovaleric aciduria (IVA) and glutaric aciduria (GA1), which belong to the group of organic acidurias leading to severe life-threatening (neonatal) metabolic compensations if untreated. IVA and GA1 are both rare diseases, with IVA having a prevalence of $\sim 1:90,000 - 100,000$ newborns [4], [5] and GA1 having a prevalence of $1:140,000$ newborns in Germany [3]. For all newborn screening diseases, the prompt evaluation of newborns with a suspected diagnosis is time-sensitive, as some treatments need to be initiated as soon as possible to prevent long-term harm [3]. Besides correctly identifying all newborns suffering from a condition, a low number of false-positive screening results is essential. However, only about one out of five newborns with positive screening is subsequently confirmed [3]. Moreover, due to the substantial variability of IMDs, there is a need for individualized disease management. Hence, new methods are needed to reduce the false-positive rate and improve the understanding of these diseases on a systemic level. To this end, mathematical models can simplify complex processes such as human metabolism and associated diseases by modeling metabolic connections.

Data-based modeling Due to the wide availability of large data sets, data-based models such as machine learning (ML) methods, a sub-field of artificial intelligence (AI), have been applied to process large amounts of data and detect valuable patterns within this data. In the medical domain, the application areas of data-based models range from image classification tasks such as mammography interpretation [6], [7], to disease prediction such as diabetes prediction [8], and lung cancer screening [9], [10]. In the context of newborn screening, various supervised classification ML methods have been applied to newborn screening data to predict whether or not a newborn suffers from a condition. Here, not only the known biomarkers (so-called first-tiers) for a disease are investigated, but several screened metabolites and physiological parameters which are usually not utilized for the interpretation of screening results. These can improve the classification accuracy

as the metabolite concentrations are influenced by various factors such as ethnic and demographic factors [11]. In newborn screening, ML methods showed promising results for improving specificity for different conditions [12]–[14].

Nevertheless, the black-box nature of advanced ML algorithms poses a significant challenge in their acceptance by clinicians and integration into newborn screening practices. This underscores the critical need for explainable AI (XAI) methods in the medical context. Therefore, XAI methods have been utilized to provide human-understandable interpretations of advanced ML algorithms [15]. XAI methods are applied in different areas either using intrinsic or post-hoc mechanisms to identify patterns and important features an ML algorithm uses to classify an instance correctly [16]. They aim to become trustworthy, transparent, interpretable, and explainable for human experts.

Mathematical metabolic modeling Mathematical models simulate processes from the real world, often applying ordinary differential equations (ODEs) or partial differential equations to capture time or spatial dependencies within a model framework derived from experiments and theory. The human metabolism is an extensive network capturing various information on metabolic reactions, fluxes, and corresponding metabolites. For this complex network, mathematical models have been developed, assembled from a genome level to investigate the complexity of an organism’s metabolism in health and disease. Starting with the Recon1 [17] human reconstruction, several successive cell-based generic reconstructions of human metabolism have been published [18], [19]. Extending this work, sex-specific and organ-resolved whole-body models (WBMs) of human metabolism have been developed [20]. Using the constraint-based reconstruction and analysis (COBRA) approach, they can predict emergent metabolic properties in human metabolism.

Regarding IMDs, computational metabolic models have been applied to predict known IMD biomarkers by comparing predicted reaction fluxes of different models. The IMD biomarker prediction has been performed on cell-based models [21], organ-specific models [22], and whole-body models [20]. All these models assume that the metabolite concentrations do not change over time. In pharmacokinetic modeling, time-dependent mathematical models simulating the dynamic behaviors of metabolism concerning drugs have been developed. These models enable the analysis of metabolite concentrations and the simulation of changes on small time scales. For time-dependent evaluation of IMD biomarker prediction in metabolic models an approach combining pharmacokinetic modeling with cell-based metabolic human models has been developed [23].

1.2 Current research challenges

Newborn screening and IMD research have evolved significantly since the 1960s. However, combining this field of medical research with mathematical modeling is still a relatively new field and faces some challenges.

1. Large number of false positives in newborn screening for IVA and GA1

In general, newborn screening is a highly effective program. However, newborn screening for IVA and GA1 is hampered by limited diagnostic accuracy. In newborn screening for IVA, there is an increasing number of false positives due to the increasing use of pivmecillinam, an antibiotic used in urinary tract infections in pregnant women [24]. Furthermore, newborn screening for GA1 is also challenged by a high number of false positives. Due to the low prevalence of both diseases, the newborn screening data sets are highly imbalanced. At the same time, a sensitivity of 100%, translating to identifying all newborns with IVA or GA1, and a very high specificity, are essential. For IVA, there is another challenge as its inclusion in the newborn screening disease panel also resulted in the identification of individuals with an attenuated, possibly asymptomatic, disease variant ("mild" IVA [25]), which was unknown in the pre-screening era. By this, the estimated birth prevalence of IVA increased from 1 in 280,000 newborns to 1 in 90,000 - 100,000 newborns worldwide [26].

2. Black-box nature of machine learning methods

In many medical ML applications and also in newborn screening, non-interpretable AI methods are used [13], [27]. However, these methods often rely on a mixture of features or complex learning algorithms that lack interpretability, as they are black-box algorithms. In critical areas such as the health care domain, there is a high demand for medical professionals to understand how and why AI methods work and what their decisions are based on since the corresponding lack of interpretability could prevent the usage of complex algorithms. Clinicians require good model explanations to interpret a system's confidence, estimate the reliability of the system's decision, and understand the features driving a prediction [16], [28].

3. Understanding whole-body metabolism of newborns on a systematic level

The subsequent disease management of IMDs after identification of affected individuals through newborn screening is highly individualized due to the substantial variability of IMDs. Mathematically modeling the affected newborn's whole-body metabolism could support clinicians by understanding the diseases and individualized treatment planning. For male and female adults, metabolic whole-body models already exist [20]. However, these adult models are unsuitable for investigating *in silico* infant metabolism. To correctly model newborn metabolism and the effect of IMDs on it, newborn-specific metabolic whole-body models would be required. Although generic cell-level metabolic models do exist for infant metabolism [18], [29], so far, no metabolic whole-body models for infant metabolism have been developed.

4. Lacking time-dependency in whole-body models

Cell-based metabolic mathematical models such as Recon [17] and the WBMs [20] operate on a steady-state assumption. This assumes that the concentration of metabolites in cells or organs remains constant over time, indicating an overall balance between metabolite inflow and outflow. These models do not account for short-term effects, such as responses to perturbations or time-dependent variations in gene expression and metabolite levels. Evaluations on small time scales with dynamic metabolite concentrations are not possible since these models cannot capture the dynamics of the metabolic system and essential nutrients such as the concentrations of essential amino acids before and after feeding.

5. Reliable mathematical models for newborn screening

Mathematical models rarely resemble an underlying natural system perfectly. Hence, necessary assumptions and simplifications must be employed to build a model. However, this introduces uncertainties into the mathematical model. WBMs for infants are parameterized with various parameters from literature and experiments. Parameter values suggested by different sources in the literature often vary, either within a wide range or with conflicting values that do not align or agree with each other, sometimes being limited to a specific range. However, for mathematical models of infant metabolism to be applicable and have a positive impact, the reliability of the

models must be estimated, along with the uncertainty associated with parameters and possible measurement errors coupled with their impact on the model's output.

1.3 Main goals of this work

The explicit goal of this dissertation is to develop data-based classification models to improve the diagnostic accuracy for newborn screening diseases IVA and GA1 by reducing the false-positive rate in the disease diagnosis. Furthermore, this work strives to develop a mathematical model for the metabolism of newborns and infants, using newborn screening data to personalize these. Here, the goal is to enable a systematic analysis of IMDs and test *in silico* treatment strategies for these diseases. In doing so, the aim is to overcome existing challenges by advancing the following five research goals.

1. Develop data-based models for IVA and GA1 to reduce false-positive screening results

To improve newborn screening for IVA and GA1, the goal is to develop a data-based classification model using previously collected newborn screening data. The ML model should be able to identify all patients suffering from one of the diseases, achieving 100% sensitivity and reducing the false-positive rates for both IVA and GA1. Here, two approaches will be developed: a digital-tier method that simulates the scenario when the ML classification is applied as an additional tier after regular newborn screening and a full data approach, where the ML method is applied to all screened data points before newborn screening.

2. Develop explainable AI models for newborn screening

Adding upon the first goal, the second goal is then to develop an explainable AI method for newborn screening classification, allowing the interpretation of black-box classification methods used for data-based modeling. In addition, the goal is to develop a framework that uses explainable AI to detect patterns in different patient groups within an IMD data set. This framework will be model-agnostic so that it can be applied to other newborn screening conditions and ML methods, and, hence, promote a more common application of XAI methods in ML-based newborn screening.

3. Develop metabolic whole-body models of infants

The third goal is to develop sex-specific, organ-resolved whole-body models of infant metabolism (infant-WBM) by applying a constrained-based modeling approach. These models will use detailed knowledge of infants' physiology and metabolic processes, such as organ-specific parameters. Furthermore, they will include the infant's energy demand for brain development, heart function, muscular activity, and thermoregulation while enabling reasonable growth comparable to WHO growth recommendations [30]. The models will be personalizable utilizing parameters collected in newborn screening, such as sex, body weight, and measured metabolite concentrations. To show the applicability of the models for newborn screening, they will be tested on their ability to predict known biomarkers and dietary treatment strategies for IMDs correctly.

4. Develop time-dependent whole-body models of infants

Besides evaluating the infant-WBMs on a daily time scale, another goal is to develop time-dependent whole-body models of infants. Therefore, one goal of this dissertation is to integrate a physiology-based pharmacokinetic (PBPK) modeling framework into whole-body models for infants. For this, an infant-specific PBPK model will be developed, which models the change of biomarker concentration over time in different organs and the blood compartment. Then, this model will be integrated into the previously developed whole-body models for infant metabolism. These integrated models will then be evaluated on their predictions of the concentration of a known IMD biomarker over a specified time interval.

5. Perform uncertainty quantification for whole-body models of infants

Uncertainty quantification will be performed to ensure the reliability of the time-dependent infant-WBMs and to understand the impact of parameter uncertainties on the model output. Therefore, Monte Carlo simulations will be used to estimate the model's sensitivity toward uncertain input parameters establishing the overall reliability of the time-dependent infant-WBMs concerning these parameter estimations.

1.4 Main contributions of this work

In the course of this research, four scientific papers were published in peer-reviewed journals. A novel data-based method, combining linear discriminant analysis and logistic regression, was presented, which reduced the false-positive rate for IVA classification [31]. Moreover, to evaluate the reliability of ML-based newborn screening methods, the application of explainable AI algorithms on an ensemble and deep learning classifier was shown, using a cumulative approach to understand patterns within patient groups of IVA classification [32]. These attribution could be addressed based on a systematic literature review which was conducted to gather knowledge on state-of-the-art research in ML-based newborn screening [33]. To enable a systematic understanding of newborn and infant metabolism and its malfunctioning due to IMDs, a sex-specific, organ-resolved WBMs of infant metabolism was developed [34].

1.5 Motivation and research impact

Through the development of data-based models using a digital-tier approach, false-positive rates for IVA and GA1 could be successfully reduced. This has a positive impact on newborns and their families as it reduces their burden of false positives, including overtreatment. Furthermore, using ML methods as a digital-tier could support the cost-effectiveness of newborn screening for IVA and GA1 since false-positive screening results are associated with additional costs and effort. These human and material costs include costs for information transmission to the local hospital, clinical evaluation of the newborn, costs for these metabolic (and genetic) analyses, and time used for communicating the results. These resources could be reduced while the hard- and software costs for the developed methods are low. Moreover, reducing false-positive screening results would allow the screening for IVA and GA1 to focus on the true positives.

Furthermore, developing metabolic infant models could be the first step toward a digital metabolic twin for newborn and infant metabolism, which would allow for personalized systematic simulations and treatment planning. These methods could eventually be used to test therapies and treatments *in silico*, as due to the extreme variability of IMDs, their management and treatment have to be personalized for each patient, based on the patient's diagnosis and phenotype [35]. Certain IMDs investigated in newborn screening, such as IVA and GA1, can be treated effectively with dietary intervention [35], which can be simulated using the developed infant-WBMs by adapt-

ing the dietary intake constraints. Moreover, the proposed time-dependent framework could be used for modeling pharmacokinetics of drugs specifically for infants as these processes are often different from adults [36], [37]. Hence, being able to model drug-related infant metabolism could be very beneficial for both researchers and patients. In the era of precision medicine, the ability to accurately predict an infant's metabolic response to dietary and drug-related interventions holds immense potential for personalized nutritional strategies, clinical decision-making, and improving the management of IMDs in infants [38].

1.6 Thesis structure

This dissertation is structured in the following order. Chapter 2 introduces the medical background on newborn screening and the analyzed diseases IVA and GA1, as well as the state-of-the-art research on ML methods for newborn screening and whole-body metabolic models. Based on this foundation, chapter 3 presents the results of the developed data-based algorithms for newborn screening classification of IVA and GA1. Chapter 4 builds upon the latter chapter by integrating an XAI analysis for ML-based newborn screening to interpret the algorithm's decision patterns. Then, chapter 5 presents the developed mathematical models for infant metabolism, infant-WBMs, and their application to newborn screening and IMDs. The subsequent chapters build upon enhancing these models by expanding the infant-WBMs for time-dependent metabolic analysis through integrating PBPK models and performing a singular value decomposition of the underlying stoichiometric matrix, chapter 7. In each chapter, the respective results are discussed. Additionally, chapter 8 provides an overall conclusion with a discussion of the main findings, limitations, and implications for future research.

2

State-of-the-art

This chapter will present an overview of the background and current research in newborn screening for IVA and GA1, machine learning-based newborn screening, and metabolic whole-body modeling. By emphasizing the most relevant concepts, methodologies, and findings, it aims to establish the foundational knowledge necessary for the subsequent chapters.

2.1 Medical background on newborn screening

This section introduces the medical and biochemical background of newborn screening for IMDs and especially for IVA and GA1.

2.1.1 Newborn screening

In 1961, Robert Guthrie introduced the first method for population-based mass blood screening for phenylketonuria (PKU) [39], laying the foundation for newborn screening. Newborn screening programs aim at the early detection of treatable rare diseases that pose a risk to the physical and mental development of affected children. Ideally, this identification should occur pre-symptomatically to significantly reduce morbidity and mortality, making these programs highly successful instruments of secondary prevention [40]. Therefore, blood samples from newborns are collected on the first days of life (i.e., in Germany at 36 - 72 h of life) and sent to a newborn screening center for analysis [1]. Starting with PKU, nowadays, there is a growing panel of different newborn screening

conditions [1], [41]. The number of conditions examined in newborn screening varies significantly among different countries worldwide and within Europe [2]. Loeber et al [1] conducted a detailed inventory of newborn screening programs in Europe since 2010. The results revealed that from the 51 examined countries 49 have some kind of newborn screening program in place [1]. The number of diseases screened ranged from one to 35 in different European countries including pilots [1]. The sensitivity and specificity of the testing methods for each disease play a crucial role in assessing the program's effectiveness [2]. Since 2021, the national German newborn screening panel has comprised 19 conditions [3]. These conditions include 13 metabolic diseases, two endocrinopathies, cystic fibrosis, severe combined immunodeficiency, 5q-associated spinal muscular atrophy, and sickle cell disease [3]. The screening for these conditions is based on the analysis of biochemical and genetic markers in blood or urine, aiming for a diagnostic sensitivity of 100% to avoid false negatives and a very high specificity ($> 99.5\%$) to minimize false positives [42]. Additionally, second- and multiple-tier strategies have been established for some diseases to improve the specificity [43]. These strategies, for example, involve an extra analysis step of additional biochemical parameters in the same dried blood spot sample if the primary biomarker (first-tier) shows pathological changes [43].

Tandem Mass Spectrometry The biochemical targeted method of tandem mass spectrometry (MS/MS) was introduced to the German regular newborn screening in 2005 [2]. Before this introduction, MS/MS was tested in pilot projects for extended newborn screening in Hannover, Heidelberg, and Munich in 1998 and 1999 [2]. MS/MS consists of two mass spectrometries used sequentially to identify and quantify metabolite profiles [44]. By this, the method enables the efficient evaluation of various metabolites from small samples, such as dried newborn blood spots [35]. MS/MS analyzes more than 50 analytes, including amino acids and acylcarnitines, at once. This makes it an ideal high throughput method test for newborn screening and hereby allows early therapy and treatment, successfully preventing severe, even fatal, manifestations of the disease [2].

2.1.2 Human metabolism

The human metabolism is a complex, intertwined system of physical and chemical reactions and processes responsible for sustaining an individual's life [45]. As described by Lanpher et al, metabolism can be understood as "the totality of all the chemical reactions that operate in a living organism" [35]. The metabolic system consists of small molecules, the metabolites, which are chemically transformed during metabolism. In newborn screening, the blood concentrations of

known disease biomarkers are measured to detect abnormal increases or decreases, which may indicate a specific condition. At the newborn screening laboratory at Heidelberg University Hospital (UKHD), 52 biomarkers are screened, Table 2.1. Most of these biomarkers belong to the substance group of amino acids or acylcarnitines, Table 2.1.

Amino Acids The human body relies on amino acids as they are the building blocks of proteins. The absence of even one amino acid can disrupt the synthesis of most proteins [46]. Fortunately, the human body can produce many amino acids on its own [46]. However, there are eight essential amino acids that the body cannot produce on its own. These essential amino acids include valine, leucine, isoleucine, phenylalanine, tryptophan, lysine, methionine, and threonine. They must be acquired through nutrition, as plants and microorganisms can synthesize them [46]. Furthermore, there exist conditionally essential amino acids, which are essential during pregnancy (histidine) and in the neonatal period growth or during times of stress (arginine) [46], [47]. The non-essential amino acids can be converted into one another, which is a crucial process as the human body often does not receive the optimal amount of each amino acid through nutrition [46]. The concentrations of various amino acids in the blood remain relatively constant, with cells selectively taking up the amino acids they currently need for their biosynthesis. Amino acids are then utilized to build proteins.

Acylcarnitines Acylcarnitines are a large class of metabolites in the non-protein amino acid family. Their abundance and special structure make acylcarnitines play an essential role in cell physiological activities and become a fundamental substance for cell metabolism [48]. They are the transport form of activated long-chain fatty acids, which enables them to be carried into and out of the mitochondria. Here, the metabolism of fatty acids takes place in the form of β -oxidation, the degradation of fatty acids as source for energy production [46]. Without the transport molecule carnitine, the outer mitochondrial membrane would be impassable for the fatty acids [46]. In particular, for the brain, carnitine is an essential factor as it highly relies on oxidative metabolism [49].

2.1.3 Inherited metabolic diseases

Inherited metabolic diseases (IMDs) is a growing group of disorders with a variable but often severe clinical phenotype which is caused by inherited deficiency of enzymes or transporter proteins in a specific metabolic pathway. This can entail pathogenic gene variants causing either a loss or

Table 2.1: Metabolites measured for regular newborn screening at the newborn screening laboratory at Heidelberg University Hospital showing their respective abbreviations and substance groups.

Abbreviation	Name	Substance group
17p	17-hydroxyprogesterone	steroid
TSH	Thyroid-stimulating hormone	hormon
BIO	Biotinidase activity	enzyme
3HMG	3-OH-2-methyl-butyrylcarnitine	acylcarnitine
Ala	Alanine	amino acid
Arg	Arginine	amino acid
Asa	Argininosuccinate	amino acid
Asp	Aspartate	amino acid
C0	Free carnitine	acylcarnitine
C10	Decanoylcarnitine	acylcarnitine
C10:1	Decenoylcarnitine	acylcarnitine
C12	Dodecanoylcarnitine	acylcarnitine
C14	Tetradecanoylcarnitine	acylcarnitine
C14:1	Tetradecenoylcarnitine	acylcarnitine
C14OH	3-OH-Tetradecanoylcarnitine	acylcarnitine
C16	Hexadecanoylcarnitine	acylcarnitine
C16:1	Hexadecenoylcarnitine	acylcarnitine
C16:1OH	3-OH-Hexadecenoylcarnitine	acylcarnitine
C16OH	3-OH-Hexadecanoylcarnitine	acylcarnitine
C18	Octadecanoylcarnitine	acylcarnitine
C18:1	Octadecenoylcarnitine	acylcarnitine
C18:1OH	3-OH-Octadecenoylcarnitine	acylcarnitine
C18:2	Octadecadienoylcarnitin	acylcarnitine
C18OH	3-OH-Octadecanoylcarnitine	acylcarnitine
C2	Acetylcarnitine	acylcarnitine
C3	Propionylcarnitine	acylcarnitine
C4	Butyrylcarnitine	acylcarnitine
C5	Isovalerylcarnitine	acylcarnitine
C5:1	Tiglylcarnitine	acylcarnitine
C6	Hexanoylcarnitine	acylcarnitine
C8	Octanoylcarnitine	acylcarnitine
C8:1	Octenoylcarnitine	acylcarnitine
Cit	Citrulline	amino acid
GALT	Galactose-1-phosphate-Uridyltransferase activity	enzyme
Gln	Glutamine	amino acid
Glu	Glutamic acid	amino acid
Glut	Glutaryl carnitine	acylcarnitine
Gly	Glycine	amino acid
Hci	Homocitrulline	amino acid
His	Histidine	amino acid
IRT	Immun Reactiv Trypsin	enzyme
Leu+Ile	Leucine + Isoleucine	amino acid
MeGlut	3-Methylglutarylcarnitine	acylcarnitine
Met	Methionine	amino acid
Orn	Ornithine	amino acid
Phe	Phenylalanine	amino acid
Pro	Proline	amino acid
SUCC-MS	Succinylacetone	organic acid
Thr	Threonine	amino acid
Trp	Tryptophan	amino acid
Tyr	Tyrosine	amino acid
Val	Valine	amino acid

gain of function in proteins, significantly impacting the overall metabolism and health of the affected individual [35]. Therefore, an IMD's overall impact at the organismic level is altering one or several metabolic fluxes within the pathway regulated by the mutant protein [35]. The metabolic flux is defined as the production or elimination of several metabolites per mass of an organism over a specific amount of time [35]. IMDs are diseases that are individually rare (from 1:10,000 to 1:1,000,000 newborns) but cumulatively account for over 1,904 recorded IMDs in 2023 (<http://www.iembase.org>). Ideally a diagnostic biomarker for an IMD is a measurable metabolite that might provide information about the presence or severity of the disorder. These biomarkers are valuable in diagnostics as they consistently exhibit changes in concentration during a disease state [21]. An IMD's impact on the human body can manifest through various mechanisms, including a toxic accumulation or deficiency of metabolites, and a redirection of the metabolic flow to secondary metabolic pathways [35]. Two IMDs screened for in newborn screening are IVA and GA1.

2.1.4 Isovaleric aciduria (IVA)

Isovaleric aciduria (IVA; OMIM #243500) is an organic aciduria leading to severe life-threatening (neonatal) metabolic decompensations. IVA has an estimated disease prevalence of 1 in 90,000 - 100,000 newborns in Germany [4], [5], [26]. The disease is inherited autosomal recessively, meaning that a child needs to receive an affected allele each from mother and father to develop the condition. IVA is caused by bi-allelic pathogenic variants in the *isovaleryl-CoA dehydrogenase* (IVD) gene (cytogenic location: 15q15.1). These variants result in a deficiency of IVD (EC 1.3.99.10) in the leucine degradation pathway and, hence, accumulation of metabolites deriving from isovaleryl-CoA. The regular way would be that leucine is broken down into isovaleric acid, which is subsequently converted into energy [4]. However, the variation in the IVA patients leads to a decrease in enzyme activity of the mitochondrial IVD, followed by an aggregation of isovaleryl-CoA and its corresponding metabolites [4]. For patients suffering from IVA, there exists a dysfunction in the activity of the enzyme isovaleryl-CoA dehydrogenase. Hence, the process of breaking down isovalerylcarnitine (C5) is impaired leading to an accumulation of isovaleric acid, 3-hydroxyisovaleric acid, C5, and isovalerylglycine [50].

If not treated, many patients experience severe life-threatening (neonatal) metabolic decompensations. Treatments aim at long-term metabolic stability of patients with IVA [51]. The primary treatment strategy is to decrease the production and enhance the excretion of isovaleryl-CoA. The

early detection enables specific treatment and reduced mortality for affected individuals with the classical disease course [4]. However, newborn screening for IVA has faced two challenges in recent years. First, there is an increasing number of false-positive screening results due to the increasing use of pivmecillinam, an antibiotic used in urinary tract infections in pregnant women [24]. Since the antibiotic's metabolite pivaloylcarnitine is isobaric to the primary marker C5, newborns with an increased pivaloylcarnitine level are falsely suspected as newborns with IVA [52]. Second, since the inclusion of IVA into the newborn screening disease panel, individuals with an attenuated, possibly asymptomatic, disease variant ("mild" IVA [4], [25]) have been identified. This identification increased the estimated birth prevalence of IVA from 1:280,000 newborns to 1:90,000 - 100,000 newborns worldwide [4], [5], [26]. Hence, patients with mild IVA constitute 80% of those screened for IVA in Germany [4]. These mild IVA patients are at risk of overtreatment, as a multi-center long-term observational study showed [4].

2.1.5 Glutaric aciduria type 1 (GA1)

Glutaric aciduria type 1 (GA1; OMIM: #231670) is an organ aciduria that is inherited autosomal recessively and has an estimated disease prevalence of 1:140,000 newborns [3]. GA1 arises due to a deficiency of the enzyme glutaryl-CoA dehydrogenase [53] which mediates degradation of lysine and tryptophan [54]. The disease is characterized by an accumulation of glutaric acid, 3-hydroxyglutaric acid, and glutarylcarnitine (Glut), as well as an accumulation of glutaconic acid, which is less frequent [53]. Through biochemical analysis, an increase of these metabolite concentrations can be detected in urine, plasma, cerebrospinal fluid, and tissue [53]. Based on their amount of urinary excretion of glutaric acid, two biochemical patient subgroups can be distinguished [55]. Low excretors are defined by a urinary glutaric acid level below 100 millimoles (mmol) per mol creatinine, while high excretors are those patients with urinary glutaric acid excretions above 100 mmol per mol creatinine [55].

Without treatment, 80 % - 90 % of GA1 patients develop neurological diseases mainly between the ages of 3 - 36 months following an acute encephalopathic crisis [53], [56]. GA1 is considered a treatable disease with the first evidence-based guidelines published in 2007 [57], which were further revised in recent years [58], [59]. Treatment of GA1 patients includes a low lysine diet with a lysine-free, tryptophan-reduced, arginine-fortified amino acid mixture and oral carnitine supplementation [59]. Furthermore, for patients with catabolic episodes an intensified intermittent emergency treatment is recommended [59]. However, newborn screening for GA1 is also challenged

by a high numbers of false positives. One known factor for false-positive screening results is renal insufficiency due to congenital or acquired renal disease in newborns since it delays the excretion of Glut via the kidneys, which increases the Glut concentration in plasma [60].

2.2 Data-based modeling in newborn screening

This section is based on and extends the published work Zaunseder et al [33]. It presents a systematic literature review on applying machine learning (ML) methods in newborn screening programs. Overall, the literature review showed considerable heterogeneity among the 17 studies, including data from ten screening centers, investigating 21 diseases, applying 12 ML classification methods, and 14 feature selection strategies.

2.2.1 Methodology of systematic literature review

The systematic literature review was performed and reported according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (www.prisma-statement.org). The methodological steps for this analysis included defining three research questions, outlining the literature search strategy and study eligibility criteria, as well as defining the data extraction and synthesis steps, which are summarized in a PRISMA flow chart, Figure 2.1.

Research questions

The primary outcome of this systematic literature review was to evaluate the applicability, advantages, and limitations of ML-based newborn screening. Therefore, the published studies were analyzed according to the following research questions:

1. *Which data pre-processing methods have been applied in newborn screening?*
2. *Which ML classification methods have been applied in newborn screening, and how did they perform?*
3. *How were pattern recognition techniques implemented in newborn screening?*

Search strategy

A two-stage search procedure was conducted to compile relevant papers. In the initial phase, five electronic databases (ScienceDirect, IEEE, ACM, Sage, PubMed) were searched in May 2021 and October 2021 to collect literature. The search keywords were “Newborn Screening” AND (“Machine Learning” OR “Deep Learning” OR “Data Mining”). In the second phase, cross-references from

eligible literature compiled in the first phase were searched via Google Scholar, and advice from pediatric experts was added to compile the final literature collection.

Inclusion and exclusion criteria for study selection

All included studies applied an ML classification method in regular newborn screening and were published between January 2000 and September 2021. Studies were excluded if they did not concern newborn screening, did not use data obtained from MS/MS, or did not apply ML algorithms for disease classification.

Study eligibility

Duplicates were removed before assessment. First, titles and abstracts were screened, and studies unrelated to the research question were excluded. Then, full-text articles were reviewed for inclusion. In case of exclusion, the reason was reported, Figure 2.1.

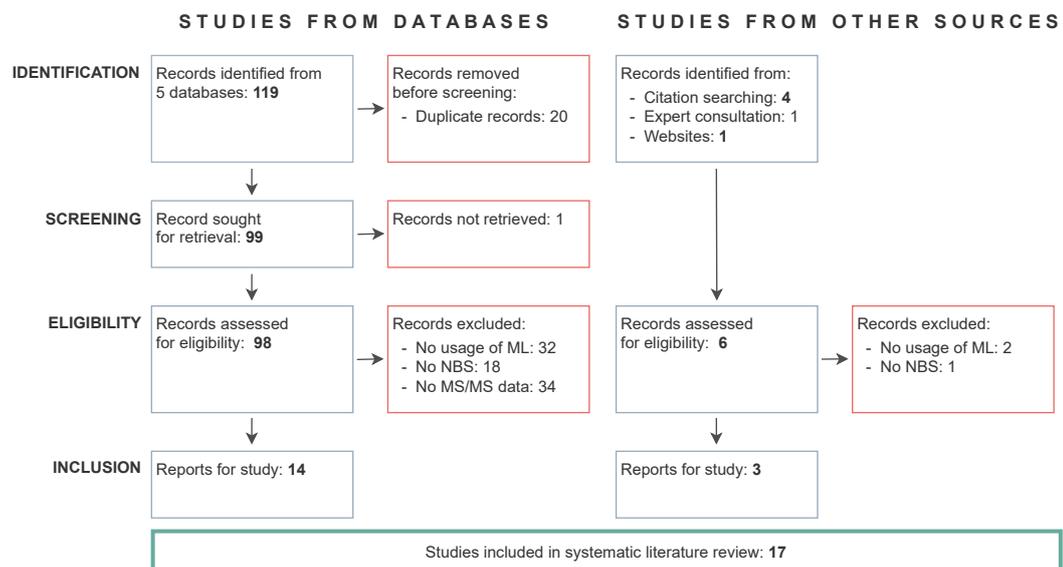


Figure 2.1: PRISMA flow diagram describing the two-stage search procedure for studies identified, screened, included, and excluded for this review. Figure taken from Zaunser et al [33].

Data extraction and synthesis

Data from all studies, including information on authors, data pre-processing, ML classification models, performance, and pattern recognition were extracted and summarized in Supplementary

Table B.1. For the data analysis, key indicators based on their underlying approaches, data requirements, interpretability on a modular level, and performance were considered. The classification performance was evaluated based on the sensitivity, specificity, and positive predictive value (PPV). These performance measures were compared to reference values and other ML methods in comparative studies. For studies lacking sensitivity or specificity values, these values were calculated based on the published contingency tables. The studies were insufficient for a meta-analysis; hence, the findings were synthesized into an overall narrative.

A total of 125 records were identified, and 17 publications were considered for the systematic literature review. The main reasons for dismissing publications were that they did not apply ML methods, did not utilize data obtained from MS/MS, or were not conducted in the context of newborn screening.

2.2.2 Literature review for data-based models in newborn screening

The reviewed literature was analyzed in four categories: examined diseases, applied pre-processing methods, ML classification models, and pattern recognition techniques.

Examined diseases

Overall, 21 diseases were examined in the 17 publications. Among these diseases, PKU [12], methylmalonic aciduria [14], and medium-chain acyl-CoA dehydrogenase deficiency [61] were the most frequently studied diseases. However, ML methods were not frequently applied in the study of IVA and GA1. IVA was only considered in one study [13], where results were not reported for individual diseases, so the predictive power of the developed algorithms could not be assessed for IVA specifically. GA1 was considered in the same cumulative study [13]. Another study developed a two-step approach to discover metabolic markers and showed that their method detects Glut as the primary biomarker for GA1 and suggests C4 as a secondary biomarker candidate [12]. Furthermore, a Random Forest (RF) algorithm was developed, reducing false positives from first-tier newborn screening from 1344 to 150 [14].

Pre-processing methods

Pre-processing methods are employed on a data set before applying an ML algorithm. This process involves analyzing and preparing the data and its features, aiming to enhance the results

when utilizing an ML algorithm. Here, the pre-processing methods for data imbalance, feature construction, and feature selection will be presented.

Data imbalance methods Various techniques are used in the literature to overcome data imbalance, which is a key characteristic of newborn screening data due to the low prevalence of the examined conditions. In previous studies, random sampling techniques were applied to change the data imbalance ratio by randomly excluding data points from the normal (unremarkable) newborn screening profiles [62]. Moreover, an advanced technique called informed sampling was employed in newborn screening. This method involves including newborns in a data set based on specific clinical criteria. These criteria were, for instance, the inclusion of newborns with a normal newborn screening profile but elevated primary markers [61], particularly removing samples close to the decision boundary [13], or normal patients with varying birth weight and gestational age [63]. These sampling methods were able to reduce false-positive classifications in newborn screening by applying additional expert knowledge on primary markers and medical background. Oversampling describes the process of enhancing the data set with synthetic samples from the minority class to decrease the data imbalance. In newborn screening, oversampling was applied by adding spiked blood samples, designed to mimic unhealthy blood samples [64], and mixed models such as SMOTE+ENN [13]. These methods are advantageous as they address the challenge of the low number of samples in the minority class. However, whether the synthetically created sample patients accurately represent a positive confirmed diagnosis remains unclear.

Feature construction Feature construction involves creating new features from existing ones, thereby expanding the feature space dimensionality. These new-build features aim to account for non-linear correlations spread over several metabolites and discover hidden interactions. Feature construction on newborn screening data mainly uses arithmetic operators on original features to construct new features. A new feature x' can be built from two features x_i, x_j by calculating their ratio [64]:

$$x' = [x_i/x_j]; i = 0, 1, \dots, n - 1; j = i + 1, i + 2, \dots, n.$$

Combining multiple features using arithmetic operators such as addition, subtraction, and multiplication was employed to generate complex new features. The features $x_i, i = 1, \dots, n$ can contain all original features [61] or a subset of disease-specific primary markers [64]. Moreover, principal component analysis [63] and self-developed algorithms [65] were applied to build new features.

Feature selection methods Feature selection techniques eliminate irrelevant and redundant information for a classification method. This aids in reducing the number of false-positive results in newborn screening. For this, either a fixed number [66] or adaptive approaches [61] were applied to select the most relevant features. The methods are usually grouped into the three categories *filter*, *wrapper*, and *embedded* methods [67]. For newborn screening, the category of *informed* methods was added, which includes using diagnostic flags developed by biochemical and medical experts [62] and expert consultation with a pediatric endocrinologist [68] to identify informative features. Furthermore, filter methods were frequently applied using statistical measures such as analysis of variance (ANOVA) [13], χ^2 tests [61], mutual information [13], Pearson-like formula [69], Fisher score [70], information gain [68], and relief-based methods [71] to identify relevant features. Moreover, embedded methods such as decision tree splitting rules [13], the discriminatory threshold from logistic regression (LR) [72], learned vector quantization [73], and underlying cost functions, such as L1 norm [13] were analyzed for feature selection. Wrapper methods were used to identify subsets of all features [64] or subsets pre-selected by another method [65], which were relevant for classification.

Implementation of pattern recognition techniques in newborn screening

Feature selection techniques are used to identify patterns within the data by comparing the results to established primary markers [13] and letting clinical and biochemical experts analyze them [62]. Interpretable methods allow to analyze patterns using their built-in decision functions [61] and discriminatory thresholds [71], whereas for non-interpretable methods, model agnostic approaches such as mean decrease accuracy are applied [14]. These pattern recognition methods could confirm established biomarkers and identify markers without other a priori assumptions or conditions [62]. Model agnostic pattern recognition can be applied to non-interpretable methods by discovering non-explainable incidents such as a higher percentage of false-positive newborns with Hispanic ethnicity [11], [74]. The application of explainable artificial intelligence (XAI) for ML methods in newborn screening was not proposed in any of the identified literature.

Evaluation of applied ML classification algorithms

Due to the availability of the confirmation diagnosis, supervised classification methods were applied in newborn screening. ML classification methods can be distinguished according to their interpretability on a modular level, which refers to methods that can inherently explain how parts

of the model affect predictions [75]. The most frequently used interpretable method was logistic regression (LR) [73], followed by the usage of ridge regression (RR) [64], decision trees [63], rule learners [61], and Naive Bayes [12]. Support vector machine (SVM) [70] was applied using linear kernels [12], radial basis function [65], or polynomial kernels [12]. From the class of non-interpretable methods ensemble learners such as random forest [74] and boosting algorithms such as adaptive boosting [68], extremely randomized trees [13], and gradient boosting [13], as well as bagging methods such as Bagging-SVM [68] were applied. Furthermore, neural networks (NNs) [68] and other methods such as K-nearest neighbor [13] and linear discriminant analysis [71] were implemented.

Table 2.2: Sensitivity, specificity, and positive predictive value (PPV) of considered ML classification methods. (A) Results of comparative studies (best-performing ML classification methods with highest sensitivity and specificity). (B) Result of single ML method studies. Where sensitivity and specificity were not stated, the results are calculated based on the published contingency table (*italics*). Results in brackets show a comparison to traditional newborn screening. Diseases with * are biochemical variations nowadays known as non-diseases. Abbreviations: CAH - Congenital adrenal hyperplasia, CH - Congenital hypothyroidism, CIT2 - Citrullinemia type II, DT - Decision Tree, GA1 - Glutaric aciduria type I, IVA - Isovaleric aciduria, LR - Logistic Regression, MCADD - Medium-chain acyl-CoA dehydrogenase deficiency, 3-MCCD - 3-methylcrotonyl-CoA carboxylase deficiency, MET - Hypermethioninemia, MMA - Methylmalonic aciduria, NN - Neural Network, OTCD - Ornithine transcarbamylase deficiency, PKU - phenylketonuria, RF - Random Forest, RR - Logistic ridge regression, SCADD - Short-chain acyl-CoA dehydrogenase deficiency, SVM - Support vector machine, VLCADD - Very long-chain acyl-CoA dehydrogenase deficiency.

Disease	ML Classification	Sensitivity (%)	Specificity (%)	PPV (%)
(A) Comparative ML classification studies				
PKU	LR [71]	100	99.793	17.41
	LR [76]	98.0	99.9	-
	LR [62]	96.809	99.905	49.46
MMA	NN [12]	98.0	-	98.0
MCADD	RR [71]	100	99.987	33.90
	LR [71]	96.83	99.992	88.41
	LR [62]	95.238	99.992	88.24
3-MCCD*	LR [62]	95.455	99.957	33.33
CH	Bagging-SVM [68]	73.33	100	-

CIT2, MET, SVM [13]		91.30	36.36	19.29
MMA, PKU, SCADD*				
(B) Single ML classification studies				
PKU	SVM [65]	100	99.997 (99.971)	-
	SVM [70]	100 (100)	99.98 (99.96)	-
	LR [73]	97.66	31.61	24.59
MMA	SVM [66]	100 (100)	100 (99.79)	-
	RF [14]	100 (100)	89.678 (81.226)	26.40 (16.40)
	RF [74]	96.117 (96.117)	65.143 (28.286)	28.9 (16.5)
	SVM [69]	95.9 (81.4)	95.6 (76.2)	-
MCADD	LR [72]	100 (100)	99.988 (99.924)	18.2 (3.4)
	RL[77]	100 (100)	99.901 (98.463)	93.75 (49.18)
GA1	RF [14]	100 (100)	94.503 (50.751)	22.30 (3.10)
3-MCCD*	SVM [65]	100	99.936 (99.711)	-
MET	SVM [65]	100	99.986 (99.958)	-
VLCADD	LR [72]	100 (100)	100 (100)	100 (100)
	RF [14]	100 (100)	92.786 (92.639)	23.40 (23.10)
OTCD	RF [14]	100 (100)	99.601 (81.983)	62.10 (3.50)
SCADD*	LR [72]	100 (100)	99.997 (99.974)	73.3 (22.0)
CAH	DT [63]	90.909 (100)	100 (87.194)	66.7 (20)

For parameter optimization, grid search is commonly applied [64]. The robustness of the methods was tested using cross-validation [65] or stratified cross-validation [64] as well as evaluation of receiver operating characteristic curves [14]. The classification performance was evaluated using classification sensitivity, specificity, and PPV. Previous ML classification results could be evaluated separately for single and comparative studies, as the latter enabled a comparison of various ML methods on the same data set, Table 2.2. Most studies applied sampling algorithms, changing the sick-to-control ratio and reduced data sets, such as only including false-positive patients from regular newborn screening. Hence, Table 2.2 presents the results from the largest or unknown validation data sets. In the comparative studies, LR, RR, SVM, Bagging-SVM, and NNs achieved the best results, whereas evaluating the single studies was more difficult due to the differences in

the data sets. The results from Lin et al [13] are presented in a separate row since they only report average evaluation results for groups of diseases. To evaluate the performance of the classification methods, the following criteria were established:

- ability to maintain 100% sensitivity and high specificity results
- frequency of application,
- performance in comparative studies and in comparison to reference values,
- interpretability on a modular level.

Considering all reviewed studies, the classification methods LR and SVM were valuable candidates for newborn screening classification according to these criteria. They were frequently applied in newborn screening, achieved 100% sensitivity for various diseases in several studies, were the best algorithms in most comparative studies, and could increase the sensitivity, specificity, and PPV compared to reference values from traditional screening, Table 2.2. LR is interpretable on a modular level, as the model and weights can be intuitively interpreted, whereas SVM is more difficult to interpret on a modular level. Furthermore, in two comparative studies, advanced versions of these methods, such as Bagging-SVM and RR, achieved the best results [64], [68].

2.3 Metabolic whole-body modeling

2.3.1 Metabolic modeling for human metabolism

For the last decades, metabolic models have played an increasing role in understanding human metabolism and IMDs. Thanks to advancements in whole genome sequencing, methods to automatically annotate the genome and to derive genome-scale metabolic models (GEMs) were created [78]. The process of annotation is based on gene–protein–reaction rules, using information that links enzymes and protein identifiers, reaction names, and reaction metabolites, Figure 2.2. The fundamental goal is to accurately define the chemical transformations that take place among chemical components in a network and build a GEM to analyze complex metabolic networks, Figure 2.2.

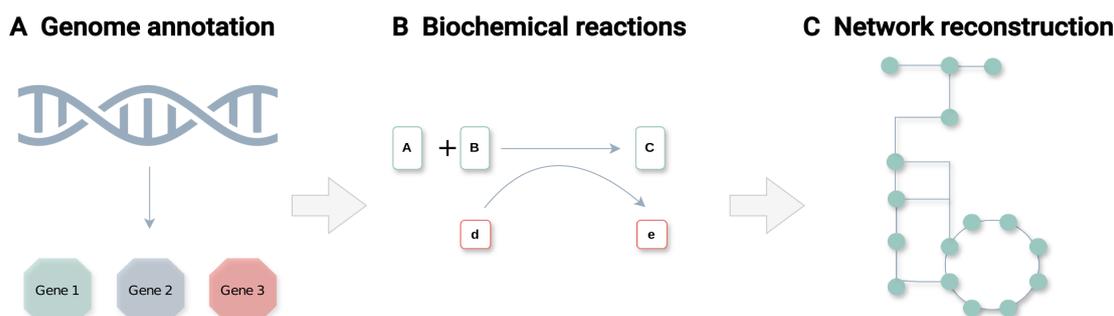


Figure 2.2: Overview of genome-scale metabolic network reconstruction. (A) Starting from genome annotations, which lead to (B) an assembly of biochemical reactions, which are then connected through (C) a metabolic network reconstruction.

A detailed understanding of molecular-level processes within a metabolic system is achieved by creating metabolic networks. This is done using constraint-based reconstruction and analysis (COBRA), as described by Palsson [79]. The network is built by compiling the identified biochemical reactions, which account for all metabolic reactions encoded within the genome [80]. These metabolic models can be used for biotechnological applications across different cell types. The applications of metabolic models extend to various organisms, including microbial models like *Escherichia coli* [81] and animal cells such as the Chinese hamster ovary [82]. Additionally, in agricultural contexts, metabolic engineering is employed to model plant metabolism, showcasing the variability of these models across different domains [83]. Human metabolic networks enable a systematic view of human metabolism and its connections to various diseases. In the medical field, a comprehensive understanding of human metabolism at a systemic level is crucial.

Human models Numerous successive generic reconstructions of human metabolism have been published, such as Recon1 [17] and the human metabolic reaction series [84], which presented the first generic reconstructions of the human metabolic network using cell-based models. Recon1 is an extensive GEM as it includes 2,766 metabolites as well as 3,311 metabolic and transport reactions [18]. Recon2 [85], Recon2.2 [80], and Recon3D [18] expanded Recon1 by utilizing several metabolic data sets to identify new metabolites as well as transport and catalyzing reactions. Furthermore, Human1 [19] presents a unified human GEM which was developed by integrating and extensively curating Recon models [17], [18] and the human metabolic reactions series (HMR1, HMR2) [86], [87]. Based on HMR2, a GEM modeling infant metabolism was developed and applied to predict the infant's growth rate over the first six months of life [29]. The simulation toolbox for infant growth (STIG-met) was created to achieve this. This toolbox simulates the growth of an infant fed with breast milk and adapts to the infant's energy needs, considering energy required for maintenance, the energy expenditure of major organs, and muscular activity [29]. In general, GEMs enable considerable personalization as they allow for an integration of various data sources, such as metabolomics [88], transcriptomic [89], proteomic [90], and metagenomic data [91]. For specific applications, the metabolism of particular pathways and the requirements of specific organs are of interest. This has led to the development of organ-specific metabolic models, including those for the liver [92], adipocytes [93], myocytes [93], erythrocyte metabolism [94], and small intestine [95].

Analyzing the impacts of diseases, drugs, or nutrition on the entire body goes beyond studying individual organs. Hence, it is crucial to capture the interactions between all organs to gain a comprehensive overview [20]. Sex-specific whole-body models (WBMs) of humans have been constructed by integrating information from Recon3D, organ-specific details, and omics data [20]. These models capture the metabolism of 26 organs and six blood cell types in two, male and female, adult models named *Harvey* and *Harvetta* [20]. The goal is to model the metabolism of the entire human, and not solely on a cellular level, as seen in the Recon models [20]. Each model was constrained with 15 physiological parameters, leading to a novel paradigm in constraint-based modeling known as physiologically and stoichiometrically constrained modeling (PSCM) [20]. The PSCM toolbox is available in MATLAB at opencobra.github.io/cobratoolbox. By this, WBMs present a first step towards a virtual metabolic human, or digital metabolic twin [20].

2.3.2 Constrained-based modeling

For the mathematical analysis and simulation of fluxes within a metabolic reconstruction, the reconstruction network is transformed into a stoichiometric matrix $S \in \mathbb{R}^{m \times n}$ where the rows correspond to the m metabolites and the columns to the n reactions, Figure 2.3 (C), (D).

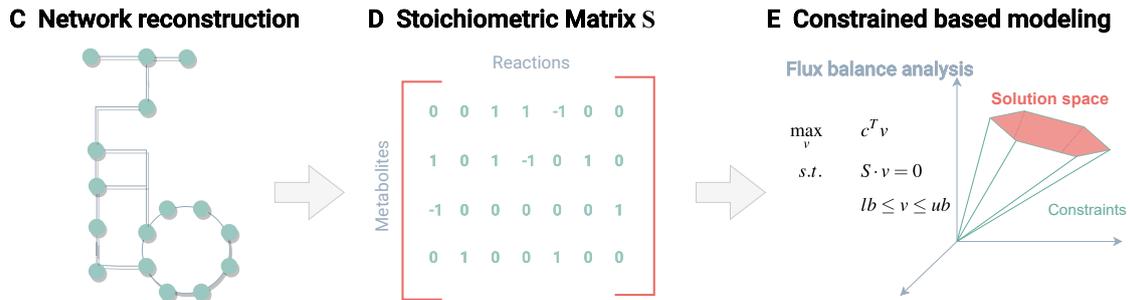


Figure 2.3: Overview of (C) genome-scale metabolic network reconstruction, the translation into (D) a stoichiometric matrix, and the subsequent (E) evaluation in constraint-based modeling.

The matrix entries s_{ij} are assigned a stoichiometric coefficient if metabolite i takes part in reaction j and zero otherwise. However, besides the connectivity of the stoichiometric matrix, further biological knowledge on metabolic reactions can be described in the form of constraints. Palsson [79] identified different types of constraints that limit cellular functions which he divided into the following four categories:

- 1. Fundamental physio-chemical constraints:** These are hard constraints that must not be breached, encompassing principles such as conservation of mass, elements, energy, and momentum. They are deduced from physio-chemical conditions within cells and their surrounding.
- 2. Spatial constraints:** These constraints describe the spatial or three-dimensional limitations on the cell determined by the accumulation of molecules within the cell. These spatial constraints and their consequences can contribute to comprehending the three-dimensional organization of cells.
- 3. Environmental constraints:** The cell is subjected to environmental constraints imposed by its surroundings, including factors like nutrient availability, pH, temperature, and osmolarity. Understanding these environmental conditions is crucial for supporting quantitative studies. While experimental setups allow for precise determination of these constraints, achieving the same precision level in natural environments becomes challenging.

4. **Regulatory constraints:** These constraints stand apart from the aforementioned, being self-imposed and time-dependent. This implies that they limit the behavior of a cell based on evolutionary factors and that, at any given time, the transcriptional activity is determined by the state of both the external and internal environments [96].

All these constraints are then incorporated into the metabolic system by setting upper ub and lower bounds lb on the reaction fluxes v ,

$$lb \leq v \leq ub, \quad (2.1)$$

where the flux describes the amount of a metabolite measured in mmol, passing through the reaction in a specific time t , relative to a mass, for instance, dry weight or human body. A specific flux v_k describes the change of a metabolite concentration x_k over time t in reaction k . The change of a metabolite concentration over time coming from all reactions in the network can be expressed by multiplying the stoichiometric matrix S with the flux vector v , resulting in the mass-balance equation for all metabolites

$$\frac{dx}{dt} = S \cdot v. \quad (2.2)$$

Flux balance analysis (FBA) [81] is a mathematical approach to calculate the flux through the metabolic network. This approach assumes that the metabolic system is in a steady state and, hence, $dx/dt = S \cdot v = 0$. The steady-state constraint indicates that the metabolite concentrations remain constant over time since the input and output fluxes must balance.

2.3.3 Modeling of inherited metabolic diseases

In medical applications, metabolic modeling is increasingly utilized, encompassing tasks such as identifying drug targets [97] and off-target drug effects [85], studying cancer metabolism [98], enhancing the understanding of microbial interactions with the host organism [99], and also emerging as a field for investigating IMDs [21]. IMD analysis is utilized in metabolic modeling to demonstrate the human metabolic model's capability to predict known biomarkers accurately. IMDs often involve a disruption of the normal metabolite flux [35]. Therefore, analyzing the metabolic flux over time is significantly more informative for IMDs than the static measurement of intermediary metabolites [35].

In this regard, Shlomi et al [21] developed a computational framework based on constraint-based modeling to predict changes in metabolite concentrations in Recon1 [17]. In their framework, they

perform gene knockouts to simulate an IMD in a metabolic model and compare it with an unchanged wild-type model. The comparison involves assessing the flux intervals, representing the maximal and minimal uptake and secretion fluxes within the feasible flux state. This is accomplished through a method known as flux variability analysis [100]. In the wild-type model, reactions are assumed to be activated, with the flux set to a non-zero value, whereas in the IMD model, they are assumed to be inactive, with the flux set to zero [21]. The fluxes of interest involve exchange reactions simulating the uptake or secretion of a biomarker metabolite between intracellular and extracellular compartments. This comparative analysis evaluates whether the biomarker is projected to be elevated, reduced, or unchanged in the IMD model compared to the wild-type model. The change is considered highly certain if the flux intervals of the two models do not overlap. The framework of Shlomi et al [21] identified a set of 233 metabolites with predictions of elevated or reduced changes based on 176 possible dysfunctional enzymes.

Recently, these biomarker prediction strategies have been incorporated to assess the validity of human reconstructions. Recon2 employed a manually compiled compendium of IMDs as a benchmark, encompassing 330 IMDs [85]. They showed that Recon2 achieved a 77% accuracy in predicting known biomarkers for IMDs. In comparison, Recon3D accurately predicted 50.2% of biomarkers in biofluid correctly [20]. This accuracy was further improved by using sex-specific WBMs for biomarker prediction, evaluated across 57 IMDs, where all organ-specific reactions carrying the known defective gene(s) were removed [20]. The analysis of 57 IMDs showed that the female WBM qualitatively predicted 215 out of 252 (85.3%) biofluid-specific biomarkers correctly, while the male WBM accurately predicted 214 out of 252 (84.9%). The authors suggest that the enhanced predictive abilities of the WBMs primarily stem from a reduction in inconclusive or unchanged predictions. Additionally, organ-specific metabolism, such as liver metabolism, was examined in the context of IMDs [22]. Pagliarini et al [22] proposed a computational workflow designed to model the changes in metabolism resulting from IMDs and predict the metabolites and reactions affected by the mutation. They adapted and examined a published metabolic model of hepatocyte metabolism [92] to detect metabolites accumulating in hepatocytes in IMDs. They employed the principle of flux-minimization as a biological objective function, which enables the comparison between the metabolic model predictions and the known biomarkers for each IMD. They developed a new method named flux difference analysis to examine 38 IMDs in the liver, demonstrating that for approximately 50% of the investigated IMDs, the flux difference analysis accurately predicted the known IMD biomarkers.

2.3.4 Time-dependent metabolic modeling

As mentioned earlier, the FBA approach assumes a steady state, which implies no changes in the metabolite concentrations over time. Models incorporating time-dependent features can simulate dynamic changes within the metabolic network. These dynamic aspects include allosteric and posttranslational regulation, alterations in metabolite concentrations, and considerations of thermodynamics [101]. Building upon the abilities of dynamic models, approaches to combining dynamic and constraint-based models have recently been developed since they increase the spatial and temporal resolution of GEMs [101]. For instance, the unsteady-state FBA approach has been developed to improve the accurate prediction of metabolic flux states for red blood cells by relaxing the steady-state assumption [102]. Mannan et al [103] introduced an approach, which integrates parameters from a genome-scale metabolic network model into a kinetic model of the central carbon metabolism of *E. coli*. Another modeling framework considered both reaction kinetics and network connectivity constraints, emphasizing the role of metabolic network connectivity in influencing cellular control over metabolite levels [104]. For human metabolism, Mohammad et al [105] integrated physiology-based pharmacokinetics (PBPK) modeling and constraint-based metabolic models to investigate the gut-brain axis for patients with an autism spectrum disorder. Moreover, Guebila et al [106] integrated a PBPK model of glucose regulation by insulin, glucagon, and incretins into whole-body models to perform dynamic flux balance analysis for type 1 diabetes. Furthermore, to analyze the effect of processes and genetic variations in ethanol metabolism, Zhu et al [107] integrated a PBPK model with whole-body models.

Furthermore, in the context of IMD biomarker prediction, Toroghi et al [23] developed the dynamic parsimonious flux balance analysis method to depict changes in biomarker concentrations within the blood compartment quantitatively. They constructed a comprehensive mathematical whole-body model, incorporating 14 organs and the human serum, and integrated data from 237 serum metabolites sourced from HMDB [108] and Recon1 [17]. They applied PBPK modeling where the metabolite concentrations C_{organ} represented the concentration in each organ. Then, the concentration change of time t in an organ receiving arterial blood was modeled as,

$$\dot{C}_{organ} = \frac{Q_{organ}}{V_{organ}} \left(C_{art} - \frac{C_{organ}}{K_{organ}} \right) - R_{organ,i}, \quad (2.3)$$

where C_{art} is the concentration in the arterial compartment, V_{organ} is the organ volume, Q_{organ} is the blood flow rate to the organ, and K_{organ} is the organ-specific tissue partition coefficient. The lung receives blood from the venous blood compartment C_{ven} , and, hence, the change of a

metabolite concentration over time t in the lung can be modeled as,

$$\dot{C}_{lung} = \frac{Q_{lung}}{V_{lung}} \left(C_{ven} - \frac{C_{lung}}{K_{lung}} \right) - R_{lung,i}. \quad (2.4)$$

For all organs, R_i describes the Michaelis-Menten kinetic model, where C_i is the concentration of a specific metabolite i in human serum,

$$R_{organ,i} = \frac{V_{organ,i} C_i}{K_{organ,i} + C_i}. \quad (2.5)$$

This model represents the production or consumption rate of a metabolite in a specific organ. All organs are connected by the blood compartment, with the venous and arterial blood being modeled as,

$$\begin{aligned} \dot{C}_{art} &= \frac{Q_{lung}}{V_{lung}} \left(\frac{C_{lung}}{K_{lung}} - C_{art} \right) + Q_{ext}, \\ C_{ven} &= \frac{\sum_{i=1}^8 \frac{Q_i C_i}{K_i}}{Q_{lung}}, \\ Q_l &= Q_s + Q_p + Q_{sp} + Q_g + Q_{l,art}, \end{aligned} \quad (2.6)$$

where Q_{ext} describes the exogenous appearance rate of a metabolite. And the abbreviations s , p , sp , g , and l correspond to the stomach, pancreas, spleen, gut, and lung. Additionally, the input of metabolite concentrations for the liver model was given as,

$$C_{l,input} = \frac{\frac{C_s Q_s}{K_s} + \frac{C_p Q_p}{K_p} + \frac{C_{sp} Q_{sp}}{K_{sp}} + \frac{C_g Q_g}{K_g} + C_{art} Q_{l,art}}{Q_l}, \quad (2.7)$$

where $Q_{l,art}$ is the arterial blood flow rate to the liver. Since the entire model consisted of 237 human metabolites in each compartment (14 organs and the human serum), this led to a system of 3555 ODEs. This network of ODEs was then coupled with the human metabolic reconstruction Recon1 [17] through the transport reactions associated with the metabolic networks. Hence, the metabolic exchange rate depended on the kinetic parameters associated with these reactions and metabolite concentrations in the human serum. The validity of this coupled model and the application of dynamic parsimonious flux balance analysis was assessed on IMD biomarker predictions. Toroghi et al [23] showed that their model correctly predicts known biomarkers of ten IMDs of amino acid metabolism over time. They demonstrated that models integrating kinetic and constraint-based modeling can effectively simulate a range of IMDs. This suggests their potential application in treatment strategies and personalized medicine.

3

Data-based modeling in newborn screening

Data-based mathematical modeling techniques, including machine learning (ML) methods, are frequently employed in the medical domain. These methods can handle large amounts of data and learn valuable rules and patterns from the data. The strength of these methods comes from the availability of large data sets and the required computational power to train these methods. In the context of newborn screening, various supervised classification ML methods have been applied to newborn screening data to classify whether or not a newborn suffers from a condition, see section 2.2. In traditional newborn screening, the metabolite concentrations in dried blood samples collected from the newborns are automatically determined using an tandem mass spectrometry. Based on these estimates, newborn screening applies internal decision rules to classify the newborns into unremarkable and suspected diagnosis newborn screening profiles. The latter are further analyzed in confirmatory diagnostic steps (see A in Figure 3.1). This work introduces two additional strategies: the digital-tier strategy and the full data strategy, both of which are compared to the traditional newborn screening procedure. The digital-tier strategy simulates the scenario where ML classification methods are applied as an additional step after traditional newborn screening (see B in Figure 3.1). Consequently, it is only applied to suspected diagnosis newborn screening profiles. This approach allows building upon the existing newborn screening procedures and focuses solely on reducing false-positive screening results. The third strategy (see C in Figure 3.1) takes a different approach. It applies ML directly to the full data set without considering the screening

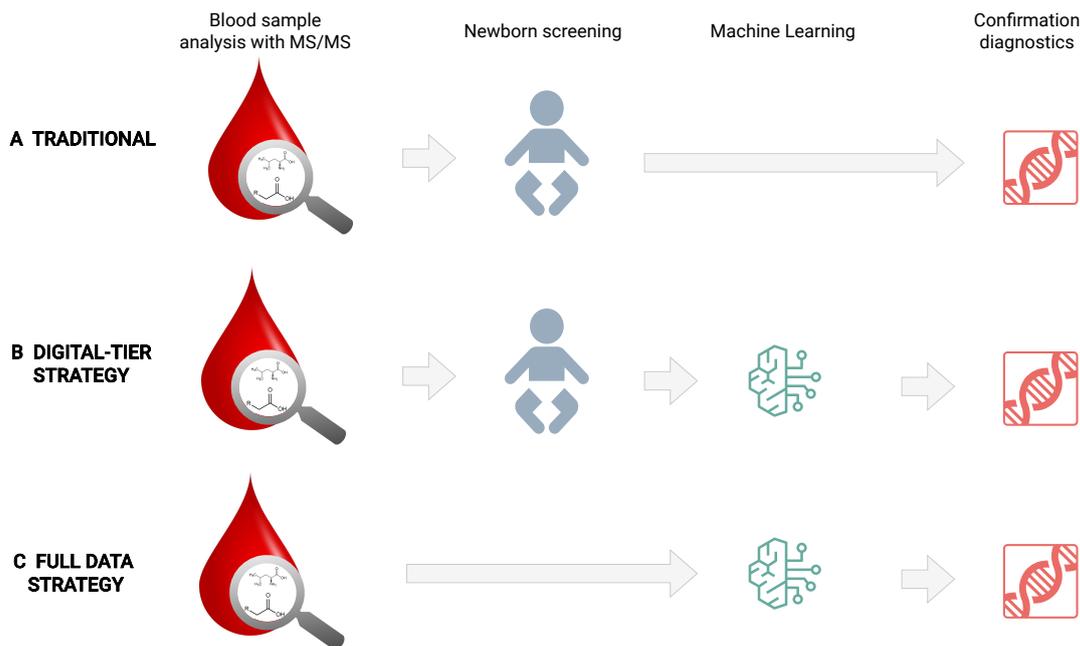


Figure 3.1: Data-based strategies for newborn screening. Comparison of (A) traditional newborn screening to (B) the digital-tier strategy and (C) the full data strategy.

results of traditional newborn screening. After both ML strategies, the suspected positive profiles are further analyzed to obtain confirmation of the results.

In this chapter, both presented data-based strategies are applied and evaluated for the diseases IVA and GA1, including respective data analysis and ML classification results. First, this chapter will introduce the applied data-based modeling methods, section 3.1. Then, the results of the ML classification for IVA, section 3.2, and GA1, section 3.3, will be presented, followed by a discussion of the results, section 3.4.

3.1 Methodological background

Data-based models solve problems and associated tasks by learning from problem-specific data, thus automating the process of constructing analytical models [109]. By this, a large number of data points can be analyzed for tasks that would be impossible or very time-consuming for humans. For these types of problems, we consider a data matrix $X \in \mathbb{R}^{m \times n}$ consisting of n data vectors $x = (x_1, \dots, x_m)^T$ which each have m features. For binary classification tasks, such as in disease prediction, there exists one feature $y \in \{0, 1\}$, the target label, which specifies the class C each data point x belongs to. For a classification problem with two classes C_0, C_1 , this could be,

$$y = \begin{cases} 0, & x \in C_0, \\ 1, & x \in C_1. \end{cases} \quad (3.1)$$

In a disease classification task, C_0 could correspond to the class of healthy individuals, while C_1 could correspond to the class of diseased individuals. The initial step in developing an algorithm for a classification task involves analyzing the underlying data.

3.1.1 Data analysis methods

Data analysis methods provide insights into the high-dimensional space of the data. For ML tasks on data sets with a large number of features, data analysis algorithms are often employed to reduce dimensionality and/or identify important features. One method to determine whether a feature is significantly different in two classes is the analysis of variance (ANOVA) [110].

ANOVA ANOVA is a statistical method applied to test whether the means μ_0, μ_1 of different classes C_0, C_1 are significantly different. ANOVA assumes that the data is normally distributed and is evaluated with a predefined p -value. Here, the null hypothesis H_0 assumes that the means of the different classes are equal,

$$H_0 : \mu_0 = \mu_1. \quad (3.2)$$

The alternative hypothesis H_A assumes that the means are different,

$$H_A : \mu_0 \neq \mu_1$$

Hence, the ANOVA can be applied to test whether the mean values differ between classes. In a disease classification task, features shown to be significantly different in healthy and sick patients can be valuable features for the classification algorithm. This method is frequently applied for feature selection in ML-based newborn screening, section 2.2 and Table B.1.

3.1.2 Machine learning classification methods

Classification algorithms are often supervised methods, meaning each data point x has a corresponding target label $y \in \{0, 1\}$ for a problem with two classes C_0, C_1 . ML classification algorithms

then attempt to detect a pattern for this classification by updating internal model parameters. The fitting of these model parameters is conducted to optimize the classification of training samples, where the model receives the true target label y as feedback in supervised learning.

Linear discriminant analysis Linear discriminant analysis (LDA) is a supervised method that is applied to maximize the separability between groups by projecting the data $X \in \mathbb{R}^{m \times n}$ onto a new axis to simultaneously maximize the distance between the class means μ_k and minimize the variation within the classes C_k [111]. The following description of the LDA is based on the presentations in [111], [112]. Suppose we have a two-class problem with classes C_0, C_1 and each data point $x \in C_0 \cup C_1$ is in the disjoint union of these classes. There are n_0 data points in class C_0 and n_1 data points in class C_1 , with $n_0 + n_1 = n$. Let μ_0 and μ_1 be the mean vectors of those classes. Let $x_j \in \mathbb{R}^m$ with $j = 1, \dots, n$ be the different data points. Then, we can consider any unit vector $v \in \mathbb{R}^m$ and $p_j = v^T x_j$ the projection of the x_j on the line with the unit vector, Figure 3.2.

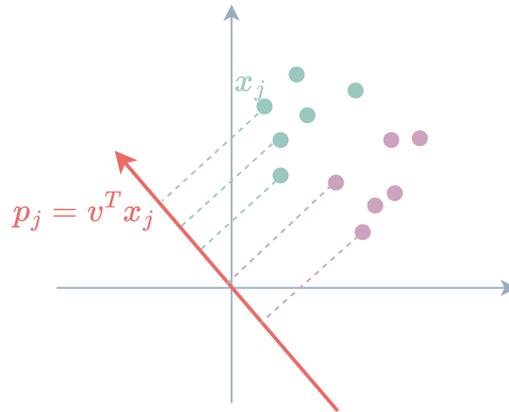


Figure 3.2: The projection p_j of data points x_j along unit vector v

This projection does not change the separability of the data points. Then, the projected means of the two classes can be written as,

$$\hat{\mu}_0 = \frac{1}{n_0} \sum_{x_j \in C_0}^{n_0} v^T x_j = v^T \mu_0, \quad (3.3)$$

and

$$\hat{\mu}_1 = \frac{1}{n_1} \sum_{x_j \in C_1}^{n_1} v^T x_j = v^T \mu_1. \quad (3.4)$$

A simple way to measure the distance between the points is to measure the distance between these

means. However, to obtain a good measure to separate groups of data, the LDA also considers the variances $\hat{\sigma}_0, \hat{\sigma}_1$ of the two classes,

$$\hat{\sigma}_0 = \sum_{x_j \in C_0} (p_j - \hat{\mu}_0)^2, \quad (3.5)$$

and

$$\hat{\sigma}_1 = \sum_{x_j \in C_1} (p_j - \hat{\mu}_1)^2. \quad (3.6)$$

Then, LDA aims to maximize the means of groups and minimize the variance within groups by maximizing $J(v)$,

$$J(v) = \frac{(\hat{\mu}_0 - \hat{\mu}_1)^2}{\hat{\sigma}_0 + \hat{\sigma}_1}. \quad (3.7)$$

The maximization of the means can be rewritten using the between-class scatter matrix

$M = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T \in \mathbb{R}^{m \times m}$. M is square, symmetric and positive semidefinite, with $\text{rank}(M) = 1$ and, hence, M only has one positive eigenvalue. Then,

$$\begin{aligned} (\hat{\mu}_0 - \hat{\mu}_1)^2 &= (v^T \mu_0 - v^T \mu_1)^2 \\ &= v^T (\mu_0 - \mu_1) \cdot (\mu_0 - \mu_1)^T v \\ &= v^T M v. \end{aligned}$$

The minimization of the variance can be reformulated utilizing the within-class matrix S_j for every class $c \in \{0, 1\}$, which is defined as $S_c = \sum_{x_j \in C_c} (x_j - \mu_c)(x_j - \mu_c)^T \in \mathbb{R}^{m \times m}$. The variance σ_c of each class can be formulated as

$$\begin{aligned} \sigma_c &= \sum_{x_j \in C_c} (p_j - \hat{\mu}_c)^2 \\ &= \sum_{x_j \in C_c} (v^T x_j - v^T \mu_c)^2 \\ &= \sum_{x_j \in C_c} v^T (x_j - \mu_c)(x_j - \mu_c)^T v \\ &= v^T \sum_{x_j \in C_c} (x_j - \mu_c)(x_j - \mu_c)^T v \\ &= v^T S_c v. \end{aligned}$$

Thus, the total within-class scatter is

$$\sigma_0 + \sigma_1 = v^T S_0 v + v^T S_1 v = v^T (S_0 + S_1) v = v^T S_w v, \quad (3.8)$$

where $S_w = S_0 + S_1$. The matrix $S_w \in \mathbb{R}^{m \times m}$ is also square, symmetric and positive semidefinite. In summary, the following optimization problem needs to be maximized,

$$\begin{aligned} \max_v \quad & \frac{v^T M v}{v^T S_w v} \\ \text{s.t.} \quad & \|v\| = 1. \end{aligned} \quad (3.9)$$

Logistic Regression Logistic regression (LR) is a discriminative method which models the posterior probability distribution $\mathbb{P}(y|x)$, of the target label y given the features x_i of a data point $x \in \mathbb{R}^m$. The following description of the LR method is based on the presentations in [113], [114]. The principles of LR rely on the linear regression method, which aims at modeling linear relationships in data by determining the best fitting linear model

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_m \cdot x_m, \quad (3.10)$$

through adjusting the regression coefficients b_0, \dots, b_m . However, for a classification problem, the target variables are discrete categorical classes and not continuous values. Hence, the LR aims to model the probability of a data point belonging to a class C_c . To obtain probabilities between 0 and 1, the logistic function,

$$f(x) = \frac{1}{1 + e^{-x}}, \quad (3.11)$$

is applied. Leading to the following probability distributions for class C_1 ,

$$\mathbb{P}(y = 1|x_1, \dots, x_m) = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + \dots + b_m x_m)}}, \quad (3.12)$$

and class C_0 ,

$$\mathbb{P}(y = 0|x_1, \dots, x_m) = 1 - \frac{1}{1 + e^{-(b_0 + b_1 x_1 + \dots + b_m x_m)}}. \quad (3.13)$$

To acquire the regression coefficients $b = b_0, \dots, b_m$, the Maximum Likelihood Estimation, a statistical method to estimate the parameters of a mathematical model, is applied. Therefore, a probability distribution for the class C_1 is assumed, and then a likelihood function L is defined. This function calculates the probability that given the features x_i and the LR model, $x \in C_1$ will be observed. The likelihood function L is then optimized to find the best fitting regression coefficients b that obtain the maximum sum of L for each data point in the training set. Hence,

the likelihood function L for LR is given as

$$L(b) = \prod_{y_i=1} \frac{1}{1 + e^{-bx_i}} \cdot \prod_{y_i=0} \left(\frac{e^{-bx_i}}{1 + e^{-bx_i}} \right). \quad (3.14)$$

Taking the logarithm and simplifying the function, the log-likelihood function l is obtained,

$$l(b) = \sum_{i=1}^m y_i bx_i - \log \left(\frac{1}{1 + e^{bx_i}} \right). \quad (3.15)$$

Several algorithms can be applied to maximize the log-likelihood by determining the best fitting regression coefficients b [114]. In this work, the limited-memory Broyden-Fletcher-Goldfarb-Shannon algorithm is applied, which is a quasi-Newton method to approximate the Broyden-Fletcher-Goldfarb-Shannon algorithm using only limited computer memory [115].

Ridge logistic regression Ridge logistic regression (RR) extends the LR method by penalizing the complexity of the resulting model. Therefore, an additional penalty parameter $\lambda > 0$ is added to the LR function, and an additional $\lambda \|b\|^2$ is added to the log-likelihood [64], [116]. In the RR optimization, coefficients are constrained by the square of the Euclidean norm of the coefficients. Hence, the penalized log-likelihood is

$$l_r(b) = l(b) - \frac{\lambda}{2} \sum_{k=1}^m b_k^2, \quad (3.16)$$

where $b \in \mathbb{R}^m$ are the regression coefficients and the penalty parameter λ controls the amount of shrinkage towards zero [117]. The intercept b_0 is omitted from the penalty term, resulting in an average predicted probability that equals the observed event rate [117].

Support vector machines Support vector machines (SVMs) attempt to find a separating hyperplane between two classes by transforming the features x_i of a data point $x \in \mathbb{R}^m$ into a higher dimensional space [12]. The following description of the SVM method is based on the presentations in [118], [119]. A linear hyperplane can be written as

$$w^T x + b = 0, \quad (3.17)$$

where w is the orthogonal vector to the hyperplane, and $b \in \mathbb{R}^m$ is the distance of the hyperplane from the origin along $w \in \mathbb{R}^m$. A margin can be defined for the data point $x \in \mathbb{R}^m$. A true positive x has a margin of $(w \cdot x + b) > 0$, and a true negative x has a margin $(w \cdot x + b) < 0$. Then, the

margin m_i , the distance between a data point with features x_i and the decision boundary, can be calculated, utilizing the Euclidean norm $\|w\|_2 = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$ of the orthogonal vector,

$$m_i = \frac{w^T x_i + b}{\|w\|_2}. \quad (3.18)$$

The data points x nearest to the decision boundary are called *support vectors*. The decision boundary of an SVM is defined as a linear combination of these support vectors. The class predictions of a linear binary classifier for a data point are then,

$$\hat{y} = \begin{cases} 0, & w \cdot x + b < 0, \\ 1, & w \cdot x + b \geq 0. \end{cases} \quad (3.19)$$

For a hard margin linear SVM classifier, the maximization of the margin leads to a quadratic, constrained optimization problem to find the optimal parameters w and b ,

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|_2^2 \\ \text{s.t.} \quad & y(w^T x + b) \geq 1 \quad \text{for} \quad x \in \mathbb{R}^m \end{aligned} \quad (3.20)$$

However, data points from different classes can often not be separated by a linear decision boundary. Therefore, kernel functions $K(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ are applied for non-linear SVM. These are used to transform the input data into a high-dimensional feature space where the data is linearly separable.

3.1.3 Experimental setup

The experimental setup describes how the ML algorithms were developed and optimized in this work. The experiments were performed using the programming language Python [120], and the Python libraries scikit-learn [121] (version 1.0.2) and scipy [122] (version 1.10.1).

Optimization In the optimization step, the internal parameters of the ML algorithms are updated to improve the model accuracy. Data imbalance is an important characteristic of newborn screening data due to the low prevalence of IMDs and the resulting low numbers of positive patients. Several methods exist to handle large data imbalances [123]. However, the newborn screening data lies on the spectrum of very high imbalance data sets. In this work, data imbalance is addressed by adapting the internal loss function of an ML method. This procedure will be shown as an example

in LR, where the logistic loss function L is used to calculate the loss,

$$L = \frac{1}{N} \sum_{i=1}^N -(y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)), \quad (3.21)$$

where y_i is the actual value of the target class and \hat{y}_i is the predicted probability of the target class. To overcome this data imbalance, a weighted logistic loss function L_w with class weight parameters w_0 and w_1 for the two classes is applied,

$$L_w = \frac{1}{N} \sum_{i=1}^N -(w_1(y_i \cdot \log(\hat{y}_i)) + w_0((1 - y_i) \cdot \log(1 - \hat{y}_i))). \quad (3.22)$$

In this work, the majority class weight parameter w_0 was set to 1. A grid search optimization strategy on the hyperparameters was applied on the training data set to find the optimal minority class weight parameter w_1 , which optimized the two objectives, maintaining 100% sensitivity S_n and maximizing specificity S_p ,

$$S_n = \frac{TP}{TP + FN} \quad \text{and} \quad S_p = \frac{TN}{TN + FP} \quad (3.23)$$

with true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP). For every algorithm and data set, the class weight parameter w_1 , which achieved 100% sensitivity and the highest specificity, was chosen.

Validation For model validation, the algorithms were first evaluated on training and validation set performance, with both data sets randomly split into an 80% training and 20% validation set. The classification performance on both data sets was then evaluated with the confusion matrix C ,

$$C = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}. \quad (3.24)$$

Subsequently, ten repeats of five-fold cross-validation were applied to evaluate the reliability of the results and estimate the effect of a single train and validation split.

3.2 Data-based newborn screening for IVA

This section is based on and extends the published work of Zaunseder et al [31] on ML-based classification for isovaleric aciduria (IVA) in newborn screening data. IVA is one of the target conditions in regular newborn screening and has faced two challenges in recent years. First, due to the increasing use of pivmecillinam in urinary tract infections in pregnant women, there is a rising number of false-positive screening results [24]. Second, since the start of newborn screening for IVA, individuals with an attenuated, possibly benign, disease variant (“mild” IVA [4], [25]) are identified. Hence, this section shows new data-based approaches to overcome these two challenges. It is divided into a description and analysis of the IVA data set, as well as the classification results on the two-fold classification (normal, IVA) and the three-fold classification (normal, mild IVA, classic IVA). Finally, the developed methods are validated on a previously unknown IVA test data set.

3.2.1 Newborn screening data set – IVA

Data lies at the core of data-based modeling, hence, the quality of the data is of great value for the development and performance of the methods. In the following, the data set composition and applied data cleaning steps are described in detail.

Newborn screening data composition and extraction

The anonymized data used for this work was screened at the newborn screening laboratory at UKHD between 2002 and 2021. The laboratory screens approximately 20% of the newborns in Germany (i.e., about 140,000 newborns per year) [5]. For this study, the data extraction was restricted to all profiles of newborns with suspected IVA and unremarkable newborn screening profiles. Hence, newborn screening profiles with confirmed or suspected positive diagnoses for other newborn screening conditions were excluded. The suspected IVA profiles contained newborns with subsequently confirmed mild IVA and classic IVA, as well as subsequently excluded newborn screening profiles, the false positives. The data extraction for the set of normal newborn screening profiles was restricted to first screenings of newborns of at least 32 weeks of gestation, at least 36 hours of life age at sampling, and unremarkable newborn screening reports (hereafter called ‘normal’). In total, the newborn screening data contained newborn screening profiles of 2, 237, 142 newborns, including 145 cases with suspected IVA. For every newborn screening profile, the data

set consisted of 60 variables, which contained 52 metabolite concentrations, and eight additional variables, such as birth weight, sex, gestational age, birth year, age at blood sample, age at sample arrival, and, if given, the suspected and the subsequently confirmed or excluded diagnosis.

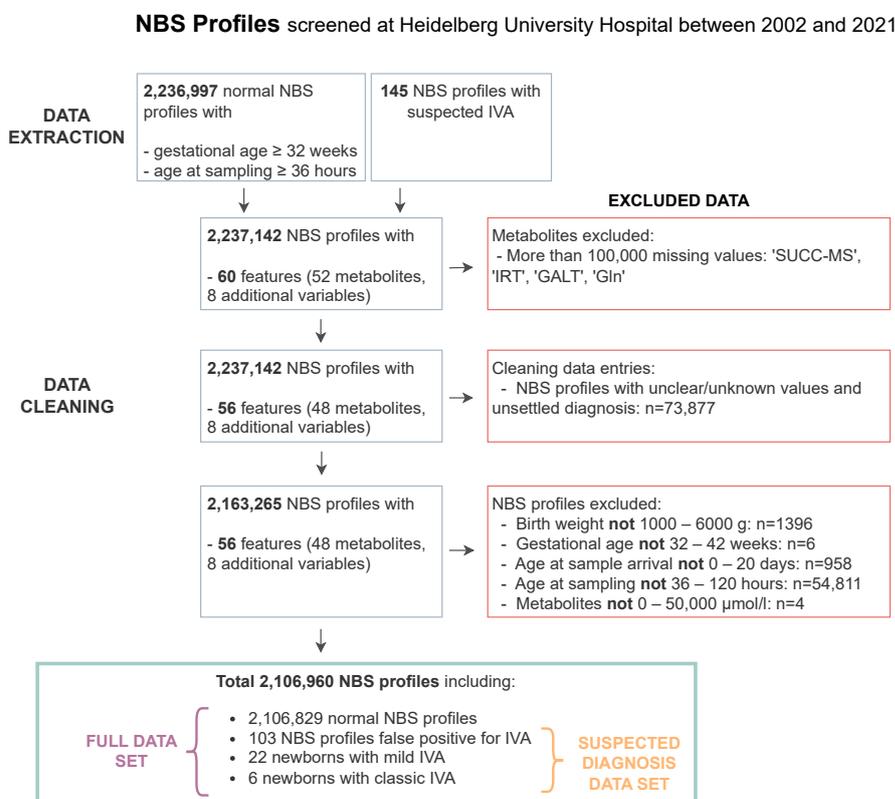


Figure 3.3: Data extraction and data cleaning flow chart for newborn screening data for IVA. Figure adapted from Zaunseder et al [31].

Newborn screening data cleaning

A data set's data quality can be reduced due to several reasons, such as manual input errors, errors from automatic scans, and data input format errors. Hence, data cleaning is of tremendous importance to ensure high data quality and remove artifacts within the data, Figure 3.3. The newborn screening data quality was evaluated based on the data set entries' completeness, correctness, and plausibility.

First, the completeness of the metabolite concentrations was investigated. Therefore, metabolic features with missing values were identified. After consulting with clinical experts, the metabolic features glutamine, succinylacetone, immune reactive trypsin, and galactose-1-phosphate uridyl-transferase were removed from the data set. These metabolites had more than 100,000 missing values, and no connection between them and IVA was reported. Second, the correctness of the

newborn screening data was investigated by evaluating the non-numeric data entries. Data entries with non-interpretable values such as '?', ' ', 'U', and 'unknown' were removed. Third, the plausibility of the data set entries was analyzed, and only data sets with plausible numeric values in the following ranges: Birth weight: 1000 – 6000 g; gestational age: 32 – 42 weeks, age at sampling: 36 – 120 hours, age at sample arrival: 0 – 20 days, and metabolites: 0 – 50,000 $\mu\text{mol/l}$ were included, Figure 3.3. For the algorithms' numerical readability, the categorical variable sex was encoded as '0' for female and '1' for male newborns. The total newborn screening data set, named the 'full data set', contained 2,106,090 newborn screening profiles. The full data sets comprised 131 cases with the suspected diagnosis IVA, including 28 subsequently confirmed IVA cases (22 mild, 6 classic) and 103 confirmed false positives. This subset was named the 'suspected diagnosis data set'.

3.2.2 Data analysis

Data analysis and statistical methods aim to reveal knowledge within a data set's feature space. For the IVA newborn screening data, first, the mean and standard deviation of the measured metabolites in dried blood samples were compared for the groups of normal and false-positive newborns as well as newborns with mild and classic IVA, Supplementary Table B.2. In newborn screening, isovalerylcarnitine (C5) is the primary biomarker to identify newborns with IVA. It is until now known to be the best metabolite to discriminate mild and classic IVA in larger cohorts [4], [25]. In the IVA newborn screening data, the mean of the measured C5 concentration was highest in newborns with classic IVA ($12.6 \pm 5.22 \mu\text{mol/l}$) and lowest in newborns with normal newborn screening profiles ($0.1 \pm 0.07 \mu\text{mol/l}$) but very similar in newborns with mild IVA ($2.6 \pm 1.16 \mu\text{mol/l}$) and false-positive newborns from newborn screening ($2.6 \pm 2.06 \mu\text{mol/l}$), Figure 3.4 and Supplementary Table B.2.

Then, an ANOVA on the full and the suspected diagnosis data sets was performed with p -value < 0.05 , Table 3.1. For both data sets, the ANOVA was performed with the target classes 'normal' and 'IVA'. The ANOVA confirmed the known biomarker C5 as a significant feature in the full newborn screening data set but not in the suspected diagnosis data set, Table 3.1. This underlined that the newborn screening profiles in the suspected diagnosis data set could not be accurately classified solely based on their C5 concentration, which is used in regular newborn screening. In the suspected diagnosis data set, other metabolites such as Trp and birth year were identified as significant features.

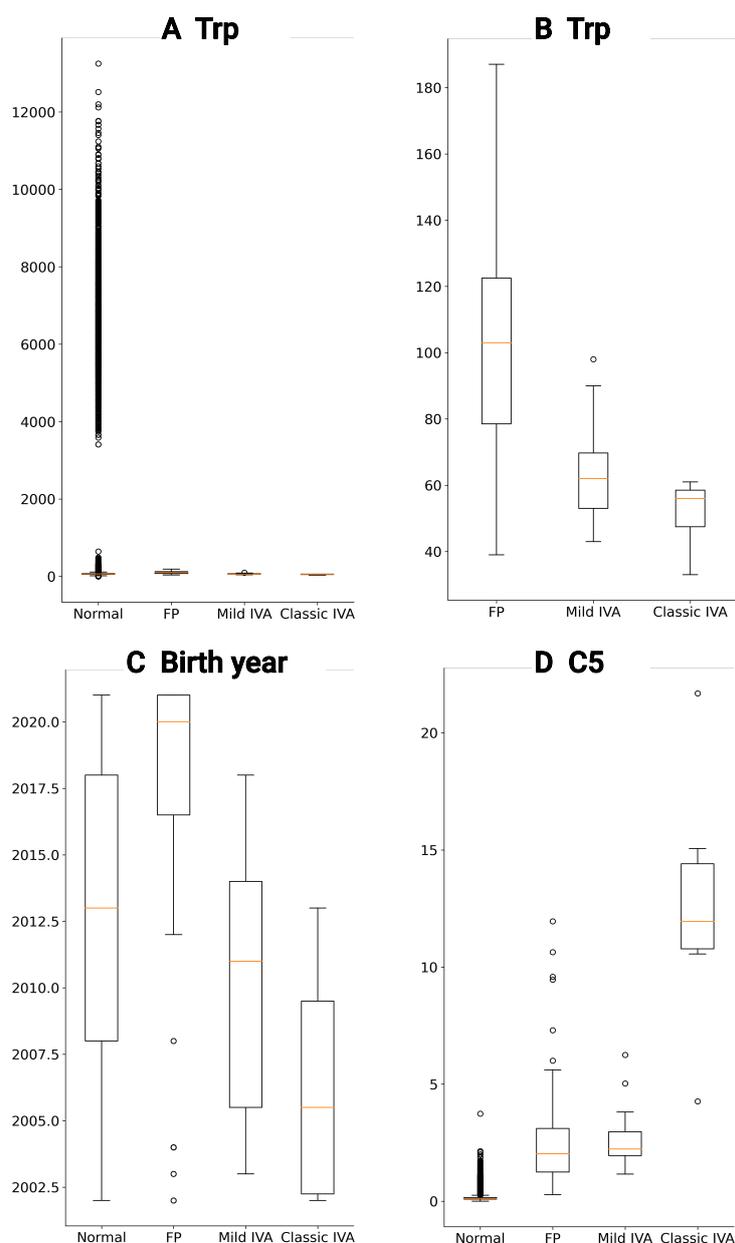


Figure 3.4: Box plot of newborn screening data for tryptophan (Trp) on (A) the full data set and (B) the suspected diagnosis data set and box plots of (C) birth year and (D) isovaleryl carnitine (C5) for the groups normal, false positive (FP), mild IVA and classic IVA. Figure adapted from Zaunseder et al [31].

The comparison of the mean and standard deviation of the metabolites showed that the groups of false positives had the highest mean value of Trp ($102.8 \pm 34.42 \mu\text{mol/l}$) from all groups. When comparing the box plots of the full and the suspected diagnosis data set, the normal and false-positive newborns showed higher Trp concentration values than newborns with IVA, Figure 3.4. Furthermore, the birth year was identified as a significant feature for both full and suspected diagnosis data sets and showed an accumulation of data points since 2016, Figure 3.4. This is

highly correlated with the increasing use of pivmecillinam, leading to an increase in false-positive screening results since 2016 [24].

Table 3.1: ANOVA results. All presented features have a p -value < 0.05 . All methods were applied to (A) the full newborn screening data set and (B) the suspected diagnosis data set. For ANOVA, five features with the largest F values with binary target variable normal or IVA are presented. Metabolite abbreviations can be found in Table 2.1.

(A) Full newborn screening data set		(B) Suspected diagnosis data set	
Feature	F value	Feature	F value
C5	97909.21	Trp	38.86
C16:1OH	28.17	Birth year	24.37
Age at blood sample	15.74	C14OH	22.55
Val	10.28	MeGlut	20.64
Birth year	8.43	His	18.22

Moreover, the newborn screening data was analyzed using data analysis methods to detect patterns and non-linear relationships within the data. The dimensionality reduction with LDA revealed a pattern within the newborn screening data. The data points of the resulting LDA dimensions were colored depending on their membership to the groups of normal, false-positive, mild IVA, and classic IVA newborn screening profiles. Figure 3.5 shows that a strict separation of clusters of the full data set was not possible based on the LDA dimensions. However, the individuals in each patient group were closely together. In particular, the newborns with classic IVA (stars) in the LDA dimensions had large negative values in the LDA 1 dimension compared to the other groups, Figure 3.5. Figure 3.5 shows that the suspected diagnosis data set was transformed into LDA dimensions, revealing three distinct clusters for groups of false positives, mild, and classic IVA. The group of newborns with classic IVA had values larger than 5.0 in the first LDA dimension, whereas the values for the other two groups were centered around 0.0. The second LDA dimension separated these two groups well, with an LDA 2 value of approximately 1.0.

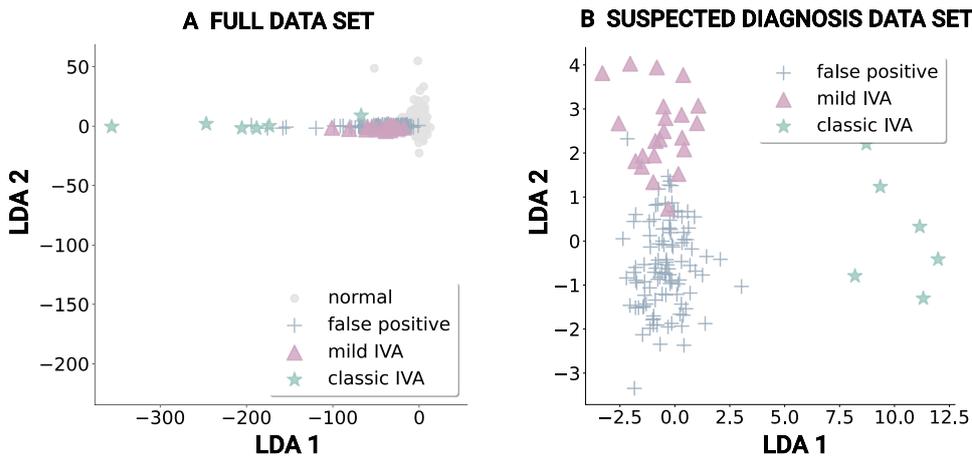


Figure 3.5: Dimensions from linear discriminant analysis (LDA) are presented. Normal newborn screening profiles are presented in gray (circle) and false-positive newborns in blue (cross), as well as newborns with mild IVA in purple (triangle) and with classic IVA in green (star) for (A) the full newborn screening data set and (B) the suspected diagnosis data set. Figure adapted from Zaunseder et al [31].

3.2.3 Machine learning for IVA classification

The application of ML methods for the classification of IVA from newborn screening data aims at detecting complex relationships within the large feature space of the data set to model the classification process enabling the classification of new and unknown data. For improving newborn screening for IVA, a new machine learning classification method based on LDA dimensions is presented in the following section.

LDA-LR Method

The application of adapted methods, such as constructing new features with arithmetic operations or combining several methods, has significantly enhanced the performance for several newborn screening conditions, surpassing the results of the application without the adaptation [61]. The data analysis with LDA revealed a good separation of false positives, mild, and classic IVA clusters in the data sets, Figure 3.5. The proposed new method, a combination of LR, RR, and SVM with LDA dimensions used as inputs are named LDA-LR, LDA-RR, and LDA-SVM. The workflow used for all three methods is presented here as an example of the LDA-LR method in Figure 3.6. First, the newborn screening data was split into 80% training and 20% validation data. Second, an LDA was trained on the training data, and then the LDA dimensions created by the model were used as input to train an LR classifier. The resulting model was evaluated to estimate the performance of

the training data. Third, the features of the validation data were transformed into LDA dimensions using the trained LDA and then these were used as input to the trained LR classifier. The labels of the validation data were then utilized to evaluate the LDA-LR method on the validation data set.

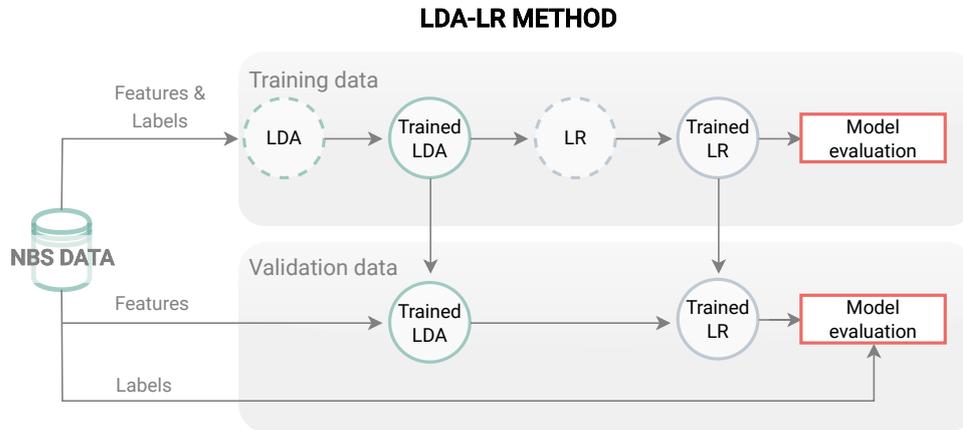


Figure 3.6: The workflow for the LDA-LR method shows the training process with features and labels of the training data and the evaluation of the trained methods on the validation data. Figure adapted from Zaunseder et al [31].

ML classification results

Each ML algorithm was applied to the full and suspected diagnosis data set using the digital-tier strategy, Figure 3.1. The algorithms were evaluated on different feature subsets with varying sizes, which contained significant features (p - value < 0.05) according to ANOVA, starting with one feature. After consultation with clinical experts, the birth year was excluded as a feature, due to its high correlation with the increasing use of pivmecillinam [24] leading to an increase in false-positive screening results. For ML classifications, the target class termed 'normal', included normal newborn screening profiles and newborn screening profiles, which were false positives in newborn screening. The target class termed 'IVA' included newborns with mild and classic IVA. The evaluation results are presented in Table 3.2.

LR was the only method obtaining 100% sensitivity on the full data set. However, this method only slightly reduced the number of false positives in the training and validation set compared to traditional newborn screening.

Table 3.2: Evaluation results of ML methods on (B) the full data set and (C) the suspected diagnosis data set compared to (A) the traditional newborn screening. Methods are applied to all 53 features, five features selected with ANOVA (Trp, C14OH, MeGlut, His, Asa), or LDA dimensions. The methods were evaluated by false negatives (FN) and false positives (FP) on the training and validation data set.

Method	Features (number)	Train		CV		Validation	
		FN	FP	S_n	S_p	FN	FP
(A) TRADITIONAL NEWBORN SCREENING							
Newborn screening	C5	0	73	100	99.9951	0	30
(B) FULL DATA SET							
LR	all (53)	0	65	100	99.9958	0	27
RR	all (53)	6	20,065	96.2143	99.8587	3	5,005
SVM	all (53)	1	35	97.9286	99.9975	0	9
(C) SUSPECTED DIAGNOSIS DATA SET							
LR	ANOVA (5)	0	35	100	99.9978	0	6
RR	ANOVA (5)	0	35	100	99.9979	0	6
SVM	ANOVA (5)	0	37	100	99.9979	0	6
LDA-LR	LDA	0	9	100	99.9981	0	10
LDA-RR	LDA	0	22	100	99.9985	0	12
LDA-SVM	LDA	0	99.2857	99.9985	12	0	10

For the suspected diagnosis data set, the digital-tier strategy, all considered methods could maintain 100% sensitivity in the training and validation set while reducing the total amount of false positives. Here, the LDA-RR and LDA-LR methods obtained the highest specificity, 99.9981% and 99.9985% compared to 99.9951% in traditional newborn screening. The evaluation of LR, RR, and SVM on the reduced feature set showed a reduction of the false-positive rate of 54 – 56% from the 103 false positives in traditional newborn screening. Accumulating the classification results of the training and validation data set, the LDA-LR method reduced the false-positive rate by nearly 70%, from 103 to 19 false positives.

3.2.4 Machine learning for IVA severity

Incorporating IVA into the newborn screening disease panel led to the identification of individuals with an attenuated, possibly asymptomatic, mild disease variant in contrast to classic IVA patients [4], [25]. In this work, ML methods were applied for the first time to investigate if these methods could be utilized to classify newborn screening profiles into three distinct groups: normal, mild IVA, and classic IVA. The normal group encompassed individuals with normal profiles and false-positive results within the full data set, while the suspected diagnosis data set solely comprised false-positive cases. For both data sets, an ANOVA was performed using three target classes, Table 3.3. This showed that on both data sets, C5 was the significant feature with the highest F values. C5 was followed by Trp, MeGlut, birth year, and C16:1OH in the suspected diagnosis data set, Table 3.3. The high C5 value correlated with the fact that classic IVA patients tend to have higher C5 values than mild IVA patients.

Table 3.3: ANOVA results. All presented features had a p -value < 0.05 . All methods were applied to (A) the full newborn screening data set and (B) the suspected diagnosis data set (B). For 3-class ANOVA, five features with the largest F values with target variable normal (false positives for (B)), mild IVA, and classic IVA are presented. Metabolite abbreviations can be found in Table 2.1.

(A) Full newborn screening data set		(B) Suspected diagnosis data set	
Feature	F value	Feature	F value
C5	90027.83	C5	57.75
C16:1OH	35.23	Trp	19.72
Tyr	17.29	MeGlut	17.00
Val	13.61	Birth year	13.30
C18:1OH	12.90	C16:1OH	13.06

The same experimental setup was used for the ML evaluation as described in the methodological background section 3.1. The algorithms performing best on the disease severity classification were evaluated with a mean confusion matrix of 100 independent runs comparing the predicted and confirmed diagnosis on both data sets, Figure 3.7. On the full data set, LR, on average, predicted 5.9 of 6 newborns with classic IVA and all newborns with mild IVA correctly. However, it falsely predicted on average 88.8 normal newborn screening profiles as mild IVA and 5.8 normal newborn screening profiles as classic IVA, Figure 3.7. On the other hand, applying RR to the suspected

diagnosis data set reduced the number of false positives on average to 17.43. At the same time, nearly all newborns with mild IVA (21 of 22) and newborns with classic IVA (5.7 of 6) were predicted correctly, Figure 3.7. With the presented LDA-LR method, the overall number of false-positive newborns was reduced to 9.85, while newborns with mild and classic IVA were predicted similarly well.

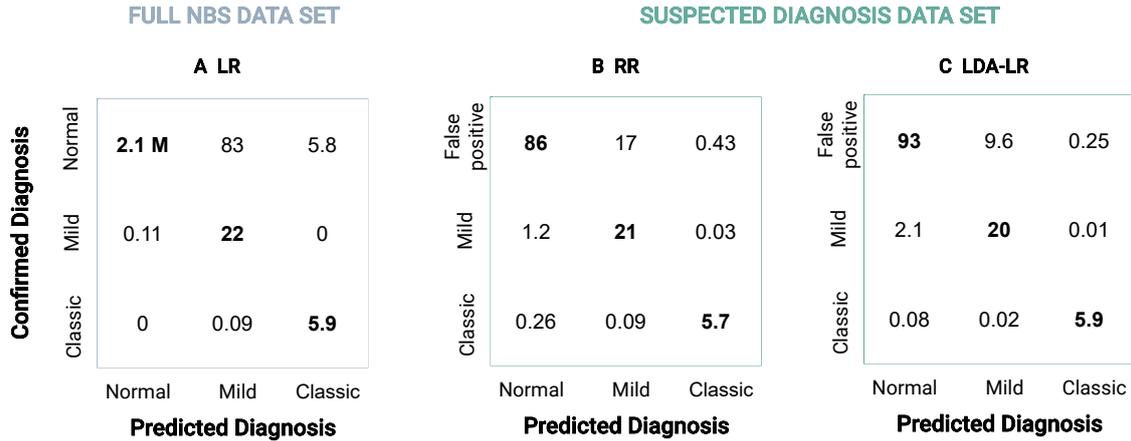


Figure 3.7: Mean confusion matrix of 100 independent runs of (A) LR classification on full data set, as well as (B) RR, and (C) LDA-LR classification on suspected diagnosis data set for normal, mild, and classic IVA are presented. Figure adapted from Zaunseder et al [31].

3.2.5 Proof of concept on IVA test data set

For the proof of concept of the developed methods for improved IVA classification, the methods were evaluated on an additional test data set from the UKHD. This test data set contained 73 newborn screening profiles screened in Heidelberg between 01.12.2021 and 31.12.2022 with suspected IVA. The test data set was curated based on the same data cleaning and exclusion criteria established in section 3.2.1 for the original suspected diagnosis data set, which will be called 'original data set' hereafter. Five newborn screening profiles had to be excluded, resulting in 68 newborn screening profiles being included in the test data set. This test data set contained one mild and one classic IVA patient, corresponding to 3% true positives in the test data set, compared to 20% (28 of 131 newborn screening profiles) in the original data set. For the proof of concept of the LDA-LR, LDA-RR, and LDA-SVM method, first, the trained LDA was applied to the test data set. Then, these data points were compared in the LDA dimension to the original data set, Figure 3.8.

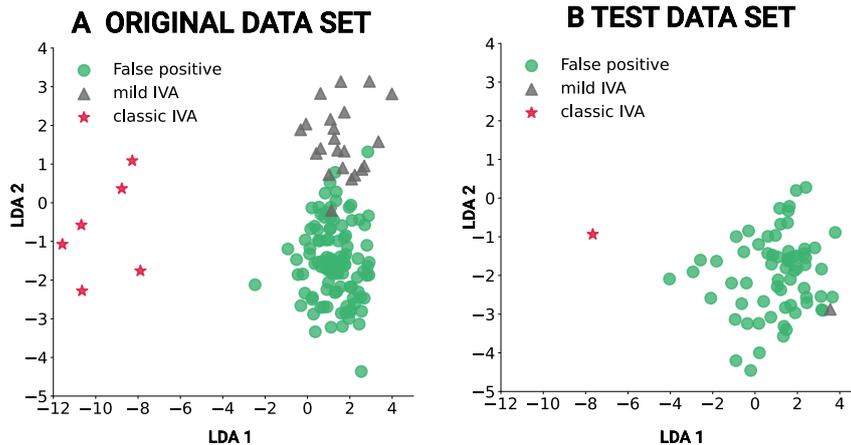


Figure 3.8: Dimensions from LDA labeled with false-positive newborns in green (circle), newborns with mild IVA in grey (triangle), and with classic IVA in red (star) for (A) the original data set and (B) the test data set.

This comparison showed that the classic IVA patient from the test data set had a low value (-7.66) in the LDA 1 dimension, such as the IVA patients in the original data set, which had a mean LDA 1 value of -9.64 ± 1.39 , Figure 3.8. Moreover, the false-positive newborn screening profiles in the test data set had a higher LDA 1 mean value (1.02 ± 1.57), which was in agreement with the mean value from the false positives in the original data set (1.30 ± 0.96). Also, the mild IVA patient showed a higher but similar LDA 1 value (3.55) compared to the mean of the LDA 1 value of the mild IVA patients in the original data set (1.59 ± 1.05). This led to a clear separation of the classic IVA patients and the other two groups based on the LDA 1 dimension in both data sets. However, the LDA 2 value was used to visually distinguish the mild IVA and false-positive newborn screening profiles in the original data set, with the mean of the LDA 2 value of mild patients being 1.61 ± 0.86 and of false positives being -1.53 ± 1.02 . This was different in the test data set. Here, the LDA 2 value of the false positives was similar (-1.97 ± 1.01), whereas the LDA 2 value of the mild IVA patient was -2.88.

Based on the trained LDA, the LDA-LR, LDA-RR, and LR-SVM methods were applied to the test data set. The results presented in Table 3.4 show that all three methods were able to classify the classic IVA patient and 74 - 80% of the formerly false-positive newborn screening profiles correctly. However, the mild IVA patient was not correctly identified and classified as unremarkable by all three algorithms.

Table 3.4: Evaluation results of ML methods on the test data set. Methods are applied to the LDA dimensions. The methods were evaluated by true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP) on the test set.

Method	Features (number)	TP	TN	FN	FP
TEST DATA SET					
LDA-LR	LDA dimensions	1	51	1	15
LDA-RR	LDA dimensions	1	49	1	17
LDA-SVM	LDA dimensions	1	53	1	13

This misclassification was probably due to the assignment of the LDA 2 value in the LDA. Reasons for the difference in LDA 2 value and, therefore, the incorrect classification could lie in the change of feature distributions over time. To analyze this effect the distributions of all measured metabolite concentrations in the full data set were compared between different birth years, Figure 3.9. Figure 3.9 presents box plots of the metabolite concentrations ($\mu\text{mol/l}$) of Asa and Trp per year from 2011 to 2021. For both metabolites, the mean values of year 2020 and 2021 (Asa: $2.93 \pm 2.1 \mu\text{mol/l}$, Trp: $112.25 \pm 30.8 \mu\text{mol/l}$) were significantly higher (Wilcoxon rank sum test: $p < 0.0001$) than the mean values from years 2005 to 2019 (Asa: $0.26 \pm 0.11 \mu\text{mol/l}$, Trp: $73.73 \pm 252 \mu\text{mol/l}$) of all newborn screening profiles. For further analysis, the mean values of the different groups of newborn screening profiles were compared to see whether the trend was also reflected in the different patient groups. A comparison of the Trp levels revealed that the levels in the test data set were higher in the false positives ($127.63 \pm 28.54 \mu\text{mol/l}$), mild IVA ($224 \mu\text{mol/l}$), and classic IVA ($123 \mu\text{mol/l}$) than compared to the original data set (false positives: $102.8 \pm 34.42 \mu\text{mol/l}$, mild IVA: $63.6 \pm 13.89 \mu\text{mol/l}$, classic IVA $51.55 \pm 9.78 \mu\text{mol/l}$). The metabolite concentration values of Asa revealed a similar trend, as the mean values in the test data set for the group of false positives ($3.76 \pm 2.12 \mu\text{mol/l}$), mild IVA patients ($0.8 \mu\text{mol/l}$), and classic IVA patients ($1.39 \mu\text{mol/l}$) were significantly (Wilcoxon rank sum test: Asa: $p < 0.0001$ and Trp: $p < 0.0001$) higher than the mean values in the original data set (false positives: $2.05 \pm 2.34 \mu\text{mol/l}$, mild IVA: $0.26 \pm 0.07 \mu\text{mol/l}$, classic IVA $0.21 \pm 0.08 \mu\text{mol/l}$). As the other two groups each consisted of one patient in the test data set, no statistical test was performed. Therefore, the training data from 2020 and 2021 should be particularly valuable for the ML method. However, although 55% (57/103) of the false-positive screened newborns were born between 2020 and 2021, no mild or classic IVA patients were screened in these two years. This unequal distribution of the

different patient groups in the data sets probably led to the ML algorithm's incorrect classification of the mild IVA patient.

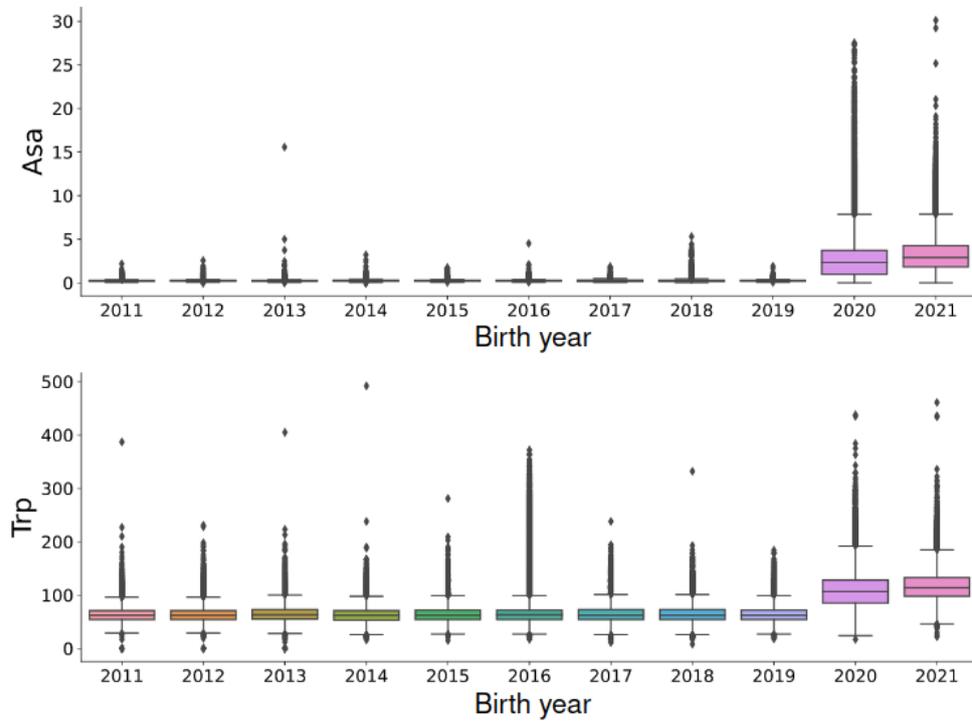


Figure 3.9: Boxplots of yearly concentration (measured in $\mu\text{mol/l}$) of argininosuccinate (Asa) and tryptophan (Trp) in the full data set compared between years 2011 and 2021.

3.3 Data-based newborn screening for GA1

Glutaric aciduria type 1 (GA1) has been one of the target conditions in regular newborn screening in Germany since 2005 [3]. The primary marker for GA1 is glutarylcarnitine (Glut). Newborn screening for GA1 is suffering from high numbers of false-positive screening results. Renal insufficiency due to congenital or acquired renal disease in newborns is one known factor for false-positive screening results since it delays the excretion of Glut via the kidneys, which increases the Glut concentration in plasma [60]. Hence, this work aims to improve newborn screening for GA1 by reducing the false-positive rate while maintaining 100% sensitivity using ML algorithms that consider several of the measured metabolite concentrations.

3.3.1 Newborn screening data set – GA1

The newborn screening data set for GA1 was obtained from the newborn screening laboratory at UKHD. The UKHD data protection officer checked the data set for anonymization. Further, the data extraction and evaluation were checked to be in accordance with the European GDPR. Similar to the newborn screening data set for IVA, section 3.2.1, this data set consisted of 60 variables (52 metabolite concentrations, birth weight, sex, gestational age, birth year, age at blood sample, sample arrival, and specification, and if given: suspected and the subsequently confirmed diagnosis). The data set was restricted to first screenings of newborns of at least 32 weeks of gestation, at least 36 hours of life age at sampling, and unremarkable newborn screening reports (hereafter called 'normal'). Additionally, all profiles of newborns with suspected GA1, subsequently confirmed ('GA1') or excluded ('false positives') were extracted. The newborn screening data set comprised newborn screening profiles of 1,055,885 newborns (including 603 cases with suspected GA1, 9 with confirmed GA1) born between 2014 and 2022. Data cleaning of the extracted data set was performed by defining the following ranges to exclude data sets with implausible values: Birth weight: 1000 – 6000 g; gestational age: 32 – 42 weeks, age at sample arrival: 0 – 20 days, and metabolite concentrations: 0 – 50,000 $\mu\text{mol/l}$. Furthermore, categorical values were converted to numerical values. The metabolite variables, Glutamine, succinylacetone, and immune reactive trypsin, were removed from the data set due to the large number of missing values ($>100,000$) as they were not measured continuously within the time frame. Finally, the full data set applied for the data analysis (hereafter "full data set") contained 1,025,953 newborn screening profiles (including 494 cases with the suspected diagnosis GA1, hereafter: "suspected diagnosis data set").

The suspected diagnosis data set contained nine subsequently confirmed GA1 cases, including six low excretors and three high excretors as well as 485 confirmed false positives. Note that the suspected data set was reduced disproportionately from 603 to 494 newborn screening profiles. The main reason for this was that the initial suspected GA1 data set included premature-born newborns and profiles with incomplete first or second screening cards.

Feature sets

After the data analysis, the feature set was prepared based on the following procedure. First, all input features used in the experimental setup were normalized between 0 and 1 to allow for direct comparability of the features. Then, feature selection techniques were applied. Hence, besides using all features, the algorithms were also applied to different data subsets that only consisted of features with significantly different values in the different groups of newborns (ANOVA p -value < 0.05). To improve newborn screening for GA1, each algorithm was applied to the full and suspected diagnosis data set (digital-tier strategy).

3.3.2 Data analysis

Statistical methods were applied to analyze the large GA1 data set. First, the mean and standard deviation of the measured metabolites in dried blood samples were compared for the three groups of normal newborn screening profiles, false-positive newborn screening profiles, and newborns with GA1, Supplementary Table B.3.

Table 3.5: ANOVA results with five features that have a p -value < 0.05 with binary target variable normal or GA1 are presented. All methods were applied to (A) the full newborn screening data set and (B) the suspected diagnosis data set. Metabolite abbreviations can be found in Table 2.1.

(A) Full newborn screening data set		(B) Suspected diagnosis data set	
Feature	F value	Feature	F value
Glut	17270.68	Glut	757.69
Hci	17.42	C10	10.07
C5	13.18	C14:1	5.71
Glu	12.55	C8	5.55
C18:1	10.07	C12	5.24

In newborn screening for GA1, Glut is the primary biomarker to identify newborns with GA1. A comparison of the Glut values revealed that the Glut mean was significantly higher in newborns with GA1 ($2.698 \pm 1.548 \mu\text{mol/l}$) than in newborns with suspected but not confirmed GA1 ($0.526 \pm 0.106 \mu\text{mol/l}$), and newborns with normal NBS profiles ($0.157 \pm 0.057 \mu\text{mol/l}$), Supplementary Table B.3. The ANOVA on the full and the suspected diagnosis GA1 data sets was performed with $p\text{-value} < 0.05$, Table 3.5.

ANOVA identified Glut as a significant feature in the full and the suspected diagnosis data set, Table 3.5. However, further significant features were present in the full data set, while different features were present in the suspected diagnosis data set, Table 3.5 and Figure 3.10.

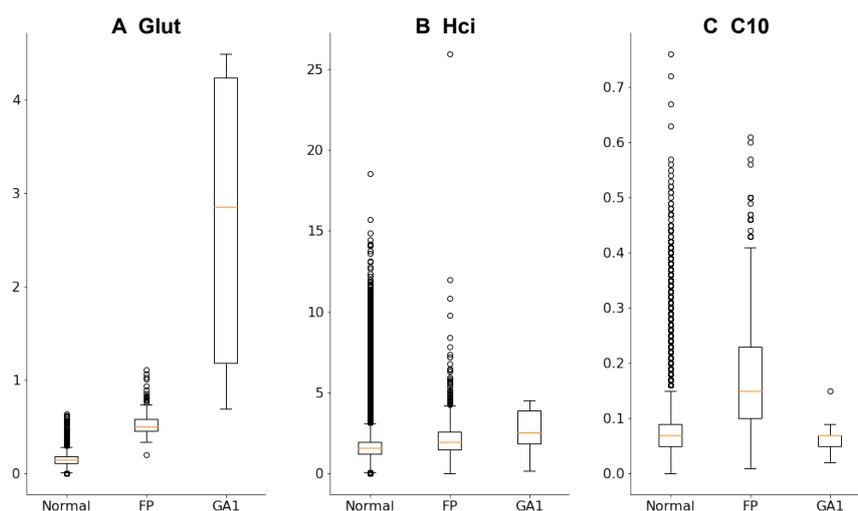


Figure 3.10: Boxplots of significant features (A) glutaryl carnitine (Glut), (B) homocitrulline (Hci), and (C) decanoyl carnitine (C10) compared for the three groups of normal newborn screening profiles, false-positive (FP) newborn screening profiles, and newborns with GA1.

The box plot of the Glut feature showed a distribution of high Glut values in newborns with GA1 and elevated but lower values for the suspected diagnosis profiles, Figure 3.10. The mean of decanoyl carnitine (C10) was higher in the suspected diagnosis data set ($0.19 \pm 0.11 \mu\text{mol/l}$) than in the normal data set ($0.07 \pm 0.03 \mu\text{mol/l}$) and in patients with GA1 ($0.07 \pm 0.03 \mu\text{mol/l}$), Table B.3 and Figure 3.10. The homocitrulline (Hci) values were on average higher ($2.25 \pm 1.77 \mu\text{mol/l}$) in the false positives than in the normal newborn screening profiles ($1.66 \pm 0.78 \mu\text{mol/l}$) but lower

than in newborns with GA1 ($2.72 \pm 1.41 \mu\text{mol/l}$), Figure 3.10.

Furthermore, dimensionality reduction with LDA was applied to both data sets to reveal patterns within the large newborn screening data set, Figure 3.11. On the full data set, three groups of normal, false positive, and newborns with GA1 could be visually identified, however, these groups were overlapping and did not offer a clear distinction, Figure 3.11. Both groups overlapped completely on the suspected diagnosis dataset, and no visually distinct clusters could be identified, Figure 3.11.

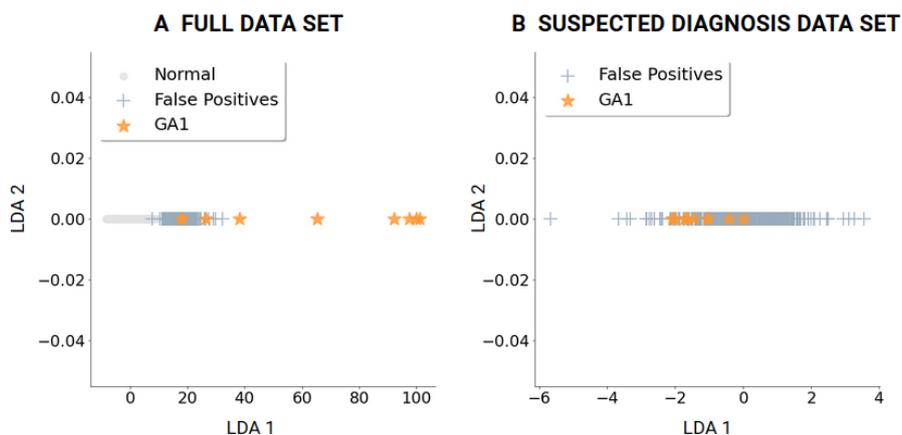


Figure 3.11: Dimensions from LDA dimensionality reduction presenting normal newborn screening profiles in gray (circle) and false positive profiles in blue (cross) as well as newborns with GA1 orange (star) for (A) the full newborn screening data set and (B) the suspected diagnosis data set.

3.3.3 Machine learning for GA1 classification

The same experimental setup, optimization, and validation strategies as in the analysis of the newborn screening data for IVA, described in section 3.1.3, were applied. The ML algorithms under investigation were LR, RR, and SVM, the best-performing algorithms in comparative studies on ML-based newborn screening, section 2.2. Table 3.6 presents the classification results of these three algorithms on the full and suspected diagnosis data (digital-tier strategy).

Overall, for both the full data set and the suspected diagnosis data set, the best methods could decrease the number of false positives while having no false negatives in the respective training and validation data set, Table 3.6. For most algorithms, basing the evaluation on Glut and C10 led to the best results. On both data sets, the RR method performed worse than LR and SVM in terms of overall false-positive rate reduction. SVM reduced the false positives by 82.06% on the full data set and by 93.4% on the suspected diagnosis data set while having no false negatives.

The LR using the features Glut and C10 presented the best-performing algorithm with respect to specificity while minimizing the number of false negatives for both data sets. LR reduced the number of false positives by 93.61% on the full data set and by 95.05% on the suspected diagnosis data set. Although the LR and SVM algorithms showed no false negatives on a randomly stratified split of the training and validation set, 5-fold cross-validation with stratified splitting revealed that neither algorithm achieved 100% sensitivity.

Table 3.6: ML classification results for GA1 comparing (A) traditional screening results, (B) the full data set, and (C) the suspected diagnosis data set. The best-performing features were selected with ANOVA. The methods were evaluated by sensitivity S_n and specificity S_p calculated from the mean results of ten repeats of 5-fold cross-validation (CV), as well as the number of false positives (FP) and false negatives (FN) (real numbers are rounded up) for training and validation set. For the suspected diagnosis data set, these evaluations were calculated based on the full data set to allow comparability between both data sets.

Method	Features	Train		CV		Validation	
		FN	FP	$S_n(\%)$	$S_p(\%)$	FN	FP
(A) TRADITIONAL NEWBORN SCREENING							
Newborn screening	Glut	0	388	100	99.9527	0	97
(B) FULL DATA SET							
LR	Glut, C10	0	26	99.11	99.9962	0	5
RR	Glut, C10, C14:1	1	583	92.44	99.25	2	152
SVM	Glut, C10	0	73	92.44	99.9987	0	14
(C) SUSPECTED DIAGNOSIS DATA SET							
LR	Glut, C10	0	19	86.67	99.9987	0	5
RR	Glut, C10	0	51	84.67	99.9965	0	18
SVM	Glut, C10	0	28	90.89	99.9981	0	5

The LR methods for the full and suspected diagnosis data set could be explicitly determined. On the full data set, the best parameters for LR were $\beta_0 = -9.03$, $\beta_1 = 68.16$ and $\beta_2 = -9.04$ with \tilde{x} the normed 'Glut' and 'C10' concentrations of newborn x leading to the following decision equation,

$$\mathbb{P}(GA1 \text{ profile} | X = x) = \frac{1}{1 + e^{-(-9.03 + 68.16\tilde{x}_1 - 9.04\tilde{x}_2)}}. \quad (3.25)$$

On the suspected diagnosis data set, the best parameters for LR were $\beta_0 = -0.65$, $\beta_1 = 8.72$

and $\beta_2 = -6.77$ with \tilde{x} the normed 'Glut' and 'C10' concentrations of newborn x leading to the following decision equation,

$$\mathbb{P}(GA1 \text{ profile} | X = x) = \frac{1}{1 + e^{-(-0.65 + 8.72\tilde{x}_1 - 6.77\tilde{x}_2)}}. \quad (3.26)$$

Since the cross-validation results for the algorithms did not obtain 100% sensitivity, another grid search was applied for this. A 5-fold stratified cross-validation over the applied class weight parameter w_0, w_1 was used. The parameter w_0 corresponding to the majority ('healthy') class was set to $w_0 = 1$. The optimal parameter w_1 corresponding to the minority class was searched in the interval $I_F = [1, 50000]$ for the full data set and in the interval $I_S = [1, 500]$ for the suspected diagnosis data set. The best-performing methods in terms of highest mean sensitivity and specificity are presented in Table 3.7.

Table 3.7: ML classification results optimized for 100% sensitivity for GA1 on (B) the suspected diagnosis data set compared to (A) the traditional newborn screening results. The methods were evaluated by sensitivity S_n and specificity S_p calculated from the mean results of ten repeats of 5-fold cross-validation (CV), as well as the number of false positives (FP) and false negatives (FN) (real numbers are rounded up) on the training and validation set. For the suspected diagnosis data set, these evaluations are calculated based on the full data set to allow comparability between both data sets.

Method	Features	Train		CV		Val	
		FN	FP	$S_n(\%)$	$S_p(\%)$	FN	FP
(A) TRADITIONAL NEWBORN SCREENING							
Newborn screening	Glut	0	388	100	99.9527	0	97
(B) SUSPECTED DIAGNOSIS DATA SET							
RR	Glut, C10	0	184	100	99.9809	0	51
LR	Glut, C10	0	112	100	99.9889	0	35
SVM	Glut, C10	0	129	100	99.9870	0	35

On the full data set, no method achieved 100 % sensitivity. Hence, only the results of the suspected diagnosis data set are presented. By increasing the class weight w_1 to values between 180 and 183, all three methods—LR, RR, and SVM—achieved 100% sensitivity in cross-validation. However, this weight adjustment also decreased the specificity of all three methods. LR classification had the best results with 114 false positives overall, while SVM had 135 false positives and

RR had 195, as shown in Table 3.7. Consequently, the total false positives were reduced by 59.8% with RR, 76.5% with LR, and 72.4% with SVM.

3.3.4 Proof of concept on GA1 test data set

For the proof of concept an independent test data set, including data from January 2022 until October 2023, was extracted from the newborn screening laboratories in Heidelberg. This test data set consisted of 257,414 newborn screening profiles, including 268 false-positive screened newborns and two patients with GA1. The test data set was curated based on the same data cleaning and exclusion criteria established for the original GA1 data set. This resulted in a test data set of 236,210 newborn screening profiles, including 235 false positives and two patients with GA1.

Table 3.8: Proof of concept of ML classification results for GA1 compared from (A) traditional screening results, (B) the full data set and (C) the suspected diagnosis data set. The best-performing features were selected with ANOVA. The methods were evaluated by sensitivity S_n and specificity S_p , as well as the number of false positives (FP) and false negatives (FN) on the test data set. For the suspected diagnosis data set, these evaluations were calculated based on the full data set to allow comparability between both data sets.

Method	Features (number)	Test data set			
		FN	FP	$S_n(\%)$	$S_p(\%)$
(A) TRADITIONAL NEWBORN SCREENING					
Newborn screening	Glut	0	235	100	99.90
(B) FULL DATA SET					
LR	Glut, C10	0	16	100	99.99
(C) SUSPECTED DIAGNOSIS DATA SET					
LR	Glut, C10	0	18	100	99.99

The LR presented the best classification results on the training and validation data set, Table 3.6. Hence, the test data was only evaluated with the trained LR models. The results are presented in Table 3.8. On the test data set, the LR classification method, which was initially trained on the full data set, Eq. (3.25), reduced the false positives by 93.19% compared to traditional newborn screening (from 235 to 16 FP results), Table 3.8. On the test data set, the LR classification method, which was initially trained on the suspected diagnosis data set, Eq. (3.26),

reduced the false positives by 92.34% compared to traditional newborn screening (from 235 to 18 FP results), Table 3.8. Both algorithms identified the patients with GA1 correctly, which led to 100% sensitivity, Table 3.8. Moreover, the LR method optimized to achieve 100% sensitivity in cross-validation, Table 3.7, classified all newborns with GA1 patients correctly. However, it classified 115 newborn screening profiles incorrectly as GA1, which was a false-positive reduction of only 51%.

3.4 Summary and outcome

Newborn screening is marked by high data imbalance due to the low prevalence of the investigated diseases. At the same time, there is a strong requirement for high sensitivity and specificity in newborn screening methods and recently, a relatively high number of false-positive screening results challenges newborn screening for IVA and GA1. Hence, this chapter evaluated the application of data analysis and machine learning methods as potential digital-tier and full data strategies to improve the specificity of newborn screening for IVA and GA1.

3.4.1 Digital-tier strategy in newborn screening

The current standards in newborn screening for IVA and GA1 are well developed and obtain 100% sensitivity. A large number of false-positive screening results hampers newborn screening for IVA and GA1. The developed digital-tier approach simulates the scenario where the ML classifier is used after traditional newborn screening. For both diseases, this strategy showed the best results regarding sensitivity and specificity. For IVA, the proposed LDA-LR and LDA-RR methods achieved the best sensitivity, specificity, and most reliable results, leading to a false-positive rate reduction of nearly 70%. For GA1, the LR method achieved the best results regarding sensitivity, specificity, and reliability, leading to a false-positive rate reduction of over 76%. This work showed that the grid search for optimal class weight parameters was essential for increasing specificity while maintaining 100% sensitivity. These high sensitivity results are important as newborn screening needs to identify all newborns suffering from a disease. The proposed algorithms relied on several metabolites to improve the disease classification. These identified metabolites present a starting point for future metabolic research to identify metabolites that second the current disease biomarkers.

Moreover, additional costs and effort accompany false-positive newborn screening results, as these results entail the information transmission about the suspicious newborn screening result to the local hospital and the families by a physician, clinical evaluation of the newborn, and sampling for the confirmatory diagnostics by a pediatrician, as well as costs for metabolic (and genetic) analyses. Hence, the increased specificity, i.e., reduction of false positives, could reduce harm to the infants and their families and support the cost-effectiveness of newborn screening. The results for both IVA and GA1 suggest that adding ML methods as digital-tier to traditional newborn screening, similar to implementing a biomarker-based two- or multiple-tier strategy, can benefit

newborn screening.

3.4.2 Method validation

Method validation is particularly important in data-based modeling for newborn screening due to the data imbalance and low number of patients with a disease. Despite its vast size of over one million screening profiles, the newborn screening data set was marked by the low prevalence of the conditions. Specifically, there were only 28 (22 mild, 6 classic) newborns with IVA and 9 (3 high excretor, 6 low excretor) newborns with GA1. Usually, having more data samples can always improve a model. However, there is no fixed minimum number of samples required to train a successful classification system. It is highly dependent on the underlying task and, notably, the significant variability within the data. These imbalanced newborn screening data sets can pose significant challenges. One such challenge is that the provided data set may not sufficiently cover the entire feature space, potentially leading to skewed results. The developed methods for both diseases IVA and GA1 were tested on additional unknown test sets from 2022 and 2023. The results showed that the models achieved the same sensitivity and specificity for the classic IVA, GA1, and false-positive newborn screening profiles as in the original data sets. However, the mild IVA patient was not classified correctly, although identifying all affected individuals is essential for successful newborn screening. Therefore, these results highlighted the importance of model validation on external data sets and the need for a larger training data set. The test set results further revealed the complexity of the newborn screening data set. Newborns with mild and classic IVA were not evenly distributed over the birth years, and existing time trends were revealed. Hence, additional data from recent years with classic IVA patients are required. The evaluation of the test data also showed how important it is that the test data and the training data have similar feature distributions, such as the Trp and Asa values, which showed significantly different values in the test data set.

In the future, the proposed methods should undergo additional validation on independent data sets from recent years to ensure their validity. Moreover, based on the developed methodology, promising avenues for future research open up. For instance, other newborn screening conditions with high false-positive rates can be evaluated, potentially leading to significant advancements in the field.

4

Explainable AI (XAI) in newborn screening

This chapter describes the application of explainable artificial intelligence (XAI) for model interpretation and pattern recognition in newborn screening for IVA. This chapter is based on and extends the published work Zaunseder et al, Copyright ©2023, IEEE [32]. The previous chapter, chapter 3, showed how applying ML methods could improve newborn screening classification for IVA and GA1. However, the applied methods did not include ensemble and deep learning methods, which showed remarkable results for disease prediction tasks such as diabetes prediction [8] and prediction of colorectal cancer [124] as well as for newborn screening [12]. The drawback of these complex ML classifiers is their black-box nature, implying a lack of or reduced intuitive understanding of the classifiers' output. In the medical domain, there is a high demand for medical professionals to comprehend how and why AI methods work and what their decisions are based on. Clinicians require explanations to interpret a system's confidence and estimate the reliability of the system's decision [16]. Through exploratory interviews, Tonekaboni et al [28] identified that the following features make a model explainable for clinicians:

- provides information about the model's context,
- understands and rationalizes the predictions,
- augments or supplements clinical ML systems.

The absence of interpretability could potentially hinder the use of complex algorithms, thereby limiting their potential benefits to the medical field. This leads to a growing demand for integrating XAI methods in ML-based medical decision-making to make them trustworthy, transparent, interpretable, and, therefore, explainable to human experts. There are different examples in the literature where ML methods were implemented to work alongside clinicians, such that they can review each other's steps, which contributes to more explainability [16]. In particular, in the area of medical imaging, there exist several promising approaches for the integration of XAI methods, such as a method to classify cardiac pathology [125] or an initial XAI-enabled software for breast core biopsies [126]. In newborn screening, methods to interpret ML algorithms have been mainly applied to inherently interpretable models by analyzing model internal parameters such as built-in decision functions [68], [71] and discriminatory thresholds [12], [71]. These were used to identify important metabolite concentrations. However, for ML-based newborn screening, XAI methods have yet to be applied.

In this chapter, first the applied ensemble and deep learning methods as well as the XAI algorithms are described. Then, the results of IVA classification, model interpretation, and pattern recognition are presented. Finally, the methods are discussed regarding their applicability and implications.

4.1 Methodological background

The methodological background describes the known classification and XAI methods applied for newborn screening in IVA. The experiments were performed using the programming language Python [120] and the Python libraries xgboost (version 1.5.0) [127], pytorch (version 1.12.0) [128], shap (version 0.40.0) [129], captum lime (version 0.5.0) [130], and scikit-learn (version 1.0.2) [121].

4.1.1 Ensemble and deep learning methods

Extreme Gradient Boosting Trees

The scalable end-to-end tree boosting system extreme gradient boosting (XGBoost) was introduced by Chen et al. [131]. It belongs to the ensemble methods, which combine the power of several base learners to obtain one powerful algorithm. Two widely applied methods to ensemble the learners' results are Bagging and Boosting. XGBoost uses the latter, where each base learner aims to reduce

the error of its predecessor. In the following, the XGBoost method is presented based on the work of Chen et al [131]. A tree ensemble model is described to consist of K additive functions to predict the output for a data point i with m features $\mathcal{D} = \{(x_i, y_i)\}$ where $x_i \in \mathbb{R}^m, y_i \in \mathbb{R}$,

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F}. \quad (4.1)$$

Here, $\mathcal{F} = \{f(x) = w_{q(x)}\}$ with $q : \mathbb{R}^m \rightarrow T$ and leaf weight $w \in \mathbb{R}^T$ is the space of regression trees with T the number of leaves in a tree. Each f_k corresponds to an independent tree structure q that maps an example to the corresponding leaf index with leaf weight w [131]. To learn the corresponding parameters the regularized objective \mathcal{L} is minimized,

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \quad (4.2)$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ penalizes the complexity of the model to avoid overfitting and l is a differentiable convex loss function [131]. As this objective cannot be optimized using traditional methods in Euclidean space, it is trained in an additive manner [131]. For the prediction of the i^{th} instance at the t^{th} iteration, $\hat{y}_i^{(t)}$, the term f_t needs to be added to minimize the following objective,

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (4.3)$$

and by this adding greedily the tree f_t that minimizes the loss in Eq. (4.2). This problem can be solved using exact greedy or approximated algorithms, both available in XGBoost [131]. XGBoost improves general gradient boosting by using parallel and distributed computing to enable faster model exploration and offers efficient memory usage [131]. Furthermore, introducing regularization generates simpler base learners, increasing speed and interpretability. XGBoost applies a distributed weighted quantile sketch algorithm to handle weighted data effectively [132]. The algorithm can be applied to different data types and has achieved good results for classification tasks on tabular data [133].

Neural Networks

Neural networks (NNs) are part of the deep learning methods, a sub-field of AI. Feed-forward NNs try to mimic the signaling processes in the human brain and exclusively pass information forward through multilayer perceptrons [134]. The feed-forward architecture was applied in this study, since the newborn screening profiles are independent of each other. Each layer in the network consists of nodes, also known as neurons or units. The connections between these nodes are characterized

by associated weights, defining the network's ability to process input data through multiple layers of computation to yield an output. Information propagation in a feed-forward NN involves two main steps: First, the weighted sum of inputs and the application of an activation function f . The output z_j of a layer j applied to an input vector $x \in \mathbb{R}^n$ can then be described as

$$z_j = f \left(\sum_{i=1}^n w_{ij} \cdot x_i \right) + b_j \quad (4.4)$$

where $w_{i,j} \in \mathbb{R}$ and $b_j \in \mathbb{R}$ are the weights and biases of each layer. At the start of the training process, these are initialized with random, often small numbers. Common activation functions, are the $ReLU(y) = \max(0, y)$ or softmax function $s(y) = \frac{e^y}{\sum_i^n e^{y_i}}$. In the second step, the data is passed through the NN to obtain an output prediction. Based on a chosen loss function \mathcal{L} , the difference between predicted and actual value is quantified. In this work, the binary cross entropy loss function \mathcal{L}_C is applied for N training samples,

$$\mathcal{L}_C = -\frac{1}{N} \sum_{i=1}^N \hat{y}_i \cdot \log(y_i) + (1 - \hat{y}_i) \cdot \log(1 - y_i), \quad (4.5)$$

where $y_i \in \{0, 1\}$ is the actual binary class label, and $\hat{y}_i \in [0, 1]$ is the predicted probability of the class label [135]. Following the forward propagation phase, backpropagation is subsequently employed to compute the gradients of the loss concerning the model parameters. This involves the application of the chain rule of calculus, facilitating the systematic calculation of gradients by backward error propagation throughout the network. After backpropagation, the optimization process continues with gradient descent to update the weights w and biases b in a direction that minimizes the loss function \mathcal{L} . The learning rate, a crucial hyperparameter, governs the magnitude of these updates. Several well-established variants of gradient descent exist, including stochastic gradient descent [136], root mean squared propagation [137], and adaptive optimization algorithms like Adam [138]. The neural network architecture encompasses details regarding the number of layers, the arrangement of neurons within each layer, the specific activation functions denoted as f , and the optimizer employed. These network parameters are strategically selected by a comprehensive hyperparameter optimization process [139]. This iterative optimization ensures an informed choice of hyperparameters, enhancing the neural network's overall performance and generalization capabilities.

4.1.2 Explainable artificial intelligence methods

XAI methods generate interpretable, intuitive, and ideally human-understandable explanations of an AI’s decision making process [140]. For this, different types of XAI algorithms have been developed in recent years. Das et al [140] suggest three general categories to distinguish these algorithms: 1. The scope of the algorithm identifies where a method is focusing on, either *locally* mainly focusing on explaining the classification of individual instances, or *globally* to understand the model as a whole. 2. The methodology, comparing the algorithmic approach of the methods with either *BackProb* by depending on gradients that are backpropagated from output to input or *perturbation*, which is based on (forced) changes of features in the data input. 3. The usage depends on how the XAI method is developed, either *intrinsic*, where it is inherent in the architecture of the method, or *Post-Hoc*, which is independent of the model and can be applied to already trained ML methods. In this work, the latter type of XAI methods were applied, as they are classification model-agnostic methods, which enable the comparison of interpretations of different AI models.

Shapley additive explanations

From the XAI methods, the Shapley additive explanations (SHAP) method developed by Lundberg et al [129] is widely applied in several areas. In the medical domain, it was, for instance, applied to melanoma detection in skin imaging [141]. In the following, the SHAP method is explained based on the work of Lundberg et al [129]. The method derives from Shapely values in game theory, wherein the features represent players in a coalition game used to compute individual feature contributions. The method assigns payouts to players (features) depending on their contribution to the total payout. These players then cooperate in a coalition and receive a profit from this cooperation. The Shapely value is the average contribution of a feature from all possible coalitions. To achieve this, suppose we have a value function v where $v_x(S)$ is the prediction for feature values in a feature subset $S \subseteq F$, where F is the set of all features. The value function is the payout function for a coalition of players, which in this case are represented by the features. Then, the exact Shapley value of player j , which are the features in the classification task, is defined as,

$$\phi_j(v) := \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - 1 - |S|)!}{|F|!} (v_{S \cup \{j\}}(x_{S \cup \{j\}}) - v_S(x_S))$$

where v_S represents the value function trained without feature j and $v_{S \cup \{j\}}$ represents the value function trained including feature j . Lundberg et al [129] proposed SHAP values as a unified feature importance measure, which are the Shapley values calculated from a conditional expectation function derived from the original model. They demonstrate that SHAP values offer a distinct additive measure of feature importance that conforms to local accuracy, missingness, and consistency principles. However, the computation of the Shapley values for the SHAP method is expensive for a high dimensional feature space. Hence, they can be approximated using, for instance, Shapely sampling values [129].

Local interpretable model-agnostic explanations

Local interpretable model-agnostic explanations (LIME) is another XAI Method that has been applied for interpreting ML methods and explaining a classifier’s predictions by learning an interpretable model locally around the prediction [142]. It can be applied to different data types, such as text, images, and medical data. For instance, SurvLIME was developed for explaining ML survival models [143]. In the following, the LIME algorithm is presented based on the work of Ribeiro et al [142]. The general idea of LIME is to approximate a complex ML model locally around a given prediction by a less complex model such as a linear model, simple decision trees, or learned rules. To make the interpretations of the LIME algorithm understandable, Ribeiro et al [142] distinguished between the original representation of an instance $x \in \mathbb{R}^d$ and the interpretable representation of the instance in a binary vector $x' \in \{0, 1\}^d$, which would be a superpixel for image classification for example. In newborn screening, the features are the metabolite concentrations that are interpretable for clinical experts. Hence, there is no need for a transformation to an interpretable representation of the instance into a binary vector. The explanation model for a model $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined as $g \in G$ the class of interpretable models where the domain of g is $\{0, 1\}^d$. Furthermore, a proximity measure between an instance z and an instance x is defined as $\pi_x(z)$. The loss function $\mathcal{L}(f, g, \pi_x)$ is defined assessing the quality of g approximating f within the locality defined by $\pi_x(z)$. Complex models with multiple features are more difficult to interpret. Hence, the complexity is penalized with a measure of model complexity defined as $\Omega(g)$. Then, the goal is to find the function g that minimizes the loss function while also minimizing this complexity measure, which results in a general LASSO-approach [142], with loss function \mathcal{L} and regularization function Ω ,

$$\xi(x) = \operatorname{argmin}_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g). \quad (4.6)$$

A sampling approach for local exploration is used to learn the local behavior of f . Therefore, $\mathcal{L}(f, g, \pi_x)$ is approximated by drawing samples which are weighted by the proximity measure $\pi_x(z)$. The perturbed samples $z' \in \{0, 1\}^d$ contain a fraction of nonzero elements of x and are transformed to instances in the original representation $z \in \mathbb{R}^d$. The original classification model f is then used to obtain a label for this perturbed sample $f(z)$ based on the data set \mathcal{Z} of perturbed samples. Their corresponding labels, Eq. (4.6), can be optimized to obtain an explanation model g [142]. This explanation model g can be used to gain insights into the model's decision-making process for one instance x from the data set.

4.2 XAI in newborn screening

This section describes the results of the ensemble and deep learning methods applied to the IVA data set and the subsequently applied XAI method to interpret and find patterns within the classification algorithms.

4.2.1 Data set and experimental setup

For the ML experiments, the IVA full data set containing 2,106,090 newborn screening profiles and the suspected diagnosis data set containing 131 newborn screening profiles established in section 3.2 were applied. The IVA profiles are not further distinguished in patients with mild and classic IVA. For both data sets, stratified splitting was applied to obtain randomly split training, validation, and test data sets with a 65-15-20 split. The data sets were highly imbalanced and contained the same proportional number of IVA samples.

The correct classification of IVA patients is crucial and the best hyperparameters had to be established to reduce misclassification from IVA patients. A grid search optimization strategy on the hyperparameters of the XGBoost and NN models was applied to find the optimal number of estimators ([1, 50]), tree depth ([1, 25]), and scale positive weight ([1, 20]) from the respective predefined parameter ranges. This strategy revealed that the optimal number of estimators was 16 for both data sets, and the maximal tree node depth was three. For the full data set, a scaled weight $w_s = 6$ was optimal, whereas for the suspected diagnosis subset $w_s = 1$ showed to be optimal. For the NN, the grid search for hyperparameter optimization resulted in two different architectures for the two data sets, Table 4.1. In particular, the class weight parameters differ between the two architectures as the data imbalance is higher in the full data set than in the suspected diagnosis data set. The minority class weight parameter w_1 can be used to penalize the misclassification of a true positive in the cross entropy loss function \mathcal{L}_C . For the model evaluation, the same procedures

Table 4.1: Overview of hyperparameters for NN architectures in the suspected diagnosis and the full data set with four layers each. Copyright ©2023, IEEE [32].

	SUSPECTED DIAGNOSIS DATA	FULL DATA
Class weight w_0	0.64	0.520833
Class weight w_1	2.34	37608.75
Neurons	[100, 53, 6, 2]	[256, 128, 64, 2]
Optimizer	Stochastic Gradient Descent [136]	Root Mean Squared Propagation [137]

as for the ML methods in section 3.2 were chosen to allow for comparability. The algorithms were evaluated on the number of false positives and false negatives as well as the mean sensitivity S_n and specificity S_p from ten times repeated 5-fold cross-validation.

4.2.2 Ensemble and deep learning for IVA classification

The XGBoost and NN classification results on the suspected diagnosis data set and the full IVA data set are presented in Table 4.2. On both data sets, XGBoost showed a substantial decrease

Table 4.2: Classification results on training, validation, and test set with the number of false positives (FP) and false negatives (FN) as well as the mean sensitivity S_n and specificity S_p from ten times repeated 5-fold cross-validation (CV). Copyright ©2023, IEEE [32].

Method	Train		Validation		CV		Test	
	<i>FN</i>	<i>FP</i>	<i>FN</i>	<i>FP</i>	$S_n(\%)$	$S_p(\%)$	<i>FN</i>	<i>FP</i>
FULL DATA SET								
<i>Traditional</i>	0	67	0	18	100	99.995	0	18
XGBoost	0	0	1	1	93.727	99.999	2	1
NN	2	7	1	0	65.357	99.999	2	4
SUSPECTED DIAGNOSIS SUBSET								
<i>Traditional</i>	0	72	0	12	100	0	0	19
XGBoost	0	0	1	3	94.491	97.382	1	0
NN	0	0	1	3	43.357	71.237	1	2

in false positives on the training, validation, and test sets. For the full data set, the false positives decreased from 67 to 0 in the training set, as well as from 18 to 1 in the validation, and in the test set, Table 4.2. This was an overall decrease of false positives of 98.06% compared to traditional newborn screening. On the suspected diagnosis data set, the number of false positives was decreased by 97.09% compared to traditional newborn screening, from 72 to 0 in the training set, 12 to 3 in the validation set, and 19 to 0 in the test set, Table 4.2. The NN classification also showed a decrease in false positives on the full data set, decreasing the number of false positives from 67 to 7 in the training set, 18 to 0 in the validation, and 18 to 4 in the test data set. This represents an overall decrease of false positives of 89.32% compared to traditional newborn screening. On the suspected diagnosis data set, the number of false positives was decreased by 95.15% compared to traditional newborn screening, from 72 to 0 in the training set, 12 to 3 in the validation set, and 19 to 2 in the test set, Table 4.2. The cross-validation for XGBoost showed a high specificity of 99.99% on the full data set and 97.38% on the suspected diagnosis data set, Table 4.2. However, the NN showed less reliable results in cross-validation, decreasing the mean sensitivity to 43.36% and specificity to 71.24%, Table 4.2. Note, both algorithms classified at least

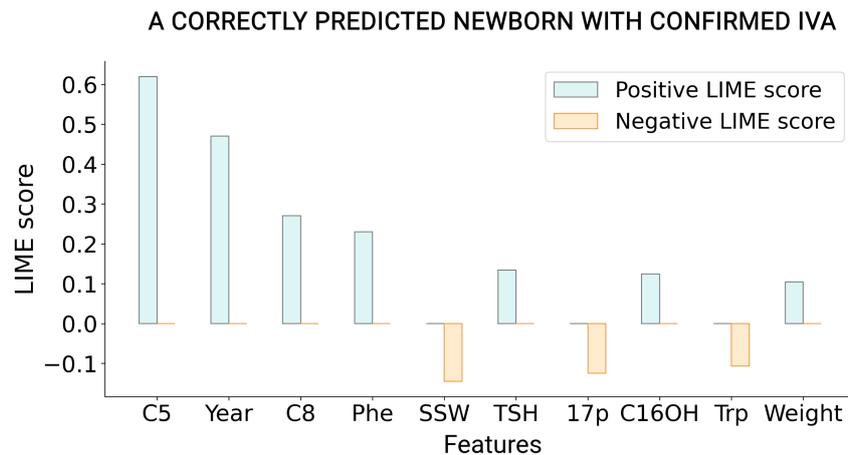


Figure 4.1: Positive and negative LIME scores l_{ix_1} for ten features i with the highest absolute LIME scores for a single newborn x_1 with IVA which was correctly classified by XGBoost trained on the full data set. Metabolite abbreviations can be found in Table 2.1. Copyright ©2023, IEEE [32].

one newborn with IVA wrongly as unremarkable in the validation and test data set, whereas in traditional newborn screening, all patients with IVA are identified correctly. This is also shown in the decreased sensitivity of XGBoost on both data sets, with 93.73% sensitivity on the full data set and 94.49% on the suspected diagnosis data set, which is 100% for traditional newborn screening. The NN showed less reliable results also for the mean sensitivity, decreasing it to 65.36% on the full data set, Table 4.2.

4.2.3 LIME method for pattern recognition

The model-agnostic method LIME is widely applied to interpret ML predictions. For one instance, x , which is the newborn screening profile in this application, the LIME method provides LIME scores for each feature depending on the best local linear model g . Figure 4.1 shows the positive and negative LIME scores l_{ix_1} for ten features i with the highest absolute LIME scores for a single newborn x_1 with IVA, which was correctly classified by XGBoost trained on the full data set. The bar chart showed that the feature values of isovalerylcarnitine (C5) (0.61), birth year (0.45), octanoylcarnitine (C8) (0.22), and phenylalanine (Phe) (0.21) had the highest positive LIME score, meaning that these features made a positive IVA prediction more likely. In particular, C5, the known biomarker for IVA, and the birth year had a high positive impact. Furthermore, the features gestational age (SSW) (-0.13), 17-hydroxyprogesterone (17p) (-0.1), and tryptophan (Trp) (-0.09) had negative LIME scores. Based on these results, the model could be interpreted as the values of C5 and the birth year being the main reasons for the model’s correct prediction of

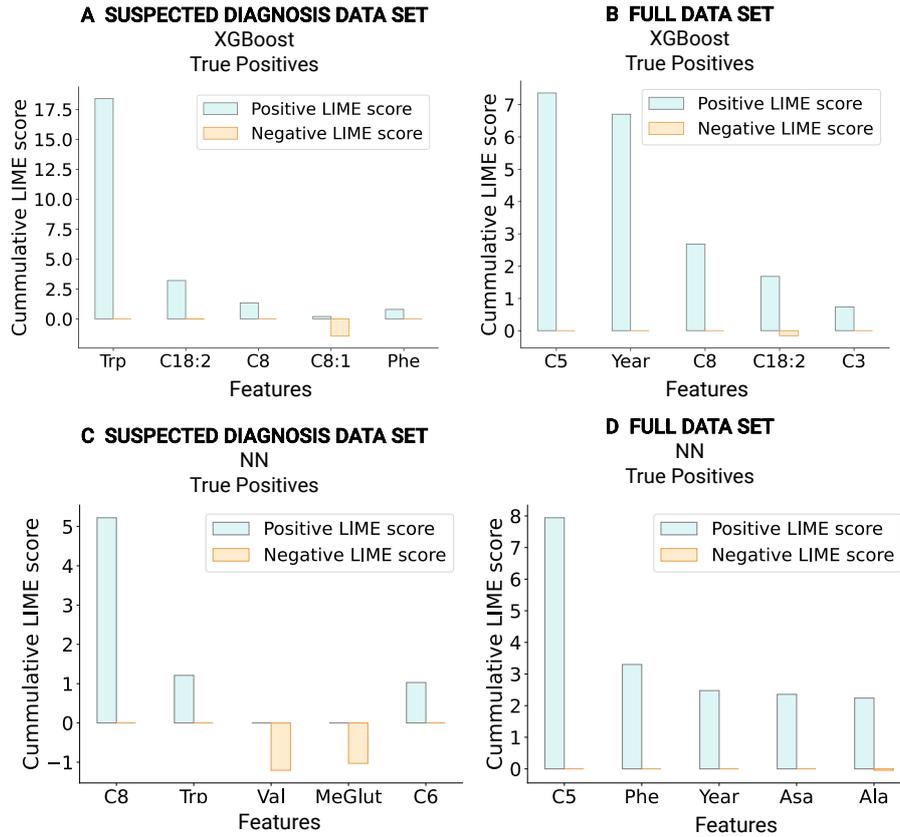


Figure 4.2: Positive and negative cumulative LIME scores L_j^+ and L_j^- for five features j with highest absolute LIME scores for the groups of true positives, the correctly classified IVA patients by (A), (B) XGBoost and (C), (D) neural network (NN) trained on (A), (C) the suspected diagnosis and (B), (D) the full data set. Metabolite abbreviations can be found in Table 2.1. Copyright ©2023, IEEE [32].

IVA. LIME presents an algorithm with a *local* scope. Another implementation of LIME, SP-LIME, aims at a *global* interpretation by selecting a representative set of explanations [142]. However, in classification tasks, the classification interpretation of a specific group of instances, such as a specific group of patients, is often more insightful than that of single instances or the whole group of instances. Hence, this research presents a novel configuration of LIME to interpret the classification model for different specific groups of instances. For every instance x from a group of n instances, where $n \leq d$ is a subgroup of the whole data set with d instances, we obtain a LIME score l_{jx} for every feature j . These were used to build two cumulative LIME scores, which add the positive L_j^+ and negative contributions L_j^- separately for every feature j ,

$$L_j^+ = \sum_x l_{xj}, \quad i = \{i = 1, \dots, n | l_{ij} > 0\},$$

$$L_j^- = \sum_x l_{xj}, \quad i = \{i = 1, \dots, n | l_{ij} < 0\}.$$

These two cumulative LIME scores could then be used to evaluate and compare the interpretation of the model classification for specific groups. Figure 4.2 shows the cumulative LIME scores for the group of newborns with IVA, which the XGBoost and NN correctly classified. For the group of true-positive IVA patients classified by the XGBoost method and trained on the suspected diagnosis data set, Trp had a LIME score $L_{Trp}^+ = 17.7$ and is by far the feature with the highest absolute cumulative LIME score (see A in Figure 4.2). For the same group classified with the NN, C8 had the highest cumulative LIME score $L_{C8}^+ = 5.1$ and Trp the second highest $L_{Trp}^+ = 1.2$ (see C in Figure 4.2). This indicates that the feature values of C8 and Trp made a positive disease prediction more likely in both classification methods. For the group of true-positive IVA patients classified by the XGBoost method and trained on the full data set, the top three features (C5, birth year, C8) with the highest absolute LIME score were the same as for the single newborn (see B in Figure 4.2). Based on the NN classification, C5 also had the highest cumulative LIME score $L_{C5}^+ = 7.9$ on the full data set, with Phe and birth year being the features with the next highest LIME score ($L_{Phe}^+ = 3.1$, $L_{Year}^+ = 2.3$). This emphasizes that the patterns in the single newborn can also be found in the whole group of newborns with correctly predicted confirmed IVA. Interestingly, different features had the highest cumulative LIME scores in the suspected diagnosis data set and the full data set. Moreover, all features either showed a high positive cumulative LIME score L_j^+ or a high negative cumulative LIME score L_j^- , indicating that the features influenced the model output in the same direction within one group.

4.2.4 SHAP method for interpretability

Further, the SHAP method was utilized to interpret the prediction mechanism of the XGBoost classifier. This method can be applied to an individual newborn to understand the impact different features or metabolite concentrations measured in the newborn's sample have on the classifier's prediction. To also understand the explanations for different groups of newborns, the full data set was divided into unremarkable newborns, newborns with suspected but not confirmed IVA, and newborns with confirmed IVA. This distinction aimed to gain insights into the differences and similarities of these groups. First, the SHAP method was applied to the three different groups from the full data set, where each group is presented in a swarm plot, Figure 4.3.

Figure 4.3 (A) shows that the feature with the strongest negative impact on the decision was the C5 value for the newborns with normal newborn screening profiles. These points were mainly colored in blue, indicating a low feature value, corresponding to the fact that an elevated level of

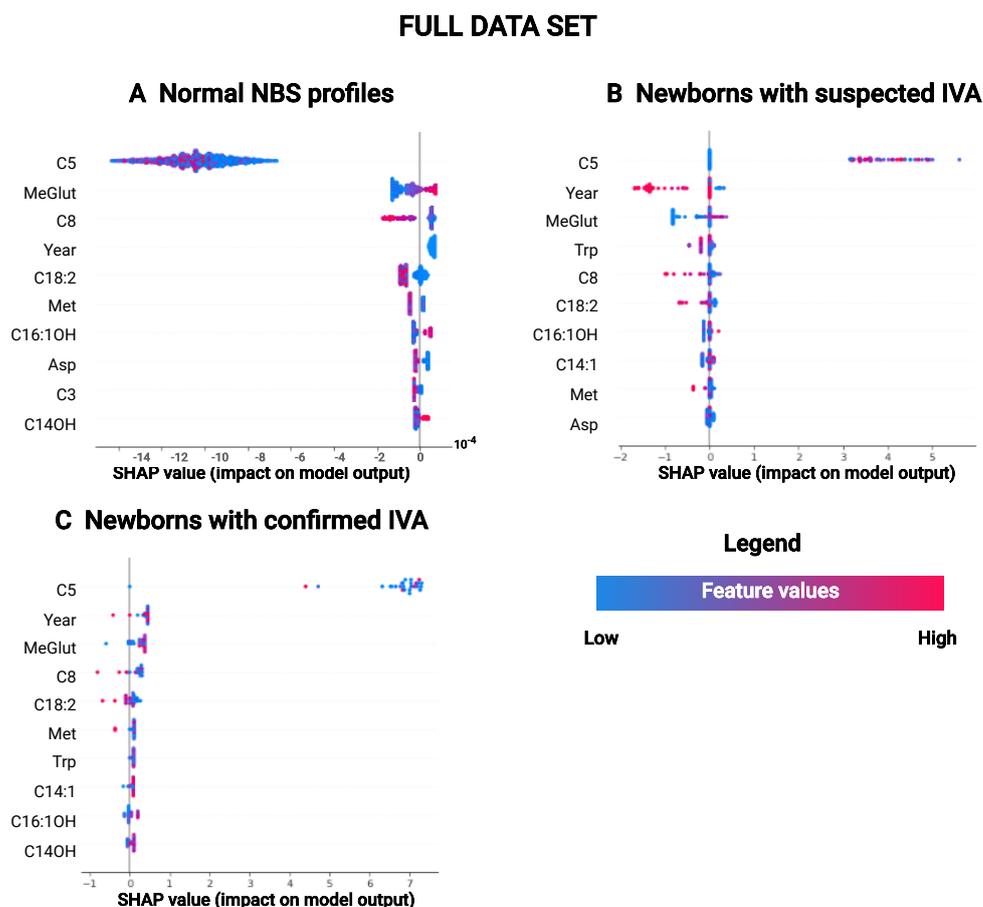


Figure 4.3: SHAP values for the full data set with the groups of (A) unremarkable newborns, (B) newborns with suspected IVA, and (C) newborns with confirmed IVA presenting the ten highest features. The data points are color-coded depending on the feature value, where blue data points show low feature values, and red data points show high feature values. Metabolite abbreviations can be found in Table 2.1. Copyright ©2023, IEEE [32].

C5 is the primary indicator for IVA [4], [25]. Also, for the group of newborns with suspected and confirmed IVA, Figure 4.3, the C5 values had the highest impact on the model output. Furthermore, the feature birth year (Year) had a high positive or negative impact on the model in the different groups. In particular, the group of newborns with suspected IVA revealed an interesting pattern, as a high birth year had a strong negative impact on the model output, Figure 4.3 (B).

For the suspected diagnosis data set, the 131 newborn screening profiles were divided into two groups, containing the newborns with suspected IVA and containing the newborns with confirmed IVA (including both mild and classic IVA), Figure 4.4. For both groups, the metabolite concentrations of Trp, and C5 had the strongest impact on the model's output. Interestingly, high Trp values had a negative impact on the model output for the group of newborns with suspected IVA.

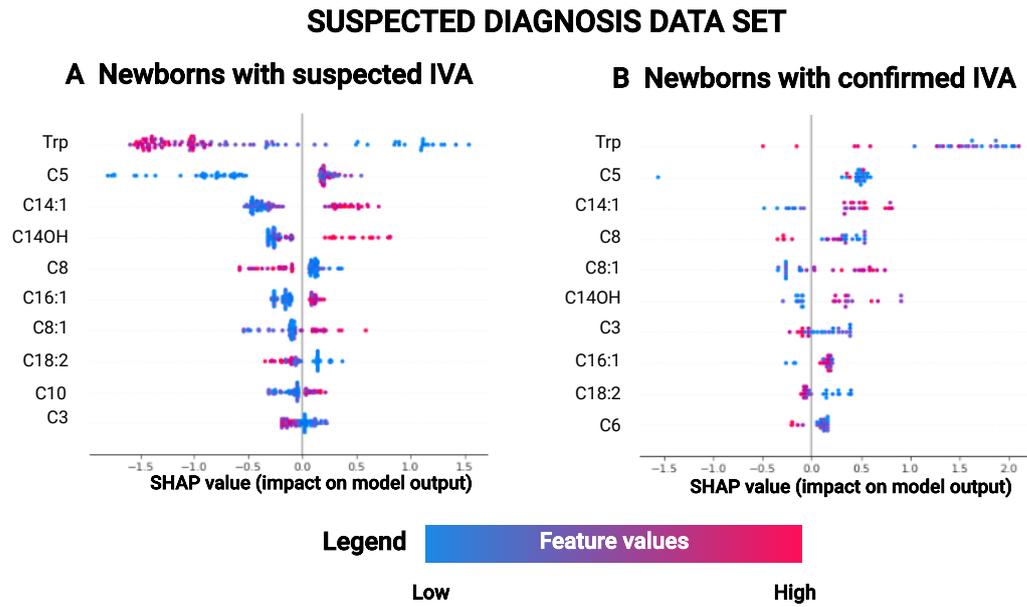


Figure 4.4: SHAP values for the suspected diagnosis data set with the groups of (A) newborns with suspected IVA, and (B) newborns with confirmed IVA presenting the ten highest features. The data points are color-coded depending on the feature value, where blue data points show low feature values, and red data points show high feature values. Metabolite abbreviations can be found in Table 2.1. Copyright ©2023, IEEE [32].

In contrast, low Trp values positively impacted the model output (see A in Figure 4.4). There was no clear distinction of this effect for newborns with confirmed IVA, although a similar trend could be seen in B in Figure 4.4. In contrast to the results on the full data set, the birth year was not a feature with high impact, which can be explained by the fact that this group already mainly included newborns with suspected IVA, who were generally born in recent years. The median birth year of newborns in the suspected diagnosis data set was 2019 since 80 of 103 newborns had a birth year of 2016 or later.

4.3 Summary and outcome

XAI methods enable to introduce interpretability into black-box AI models. The interpretability and traceability of an algorithm's actions are often key for adopting systems in the real world, particularly in system-critical areas such as the medical domain. Applying XAI methods provides a first step to tackle these problems and peek inside the AI black-box as they have been successfully applied to provide human-understandable interpretations of deep learning algorithms in the medical domain [15], [144]. However, the presented results showed the first application of XAI methods to interpret methods in machine learning-based newborn screening.

4.3.1 Ensemble and deep learning in newborn screening for IVA

In addition to applying basic machine learning methods in chapter 3, this chapter presented the application of ensemble and deep learning methods. On both data sets, XGBoost and NN classification showed promising results in reducing the false positives. However, the ensemble method XGBoost obtained good classification results more reliably than NN on both data sets. In particular, on the suspected diagnosis data set, the number of newborns falsely classified as newborns with IVA on the test set could be reduced by 100% with XGBoost, Table 4.2. When evaluating the cross-validation results, XGBoost proved to be a robust method obtaining $S_n = 93.727\%$ sensitivity and $S_p = 99.999\%$ specificity on the full data set. Whereas in traditional newborn screening, a sensitivity of 100% and specificity of 99.995% is obtained. For integrating machine learning-based newborn screening in practice, a sensitivity of 100% is essential to identify all newborns suffering from IVA. However, complex algorithms often require more data for accurate classification. In future studies, more newborn screening data could be included to train these algorithms and ensure higher positive sample sizes. In this work, the gradient boosting method XGBoost was shown to be the best performing algorithm; for other diseases also, other state-of-the-art methods in gradient boosting such as CatBoost[145] and LightGBM [146] could be evaluated.

4.3.2 XAI in newborn screening

In this work, the XAI methods SHAP [129] and LIME [142] were applied, which are commonly used in various medical applications [141], [143]. SHAP guarantees that the prediction is fairly distributed among the feature values of a patient and is based on a solid theory [129]. SHAP values

allow contrastive explanations, enabling comparisons of explanations to a subset or even to a single data point. On the limitations of SHAP, Frye et al [147] argue that SHAP values are ignorant of causal structures and have problems with correlated features within the data since they place all features equally in the model explanation by requiring attributions to be equally distributed over identically informative features. Two methods are presented to solve the problem of correlated features: Asymmetric SHAP and causal SHAP values [148], [149]. In newborn screening, all metabolic features are causally related by the metabolic network. However, how and if this causal relation influences the SHAP explanation is unclear and could potentially be investigated in future studies. When comparing SHAP to LIME, Aas et al [150] state that Shapley values explain the difference between the prediction and the global mean prediction. In contrast, LIME explains the difference between the prediction and a local mean prediction. Furthermore, Lundberg et al [129] evaluated the consistency of the model's explanation with human intuition and showed that Shapley values more strongly agree with human explanation than LIME. In general, humans prefer selective explanations. The superpixels used in the LIME algorithm can produce these, whereas the Shapley value method always uses all the features. In newborn screening, all features are interpretable for clinicians, which is not the case for single pixels in image classification.

Overall, the applied algorithms SHAP and LIME enable a model-agnostic evaluation and, by this, a simple comparison of explanations between different classification algorithms. Both methods are suitable for newborn screening, and it needs to be clarified whether one method should be preferred over the other. In the case of XGBoost, the most important features can also be analyzed with the feature importance parameter of the trained model. However, this is a measure of importance inherent to the algorithm and cannot be compared to other classification methods such as the NN classification. Moreover, this measure of importance only gives a global interpretation but no information on different groups within the data set, such as the different patient groups in newborn screening for IVA.

To enhance the benefits of these methods, the results on a group level were evaluated by introducing a cumulative LIME score, Figure 4.2, and by distinguishing the swarm plots of the SHAP values of different groups, Figure 4.3. These evaluations confirmed biological knowledge, such as highlighting C5 in both XGBoost and NN as a important feature for the full data set. C5 is the known biomarker of newborn screening for IVA in Germany, [4], [25]. Furthermore, the birth year was identified as an influential feature in both algorithms, Figure 4.2, which coincides with the increasing number of false positives in recent years due to the use of pivmecillinam as an antibiotic in pregnant women since its authorization in Germany in 2016 [24]. Moreover, both

methods highlighted additional metabolite concentrations, which were important for the correct prediction and need further analysis in future research. The features highlighted by the cumulated LIME method varied based on whether the XGBoost or NN classification method was applied. This discrepancy raises doubts about the significance of these particular features and needs in-depth analysis by biochemical experts.

In summary, this chapter demonstrated that the utilization of ensemble and deep learning techniques led to increased specificity but not 100% sensitivity in IVA classification. Applying XAI methods provided promising insights into result interpretation, offering novel prospects for further research in newborn screening. In the future, XAI methods could lead to higher acceptance of machine learning methods in newborn screening as it is still being determined how black-box machine learning methods can be applied in the clinical setting and if patients and society will accept them.

5

Whole-body modeling for infant metabolism

This chapter is based on and extends the published work of Zaunseeder et al [151]. It presents the development and validation of mathematical whole-body models for newborn and infant metabolism (infant-WBMs). These are organ-resolved, sex-specific, anatomically accurate models of newborn and infant metabolism representing the first 180 days of an infant's life. First, the methodological background is described. Then, the model development, including the modeling choices for the organ weights and the energy requirements, is presented. Subsequently, the model is validated on its quality to predict accurate infant growth and metabolic fluxes. Moreover, the model is used in three applications regarding newborn screening, including model personalization, inherited metabolic disease (IMD) biomarker prediction, and *in silico* evaluation of dietary treatment strategies for IMD patients. Finally, the presented model is discussed regarding its limitations and future work.

5.1 Methodological background

The mathematical models developed in this chapter are metabolic whole-body models (WBMs) based on the basic principles of constraint-based modeling.

5.1.1 Whole-body models

WBMs are metabolic network reconstructions of humans. They present a progression of the first human cell-based models, the Recon series [17], [18], [80], [85]. They extend this cell-based view by developing an ensemble of organ-specific reconstructions and their interactions. They have been constructed by integrating information from Recon3D [18], organ-specific details, and omics data [20]. These models capture the metabolism of 20 sex-independent organs, six blood cell types, as well as two sex-specific organs (testis and prostate) for the male model ('Harvey') and four sex-specific organs (breast, cervix, uterus, and ovary) for the female model ('Harvetta') [20]. Overall, Harvey has 81,094 reactions and 56,452 metabolites, whereas Harvetta contains 83,521 reactions and 58,851 metabolites [20]. Each of the models was constrained with 15 physiological parameters, which led to a novel paradigm in constraint-based modeling named physiologically and stoichiometrically constrained modeling (PSCM) [20], described in the PSCM toolbox in MATLAB (www.opencobra.github.io/cobratoolbox).

5.1.2 Flux balance analysis

For the mathematical analysis and simulation of fluxes within a metabolic reconstruction, the reconstruction network is transformed into a stoichiometric matrix $S \in \mathbb{R}^{m \times n}$ where the rows correspond to the m metabolites and the columns to the n reactions. The matrix entries s_{ij} are assigned a stoichiometric coefficient if metabolite i takes part in reaction j and zero otherwise. The change of a metabolite concentration x_k over time t is then represented by the k th row and the flux through the reaction by the flux vector $v = (v_1, \dots, v_n)$. Resulting in the mass-balance equation for all metabolite concentrations $x \in \mathbb{R}^m$ as,

$$\frac{dx}{dt} = S \cdot v. \quad (5.1)$$

For the simulations, the constrained-based reconstruction and analysis (COBRA) [79] approach is applied. This approach assumes that the metabolic system is in a steady state $\frac{dx}{dt} = S \cdot v = 0$. The steady-state constraint implies that metabolite concentrations remain constant over time, as the input and output fluxes must balance [101]. This concept can be seen as applying Kirchhoff's law to the metabolites within the network [101]. The constraint is then used in flux balance analysis (FBA), which presents an approach to calculate the optimal flux through the metabolic

network [81]. It is applied to optimize an objective function, leading to the linear program (LP),

$$\begin{aligned}
 \max_v \quad & c^T v \\
 \text{s.t.} \quad & S \cdot v = 0 \\
 & lb \leq v \leq ub,
 \end{aligned} \tag{5.2}$$

where $lb \in \mathbb{R}^n$ and $ub \in \mathbb{R}^n$ are the lower and upper bounds on the fluxes $v \in \mathbb{R}^n$. These are based on specific constraints such as physicochemical constraints, nutrient uptake rates, and enzyme reaction rates. The vector of weights $c \in \mathbb{R}^n$ indicates both which reactions and to what extent each reaction v contributes to the objective [81]. The lb of irreversible reactions is set to 0 mmol/day/person and $ub > 0$ mmol/day/person [20]. Negative flux through the reactions is allowed for reversible reactions, which leads to $lb < 0$ mmol/day/person. The bounds lb and ub of unconstrained reactions are set to -1,000,000 mmol/day/person and 1,000,000 mmol/day/person, respectively. FBA operates under the assumption that the network is in a state of homeostasis, signifying a stable condition with constant metabolite concentrations [101]. In this state, nutrients are consistently transformed for biomass production [101]. Additionally, FBA assumes that in each state, one or multiple biologically reasonable objectives are optimized [101]. The feasible flux distributions for a system of linear equations and inequality constraints then lie in a convex polyhedral cone [101].

In this work, the flux optimization methods are employed using the commercial solver IBM ILOG CPLEX Optimizer by the software company IBM [152] to solve the optimization problem. This high-performance programming solver implements the simplex method in the programming language C, providing interfaces to other programming languages such as MATLAB [153]. CPLEX solves LP, mixed integer programming, and quadratic programming. The simplex method is a widely applied algorithm that can solve any linear program. It operates in a two-phase manner, initially identifying a feasible solution as a starting point to determine the problem's feasibility or infeasibility. The simplex method is self-initiated from this solution, utilizing the solution as a starting point. In the second phase, an optimal feasible solution is sought, which either exists or an infinite edge, along which the objective function is unbounded, exists [154]. The simplex algorithm can identify redundant constraints in the problem formulation, efficiently solve problems with one or more optimal solutions, and addresses the dual problem, offering insights into cost coefficients, constraint coefficients, and right-hand-side data [155].

5.1.3 Quadratic flux balance analysis

The stoichiometric matrix S consists of more columns than rows, $n > m$. Hence, the problem is underdetermined, leading to a polyhedral convex steady-state solution space containing all feasible steady-state solutions. A second problem based on flux-minimization [156] is applied to reduce this feasible solution space. The main idea here is that, when provided with the values of key target fluxes, the probable arrangement of steady-state fluxes in the metabolic network tends to minimize the weighted sum of all fluxes [22]. The biological idea behind flux-minimization is based on enzyme limitations in a cell. An enzyme catalyzes each metabolic reaction, and hence, the flux through each reaction is aimed to be minimized [157]. This minimization ensures the most efficient use of the available proteome [157]. The parsimonious FBA presented by Lewis et al [158] is such a method. It aims to minimize the sum of all gene-associated reaction fluxes while maintaining optimal growth. For this, the L1 norm $\|v\|_1 = |v_1| + |v_2| + \dots + |v_n|$ is minimized leading to a bi-level LP [158],

$$\begin{aligned} \min_v \quad & \|v\|_1 \\ \text{s.t.} \quad & S \cdot v = 0, \\ & lb \leq v \leq ub, \\ & z^* = c^T v, \end{aligned}$$

where z^* is the optimal solution. However, the solution to this problem is not unique since all fluxes are penalized the same. For the flux-minimization of the whole-body models, a second, quadratic program (QP) is applied. Here, the Euclidean norm $\|v\|_2$ of the flux vector is minimized,

$$\begin{aligned} \min_v \quad & c^T v + \eta \|v\|_2 \\ \text{s.t.} \quad & S \cdot v = 0, \\ & lb \leq v \leq ub, \end{aligned}$$

where the regularization parameter $\eta = 1.0 \cdot 10^{-6}$ corresponding to the standard configura-

tion used in the COBRA Toolbox [159] is applied. The Euclidean norm enforces a more equal distribution of the fluxes and disproportionately penalizes large fluxes.

5.1.4 Biomass objective function

For FBA, it is essential to determine an objective function $z = c^T v$ that is a biologically reasonable objective function for the underlying metabolic system. For both cell-based and whole-body models, the biomass reaction is often used as the objective function to compute an optimal network state [81], [160]. The biomass objective function can be formulated depending on the composition of the cell and energetic requirements to generate biomass content from metabolic precursors and describe the growth requirements of a cell [160]. For human adult WBMs, only some organs have regenerative capacities such as the liver [161], heart [162], and kidney [20], [163]. To account for these organ-specific differences, the whole-body biomass reaction is a coupled reaction of organ-specific biomass reactions, weighted by the organ weights [20]. Organs that can regenerate are assigned a different organ-specific biomass reaction than organs that cannot. To address this, a biomass maintenance reaction was introduced for every organ. This ensures the upkeep of cellular metabolic profiles, specifically the organs' ability to produce all biomass components except nuclear deoxynucleotides [20]. Every organ includes a biomass maintenance reaction that accounts for the macromolecular precursors, such as amino acids, necessary for maintaining the organ [20].

5.1.5 Inherited metabolic diseases analysis

IMD analysis is utilized in metabolic modeling to demonstrate the human metabolic model's capability to predict known biomarkers accurately. IMDs often disrupt the normal metabolite flux. Hence, analyzing the metabolic flux over time is more informative for IMDs than the static measurement of metabolites at a single time point [35]. For the IMD analysis in the whole-body models, the method established in Thiele et al [20] and encoded by the *performIEMAnalysis.m* function in the COBRA Toolbox v3.0 [159] can be applied. This method is a two-step process. In the first step, all reactions $k = (k_1, \dots, k_n)^T$ in all model organs associated with the defect gene that causes the IMD are identified. The contribution of these reactions k_i is then summed in an artificial reaction v_d . This reaction is added to the model, and a first LP is solved, maximizing v_d . In the second step, 75% of the optimal solution z^* is used as a lower bound on v_d in a second LP. For this LP, an artificial demand reaction in the blood compartment v_D is set as objective function for flux maximization. This demand reaction is added using the *addDemandReaction.m* (COBRA

Toolbox V3.0 [159]) function for each biomarker. The second LP can be written as,

$$\begin{aligned}
 \max_{v_D} \quad & v_D \\
 \text{s.t.} \quad & S \cdot v = 0 \\
 & lb \leq v \leq ub, \\
 & 0.75 \cdot z^* \leq v_d \leq ub_d.
 \end{aligned}$$

Here, S and v correspond to the stoichiometric matrix and flux vector. The reactions v_d and v_D are added to the original model. This LP describes the evaluation for the wild-type model, which simulates the reference (healthy) model to obtain the fluxes v_{WT} . For the knock-out (disease) infant-WBM, the lower and upper bound of v_d are set to 0, and the maximization through each biomarker metabolite demand reaction v_D is determined,

$$\begin{aligned}
 \max_{v_D} \quad & v_D \\
 \text{s.t.} \quad & S \cdot v = 0 \\
 & lb \leq v \leq ub, \\
 & 0 \leq v_d \leq 0,
 \end{aligned}$$

to obtain the flux v_D for the respective biomarkers. Finally, the obtained fluxes within the biomarker reaction of the wild-type model v_{WT} and the diseased model v_{IMD} are compared. The relative flux increase f_v is calculated,

$$f_v = \frac{v_{IMD} - v_{WT}}{v_{WT}}.$$

With this ratio the flux changes between various biomarkers in a wild-type and IMD model can be compared.

5.2 Development of whole-body models for infants (infant-WBM)

The development of whole-body models for newborns and infants is based on the adult sex-specific, organ-resolved WBMs [20], Figure 5.1. The models were assembled in an anatomically and physiologically consistent manner and can be parameterized with physiological, dietary, and metabolomic data.

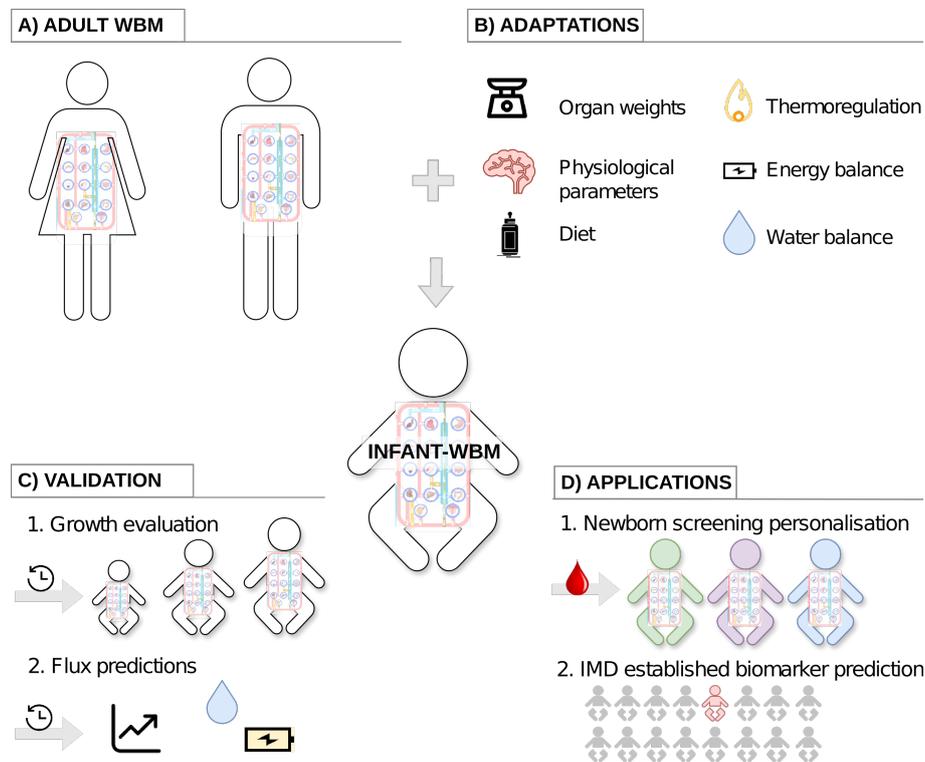


Figure 5.1: Overview of the reconstruction process and sample applications of the infant-WBMs. (A) The infant-WBMs were derived from the adult WBMs [20]. (B) Main adaptations made to the adult model to generate the infant WBMs. (C) Validation of the infant-WBMs with growth evaluation and flux predictions. (D) Application of the infant-WBMs for newborn screening personalization and biomarker prediction of IMDs. Figure taken from Zaunseder et al [34].

The development of the infant-WBMs consisted of six steps, resulting in the final model assembly. The final model is presented as a constrained-based optimization problem.

5.2.1 Organ-specific biomass growth reactions

All organ compartments in the infant-WBMs were presumed to undergo growth to mimic infant metabolism. While most organs in adult WBMs do not exhibit growth, certain organs, like the skin, already encompass biomass growth reactions in the model and, therefore, were not modified. The function *metBabies_WBM_adjustment.m* developed for the COBRA Toolbox 3.0, [159] was applied to change the biomass reactions in the male and female infant models. Therefore, the biomass growth reaction for all organs that do not undergo growth in adult models was restructured and incorporated into the models. These biomass growth reactions encompass all known components necessary for replicating the cells of an organ [160], [164]. Modeled upon the existing biomass maintenance reactions in adult WBMs, the biomass growth reactions were expanded to encompass both molecular and energetic prerequisites for replication. To accommodate organ-specific biomass growth reactions, additional reactions were introduced to the infant-WBMs, facilitating the transportation of biomass constituents into the appropriate compartments. These changes were integrated into the stoichiometric matrix representing the connections within the metabolic reaction network. By this, two new sex-specific and time-invariant stoichiometric matrices were created. The stoichiometric matrix $S^{male} \in \mathbb{R}^{57,980 \times 83,149}$ had 236 additional and the stoichiometric matrix $S^{female} \in \mathbb{R}^{60,436 \times 85,662}$ had 259 additional reactions compared to the respective adult models. These stoichiometric matrices are both very sparse matrices with not even 0.01% of non-zero entries.

5.2.2 Whole-body biomass reaction, organ weights, and blood flow

The infant-WBMs were scaled according to the infant's body weight and individual organ weights. Where available these values were obtained from literature [165]. For organs without corresponding literature values, the relative organ weight of the female and male adult WBM [20] was applied and scaled based on the infant's body weight. For the reference male infant-WBM, a body weight of 3.3 kg was assumed whereas for the reference female infant-WBM, a body weight of 3.2 kg was assumed. The whole-body biomass reaction with associated flux v_{WBB} is the weighted sum of the organ-specific biomass reactions depending on the organ weights. It represents the material and energy (adenosine triphosphate (ATP)) required to maintain the non-metabolic cellular functions of the body [20]. For the formulation of the infant-specific biomass reaction, the relative organ weights measured in literature [165] were applied. Additionally, in the female infant-WBMs, the weight of the uterus was specifically decreased, instead of adopting the relative value obtained from

the female adult model. This adjustment was made to enhance the accuracy of the representation. Moreover, the organ weights were applied to calculate the infant-specific organ blood flow rates b [165].

5.2.3 Physiological parameters

To enhance the accuracy of modeling infant metabolism, crucial physiological parameters—including heart rate, stroke volume, cardiac output, hematocrit value, creatinine in the urine, urine flow rate, glomerular filtration rate, and cerebrospinal fluid flow rate—were customized to be specific to infants. The modeling of these parameters involved the application of measured literature values and phenomenological models, based on the infant’s weight w and age a , Table 5.1.

Table 5.1: Applied physiological parameters for the infant-WBM based on measured literature values (*) or phenomenological models calculated with the infant’s weight w and age a . Comparison to parameters applied in adult WBMs.

Parameter	Adult WBM (unit) [20]	Infant-WBM (unit)
Heart rate	67 (beats/min)	140 (beats/min)* [166]
Stroke volume	80 (ml/beat)	$(1.77 \pm 0.28) \cdot w$ (ml/beat) [167]
Cardiac output	5,360 (ml/min)	Heart rate \cdot Stroke volume (ml/min) [168]
Creatinine in urine	0.5 - 1.2 (mg/dl)	27.56 - 61.05 (mg/dl)* [169]
Urine flow rate	2000 (ml/day)	$2 \cdot w \cdot 24$ (ml/day) [170]
Glomerular filtration rate	90 (ml/min)	40 (ml/min)* [171]

The cerebrospinal fluid flow rate measured in ml/day was approximated as [172],

$$CSFrate = \frac{2.78 + 0.97 \cdot \log\left(\frac{a}{365}\right) + 2.26 \cdot \log(w)}{60}.$$

In the initial phenomenological model, a subtraction of 2.23 was applied for female infants. However, for the sake of simplicity, the same phenomenological model was applied for both sex-specific infant-WBMs. The physiological parameters were then used to adapt the bounds on the related fluxes in the model. The blood flow rate and cardiac output were used to adapt the lower bound $lb_{m,organ}$ of metabolite fluxes from the blood compartment into individual organs which constrains the uptake of metabolite concentration from an organ. For this, first the organ-specific plasma

flow rate PFR was calculated as:

$$PFR_{organ} = b_{organ} \cdot CO \cdot (1 - h), \quad (5.3)$$

based on the organ-, sex-, and age-specific blood flow percentage b_{organ} , the cardiac output CO , and the hematocrit value h . The organ-specific plasma flow rate PFR_{organ} and metabolite concentrations M obtained from the Human Metabolome Database (HMDB) [20], [108] were then used to constrain the lower bound of metabolite fluxes M into an organ,

$$lb_{m,organ} = (-1) \cdot \frac{M}{1000} \cdot PFR_{organ} \cdot \frac{60 \cdot 24}{1000}, \quad (5.4)$$

given in mmol/day/person.

5.2.4 Water balance in infants

The water balance describes the equilibrium between water intake and water loss in the human body. Maintaining optimal hydration is essential for the proper functioning of various physiological processes in both adults and infants [173]. The intake and excretion of water differ significantly between infants and adults. The urinary excretion of water can be approximated as

$$2 \cdot w \text{ (in kg)} \cdot 24 \text{ (ml/day)}$$

per infant on their second day of life [170]. For the reference male infant-WBM, which weighs 3.3 kg, this is an excretion of 158 ml urine per day and for the reference female (3.2 kg) this leads to an excretion of 154 ml urine per day. These estimated values correspond to approximately 10% of the urine excretion of the male adult WBM, which has been estimated to be 1,400 ml per day [20]. Hence, the water excretion through urine feces, sweat, and air has been set to 10% of the adult WBMs on the second day of life. From these constraints set on the first days of life, the amount of excreted water was then daily increased depending on the daily water intake from the infant's diet. Since the water excretion can be variable, a margin of $\pm 10\%$ was set for each water excretion path on every day. The upper bounds ub_i and lower bounds lb_i of each water excretion reaction i were approximated based on the bound b_i on the respective reaction in the adult WBMs [20] and

set to

$$lb_i = b_i \cdot 10\% \cdot 90\% \cdot p$$

$$ub_i = b_i \cdot 10\% \cdot 110\% \cdot p,$$

where p is the milk intake percentage.

5.2.5 Energy balance in infants

In contrast to the adult WBMs, the infant-WBMs are modeled in an active, and not in a resting state, which resembles the real human metabolism more closely. By this, the aim of infant's metabolism to optimize nutrition for growth, activity, and normal organ development can be simulated under the constraint that all energy requirements necessary for normal infant development are met. Since the energy balance is very different in infants compared to adults, a large part of the model development included the integration of organ-specific energy demands [174]. These cover the energy intake by the diet and infant-specific energy expenditures to allow for a normal development of the infant, which should be able to move around, keep warm, and regulate heat. In particular, the energy demands for brain development, heart function, muscular activity, and thermoregulation over six months were included.

Dietary nutrition

In this model, the infants were assumed to be exclusively breastfed until the age of 6 months, as recommended by the WHO [175]. To model the dietary intake of breast milk correctly, the diet composition and amount of diet intake had to be defined. The diet composition of infants varies over time according to the needs of a growing child [176], as well as depending on environmental pollution [177], exogenous chemicals, such as drugs and synthetic compounds [178], and the mother [179]. In this study, the human milk diet composition $c \in \mathbb{R}^l$ of the Virtual Metabolic Human database [180] (<https://www.vmh.life/>) was applied to determine the number of components of c and their concentrations per 100 ml. Consultation with pediatric clinicians from Heidelberg University Hospital (UKHD) showed that the diet lacked lactose but was appropriate for infants otherwise. Lactose has been shown to account for approximately 7% of breast milk [181]. Hence, this component was added to the breast milk diet c . The formulated diet comprised approximately 87.5% of water, aligning with findings from studies that analyzed the composition of

human breast milk [182]. As the diet is a very complex and sensitive part of the model, it had to be modified to ensure the feasibility of the linear problem in FBA [81] over several time steps to allow adequate flux through the whole-body biomass reaction v_{WBB} . For this, the uptake bounds of dietary fluxes in the model of 12 metabolites, including six essential amino acids (L-methionine, L-isoleucine, L-valine, L-phenylalanine, L-threonine, L-leucine) and six other components (choline, phosphatidylethanolamine, homocitrulline, D-glucose, thiamin monophosphate, guanidinoacetic acid), were increased. Additionally, the model experienced growth-stunting effects due to a reduced phosphate concentration in the diet. Also, the L-lysine concentration was decreased every month starting from two months, which agrees with L-lysine concentration measurements, that showed a decrease in breast milk after two weeks of lactation [183]. Furthermore, for the female infant-WBMs, the dietary intake of L-cysteine was adapted as it presented another growth-limiting factor in the models.

The amount of daily milk intake was based on the milk model applied in the STIG-met model [29] and agreed with mean values of Swedish babies at 1, 2, 3, and 6 months of age [184], [185]. The only adaptation to this milk model was, that on the second day of life, it was increased from 100 g to 225 g per day since on their second day of life newborns consume 22-27 ml milk per feed at 8-10 feeds per day [186]. Hence, for every day the diet $d \in \mathbb{R}^l$ was obtained by multiplying the diet composition c with the corresponding daily milk intake m . Additionally, a band of $\pm 20\%$ was applied on the diet fluxes v_{diet} to account for variability of the diet,

$$d_{0.8} = 0.8 \cdot c \cdot m \leq v_{diet} \leq 1.2 \cdot c \cdot m = d_{1.2}. \quad (5.5)$$

For both male and female infant-WBMs, the same milk model was employed, as a study of a healthy infant cohort revealed that milk intake was comparable between male and female infants at both 1 month and 3 months of age [187].

Brain development

The development of the brain and its associated demand for glucose play a crucial role in the overall growth of infants and, consequently, in the construction of infant-WBMs. The extended duration and energy investment dedicated to learning and brain development have been suggested as factors contributing to the comparatively slow pre-adult life stage in humans [188], [189]. The brain glucose uptake per infant per day was calculated as the sum of the cerebral, cerebellar, and brainstem glucose uptake and has been predicted to be 19.7 grams (converted to 100 mmol

of glucose using the molecular weight of glucose 180,156 g/mol) [188]. For the infant-WBMs, it was assumed that the glycerine-3-phosphate shuttle is used to produce 30 mol ATP from 1 g glucose [190]. This conversion was applied to model the brain energy demand DM_{brain} , adapting the lower bound of the brain ATP demand reaction. In detail, the predicted accumulated brain glucose uptake was used and converted into mmol assuming the aforementioned conversion. On day 1 this resulted in a lower bound of the glucose uptake of $DM_{brain} = 100 \text{ mmol} \cdot 30 = 3,000$ mmol for the infant-WBMs. The lower limit was increased daily, initially with a rapid increase over two months, and then with a more slow ascent aligning with the typical body growth pattern of an infant. The constraint on the brain ATP demand reaction flux could be described for any day as,

$$DM_{brain} \leq v_{brain,atp}. \quad (5.6)$$

During the first six months of the infant's life, the ATP demand of the brain increased daily as shown in Figure 5.2 (blue line).

Thermoregulation

The regulation of body temperature in humans involves the capacity to manage both heat production and loss, ensuring that the body temperature is maintained within a normal range. For infants, thermoregulation of the body temperature is very important due to several factors. They have less insulating fat compared to adults making it more difficult for them to keep the heat [191], [192]. Furthermore, their body proportions make thermoregulation challenging, as they have a relatively high surface-area-to-mass ratio, and a high head-to-body size ratio [191], [192]. Moreover, newborns and infants that are younger than six months are not able to shiver to produce heat since their muscles are relatively immature [193]. Infants employ a process known as non-shivering thermogenesis for thermoregulation, which does not require muscle contraction [194], [195]. When using non-shivering thermogenesis, heat is mainly produced by the brown adipose tissue and also to a smaller extent in skeletal muscle, brain, liver, and white adipose tissue [196]. Brown adipose tissue is the main component of the infant's fat [197] and it enables thermoregulation by an uncoupling protein, i.e., thermogenin [194], [195]. As there are no precise measurements for the exact energy needs of thermoregulation in infants, the ATP demand DM_{heat} was estimated using literature references and the energy demand of the brain. This was based on the estimation that around 44% [198] to 55% [191] of the total heat production of an infant originates from brain

metabolism. For brain metabolism, it has been reported that approximately 1/3 of the glucose consumed by the brain is released as heat, while the remaining portion is utilized to generate ATP [199], [200]. In the previous paragraph, the non-metabolic energy demand of the brain was estimated as 3,000 mmol ATP per infant per day at age one day. Hence, a minimum of a third of this value was assumed to correspond to the reported 1/3 of consumed glucose leading to a lower bound on the adipose tissue energy demand reaction of $DM_{heat} = 3,000 \cdot 1/3 = 1,000$ mmol ATP per infant per day at age one day. Similar to the brain ATP demand it increased proportionally every day, Figure 5.2 (orange line), expressing the nearly equivalent contribution (44% [198] and 55% [191]) of both the brain and brown fat to the infant's heat production. More generally, the adipose tissue ATP demand reaction flux $v_{adipose,atp}$ could be described for any day as,

$$DM_{heat} \leq v_{adipose,atp}. \quad (5.7)$$

Heart

For infants and adults, the heart is a life-sustaining organ and the main organ of the cardiovascular system, responsible for pumping blood throughout the body. Besides the heart rate and the stroke volume, Table 5.1, also the size of the heart is different in infants and adults [165]. At birth, the infant's heart weight was approximated as 20 g [165], whereas the adult's heart weighs approximately 330 g [20]. The heart ATP demand for adult WBMs has been estimated to be 6,000 mmol ATP per day [20]. These two heart weight approximations were used to estimate the heart ATP demand DM_{heart} for the infant-WBMs. The infant's heart weight (20 g) corresponds to 6% of the adult's heart weight. Therefore, the lower bound on the heart ATP demand was set to $DM_{heart} = 6,000 \cdot 6\% = 360$ mmol ATP on day one and then increased daily according to the relative heart weight of the infant, Figure 5.2 (red line). More generally, the heart ATP demand reaction flux $v_{heart,atp}$ could be described for any day as,

$$DM_{heart} \leq v_{heart,atp}. \quad (5.8)$$

Muscular activity

In comparison to other cell-based [18] and whole-body reconstruction [20] of humans, the infant-WBMs are developed to model the infant's metabolism in an active and not in a resting state.

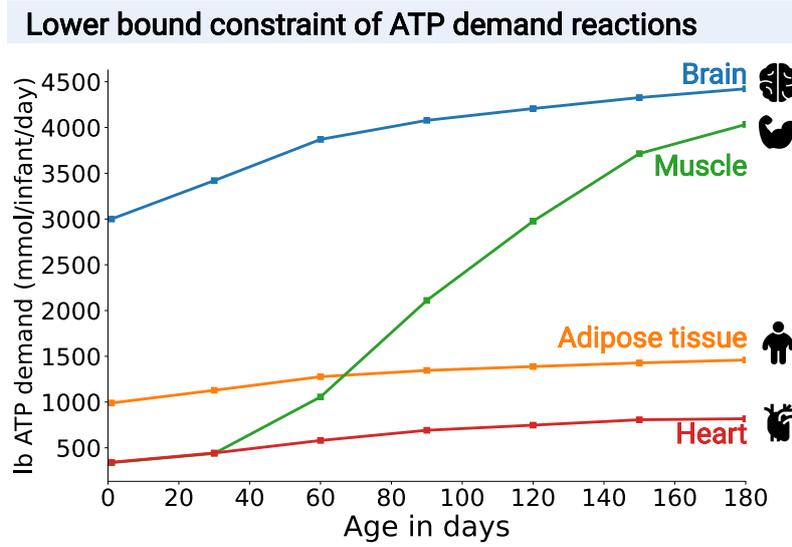


Figure 5.2: Lower bounds (lb) on ATP demand reactions in brain, heart, muscle and adipose tissue in mmol/day/person during the first six months of a male infant's life. Figure taken from Zaunseder et al [34].

Therefore, another vital mechanism considered in the model development was the energy expended by the infant in physical activity. For the integration of this energy demand, DM_{muscle} , a published activity model [29] was applied which accounted for the difference between sleeping energy expenditure and total energy expenditure. The energy expenditure varies with age and was determined to be 4.2 kcal/kg in newborns [201], 10 kcal/kg in 3-month-old infants [202], and 14.4 kcal/kg in 4 - 6 month-old infants [203]. To account for this variability in the activity model, the value of 14.4 kcal/kg was multiplied by a factor $\alpha \in [0, 1]$, which was determined based on estimated changes in physical activity sourced from the literature and interpolated by a second-degree polynomial function [29], [204]. The energy expenditure was subsequently converted to 28.12 mmol ATP hydrolyzed per kcal, a value estimated through the simulation of ATP production from glucose considering the known energy content of 4 kcal/g [29]. As a result, the obtained value was then multiplied by the infant's body weight, resulting in the calculated energy expenditure for physical activity ranging from 328 to 3,515 mmol/day/infant ATP in the female models and 338 to 4,033 mmol/day/infant ATP in the male models, depending on the age,

$$DM_{muscle} \leq v_{muscle,atp}. \quad (5.9)$$

In the infant-WBMs, this calculated activity expenditure in ATP was used to set the lower bound of the muscle ATP demand reaction $v_{muscle,atp}$, Figure 5.2 (green line).

5.2.6 Integration of newborn screening data

The newborn screening data used for integration into the infant-WBMs was obtained from the newborn screening laboratory at UKHD. The data was already applied in this work for machine learning methods in IVA diagnostics, section 3.2. To ensure data protection, the metabolite concentrations applied for the reference male and female infant-WBMs were synthetic data. Each metabolite concentration in the synthetic data was based on the sex-specific mean values μ from newborn screening, by uniform random sampling from the interval $[0.8\mu, 1.2\mu]$. From all 52 newborn screening metabolites, only 29 (17 amino acids and 12 acylcarnitines) were utilized as well as information on sex, and birth weight from each newborn. To integrate the metabolite concentration measurements, the blood uptake reactions for each organ corresponding to the metabolites m were identified to update their bounds respectively. To allow variability within the fluxes the metabolite concentrations $m_{conc,m} \in \mathbb{R}_{\geq 0}$ were multiplied with a coefficient of variation $c_m \in \mathbb{R}_{\geq 0}$ determined by the UKHD newborn screening laboratory, reflecting day-to-day variability within the tandem mass spectrometry. This range was further constrained with the coefficient $x_m = \min\{c_m, 0.1\}$. Hence, the bounds of the flux interval $m_{min,m}$ and $m_{max,m}$ were calculated for every metabolite concentrations as:

$$\begin{aligned} m_{min,m} &= m_{conc,m} \cdot (1 - x_m), \\ m_{max,m} &= m_{conc,m} \cdot (1 + x_m). \end{aligned}$$

The minimum and maximum concentrations of a metabolite were denoted as m_{min} and m_{max} , respectively, and expressed in $\mu\text{mol/l}$ of blood. These characterize the metabolite's concentration in the blood compartment. Given that an organ cannot absorb more of the metabolite from the blood than what is maximally present in the blood compartment, the lower bound $lb_{m,organ}$ of the flux into an organ was restricted by the maximum metabolite concentration m_{max} . For all organs except for the kidney, only the uptake of metabolites from the blood was constrained utilizing the organ-specific plasma flow rate PFR_{organ} , Eq. (5.3),

$$lb_{m,organ} = (-1) \cdot \frac{m_{max,m}}{1000} \cdot PFR_{organ} \cdot \frac{60 \cdot 24}{1000}, \quad (5.10)$$

given in mmol/day/person . For the kidney, the lower bound $lb_{m,kidney}$ and upper bound $ub_{m,kidney}$ on its metabolite uptake reaction fluxes were calculated using the glomerular filtration rate (GFR),

Table 5.1,

$$\begin{aligned} ub_{m,kidney} &= (-1) \cdot \frac{m_{min,m}}{1000} \cdot GFR \cdot \frac{60 \cdot 24}{1000} \\ lb_{m,kidney} &= (-1) \cdot \frac{m_{max,m}}{1000} \cdot GFR \cdot \frac{60 \cdot 24}{1000}, \end{aligned}$$

and are given in mmol/day/person. In cases where literature data for the blood flow percentage was unavailable, it was assigned the default value of 1%. These equations and flux relationships were formulated by Thiele et al [20].

5.2.7 Model assembly

After these six model development steps, the male and female infant-WBMs were assembled. The male infant-WBM accounted for 1,724 unique genes (2,071 transcripts), 83,149 reactions, and 57,980 metabolites stoichiometrically represented in S^{male} and the female infant-WBM accounted for 1,724 unique genes (2,071 transcripts), 85,662 reactions, and 60,436 metabolites stoichiometrically represented in S^{female} . The flux through the whole-body biomass functions v_{WBB}^{male} and v_{WBB}^{female} were applied as maximization objective functions for the sex-specific models respectively. The resulting LP for a male infant can then be formulated as,

$$\begin{aligned} \max_{\substack{v_{WBB}^{male} \\ v_{WBB}^{male}}} & v_{WBB}^{male} \\ s.t. & S^{male} \cdot v = 0 \\ & lb^{male} \leq v \leq ub^{male} \\ & d_{0.8} \leq v_{diet} \leq d_{1.2} \\ & DM_{brain} \leq v_{brain,atp} \\ & DM_{heat} \leq v_{heat,atp}^{male} \\ & DM_{heart}^{male} \leq v_{heart,atp} \\ & DM_{muscle}^{male} \leq v_{muscle,atp}^{male} \\ & (-1) \cdot \frac{M}{1000} \cdot PFR_{organ} \cdot \frac{60 \cdot 24}{1000} \leq v_{M,organ} \\ & (-1) \cdot \frac{m_{max,m}}{1000} \cdot PFR_{organ} \cdot \frac{60 \cdot 24}{1000} \leq v_{m,organ} \\ & \frac{-m_{max,m}}{1000} \cdot GFR \cdot \frac{60 \cdot 24}{1000} \leq v_{m,kidney} \leq \frac{-m_{min,m}}{1000} \cdot GFR \cdot \frac{60 \cdot 24}{1000} \end{aligned} \tag{5.11}$$

where lb_{diet} and ub_{diet} denote the lower and upper bounds on the intake flux of the l diet

metabolites. The lower bounds $v_{brain,atp}$, $v_{heat,atp}$, $v_{heart,atp}^{male}$, and $v_{muscle,atp}^{male}$, are the fluxes through the ATP demand reaction in the respective organs. Uptake reactions, such as dietary intake, were defined with negative flux values, while excretion reactions, including urine, fecal, air, and sweat excretion, were defined with positive flux values [20]. The heart and the muscles bounds are sex-specific, as they depend on the weight of the infant, which is lower in the female reference model. All energy demands, milk intake constraints ($d_{0.8}, d_{1.2}$), and the bounds on metabolite uptake rates are age-dependent, implying that on each day an infant's metabolism was modeled these constraints changed respectively.

5.3 Infant-WBM - Model validation

Mathematical models that simulate real processes, including biological systems, must be validated against data from the underlying real system to ensure their functionality. The underlying system for infant-WBMs is infant metabolism. Therefore, this section presents two validation steps linked to the infant's ability to grow while maintaining a functioning water balance and ATP synthase over the first six months of life.

5.3.1 Model growth prediction

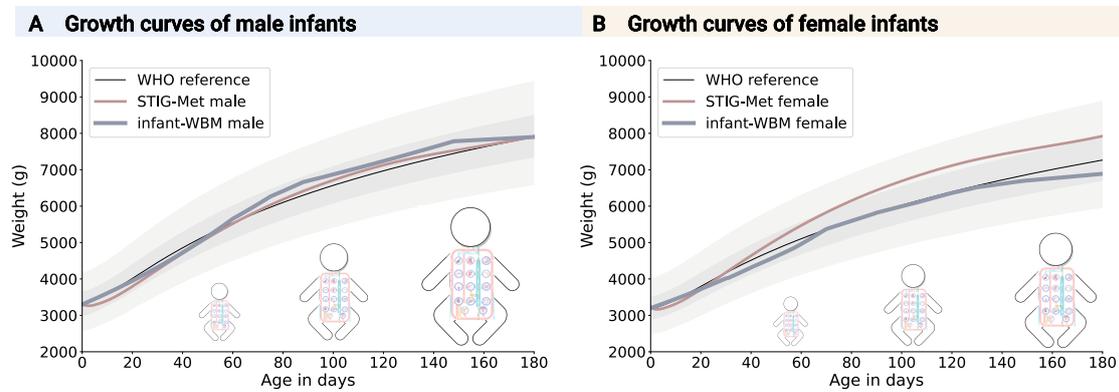


Figure 5.3: (A) Male and (B) female infant-WBM growth predictions in comparison with WHO quartiles [30] and predictions by the STIG-Met model [29]. The light gray area shows the 100% distribution, the dark gray shows the 25 - 75% distribution, and the black line presents the median of the WHO reference values. Figure taken from Zaunseder et al [34].

During the first six months, infants undergo rapid growth and aim to optimize their nutrient intake for body growth, activity, and organ development [205], [206]. This differs significantly from adults, where only a few organs can reproduce cells. Hence, the development step of the infant-WBMs included integrating organ-specific biomass reactions, which allow growth in all organs. For the validation of the infant-WBMs, the predicted growth of the models was compared to growth charts from the WHO [30] and a generic genome-scale model of infant metabolism, STIG-Met [29]. Therefore, the whole-body biomass reaction v_{WBB} , which accounts for the organ-specific biomass reactions relative to their organ weight, was maximized. This optimization simulated the maximal possible growth rate, considering limitations imposed by nutritional inputs and requirements for thermoregulation, brain development, heart function, and muscular activity. For the daily growth starting at age day 0, the LP, Eq. (5.11), maximizing the flux through the whole-

body biomass reaction, was solved. The resulting flux can be interpreted as the growth rate. A flux value of 1.01 corresponded to a 1% gain in biomass, i.e., an infant's body weight increased by 1%. According to the growth rate, the body and organ weights were adapted equally. On every following day, age-dependent parameters such as milk intake ($d_{0.8}, d_{1.2}$), energy demands ($DM_{brain}, DM_{heat}, DM_{heart}^{male}, DM_{muscle}^{male}$) and metabolite uptake rates ($lb_{m,organ}, ub_{m,organ}$) were updated. This update was performed until the infant model reached the age of six months. The growth evaluation started with a birth weight of 3300 g in the male and 3200 g in the female model, Figure 5.3. On a daily average, the male infant-WBMs gained 0.49%, reaching a weight of 7,900 g after six months, whereas the female models gained 0.43%, reaching a weight of 6,889 g after six months. Overall, both models showed continuous growth with more rapid growth in the first two months (mean male: 0.9%; female: 0.73%) and a slower increase in the subsequent months (male: 0.28%; female: 0.27%). The growth predictions of both models were within the 25 - 75% quartile of the WHO reference values [30] and were comparable to the predictions of the STIG-Met model [29]. For further validation, the weight prediction of each organ at age six months was compared to organ measurements of infants at this age [165], Figure 5.4.

A Organ weights of male infants				B Organ weights of female infants			
Organs	Weight Day 1 (g)	Ref. ^a Weight Day 180 (g)	Infant-WBM Weight Day 180 (g)	Organs	Weight Day 1 (g)	Ref. ^a Weight Day 180 (g)	Infant-WBM Weight Day 180 (g)
Blood	160	350	361	Blood	150	310	313
Brain	350	750	789	Brain	350	710	730
Heart	20	40	45	Heart	20	40	42
Lung	60	120	135	Lung	50	120	105
Liver	120	270	271	Liver	130	250	271
Kidney	30	50	68	Kidney	30	40	63
Skin	170	290	383	Skin	160	270	334
Stomach	10	20	23	Stomach	10	20	21
Gut	50	90	113	Gut	50	90	104
Muscle	800	1350	1805	Muscle	800	1350	1668
Adipose tissue	890	2970	2008	Adipose tissue	890	2850	1856

Figure 5.4: (A) Organ weights of male infant-WBM compared to reference values on day 180 (Ref.^a) [165]. (B) Organ weights of female infant-WBM compared to reference values on day 180 (Ref.^a) [165]. Figure taken from Zauneder et al [34].

For both models, the organ predictions at age six months agreed well with the measured literature values, Figure 5.4. The predicted weights of the blood compartment, brain, heart, lung, liver, and stomach were within $\pm 15\%$ of the measured organ weights for both male and female models. The models tended to overestimate kidney, skin, and gut growth. The most considerable differences were in both models' muscle and adipose tissue. The male infant-WBMs overestimated the muscle organ growth by 34% and underestimated the adipose tissue growth by 32%. In the female models, the muscle growth was overestimated by 24%, and the adipose tissue growth was underestimated by 35%.

5.3.2 Water excretion flux prediction

The water balance in the human body is maintained through dietary intake of water and water loss via excretions through urine, feces, skin, and respiration. It is crucial to properly balance these factors for adults and infants, as the appropriate functioning of several physiological processes depends on it [173]. For the validation of the infant-WBMs, the prediction of water loss was evaluated by analyzing the flux distribution through the water excretion reactions at several time points, Figure 5.5.

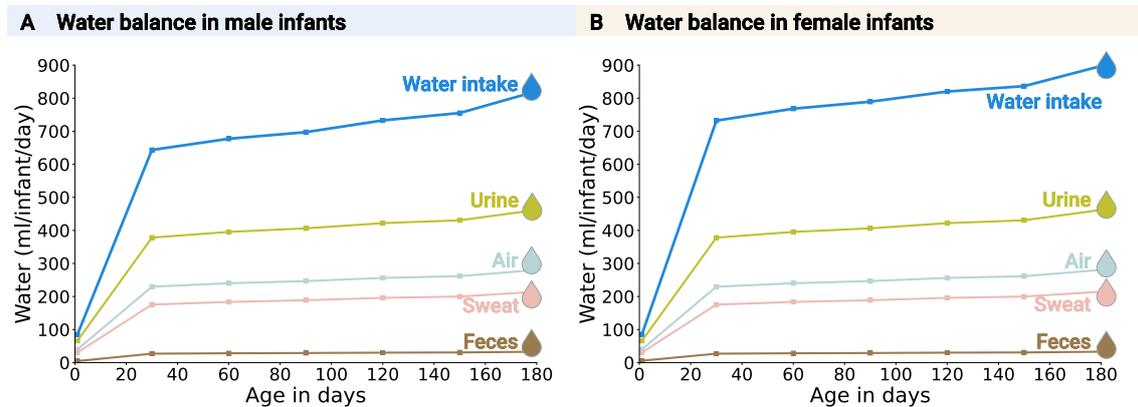


Figure 5.5: Predictions of water intake and excretion fluxes through urine, air, sweat, and feces in (A) male and (B) female infant-WBMs over first six months of an infant's life. Figure taken from Zaunseeder et al [34].

Overall, the water balance predictions were similar for male and female infant-WBMs, Figure 5.5. The predicted water excretion for both models ranged from 141 - 990 ml/day over the first six months. The average water excretion per day was 784 ml, and the predicted water excretion of both models on day 30 was 810 ml/day. This compared well with the reported mean water excretion of 900 ml/day from 78 male babies aged between 8 and 180 days (mean=36 days) [207]. From all excretion pathways, urinary excretion had the highest volume and was predicted between 66 - 462 ml per day for both models. These flux predictions are comparable with literature values, where the urine excretion of an infant was estimated as $2 \cdot \text{weight (kg)} \cdot 24 \text{ (ml/day)}$ [170]. For the male model weighing between 3.3 - 7.9 kg, this led to an approximated urine excretion of 158 - 379 ml/day, and for the female model weighing 3.2 - 6.9 kg, this led to an approximated urine excretion of 154 - 331 ml/day. The water loss through air and sweat was predicted as 71 - 495 ml in male and female models and was consistent with estimations on water evaporation through skin and air being 149 - 356 ml per day for 0 - 6 month olds [208]. The predicted fecal water loss ranged between 5 and 33 ml/day for male models and 6 - 33 ml/day for female models. These predictions

were comparable to literature values where fecal water loss for newborns was estimated as 5 ml/kg per day, which converts to 17 - 40 ml/day for male and 16 - 35 ml/day for female models over the first six months [208]. Moreover, the infant-WBMs also accounted for the metabolic production of water, which was between 56 - 169 ml/day in male models and 57 - 93 ml/day in female models.

5.3.3 ATP synthase flux prediction

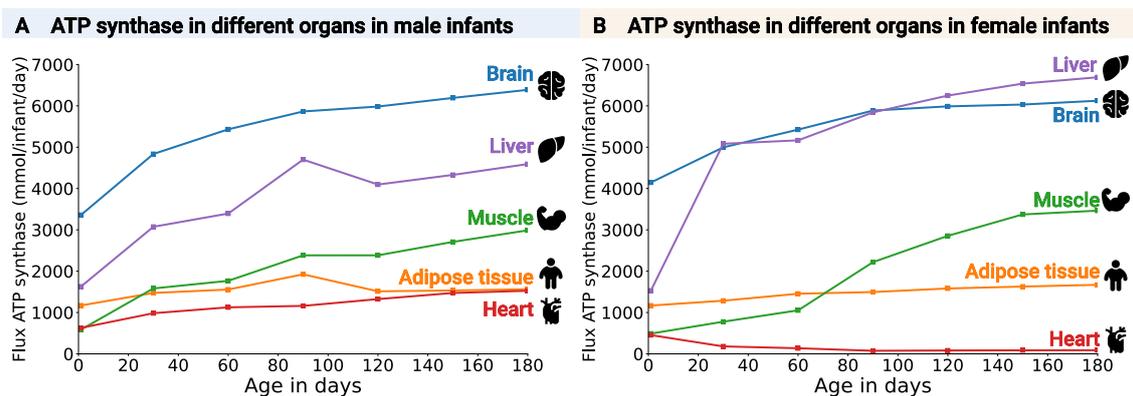


Figure 5.6: Predicted flux through ATP synthase reaction of the infant-WBM in brain, liver, muscle adipose tissue, and heart for (A) male and (B) female infants. Icons from <https://icons8.de>. Figure taken from Zaunseder et al [34].

The energy balance in the human body describes the energy intake and consumption. For infants, it is crucial to maintain sufficient energy for essential functions such as brain development, thermoregulation, heart function, and physical activity based on the energy intake from the breast milk diet. This analysis can only be performed in a whole-body model, such as the infant-WBMs, as they allow the allocation of distinct energy demands to specific organs for these functionalities. Each infant-WBM contained organ-specific ATP demand reactions, representing the ATP hydrolysis for non-growth-associated metabolic processes such as physical or brain activity. As the constraints on these reactions increased, an increase in the flux through the ATP synthase reaction was also expected. The ATP synthase reaction is involved in oxidative phosphorylation in these organs. For male and female models, the predicted flux through the brain and liver ATP synthase reaction was the highest over all time points, Figure 5.6. For male models, the brain ATP synthase flux was predicted to be 3,080 - 6,388 mmol/day/infant, and the liver ATP synthase flux was predicted to be 1,328- 4,591 mmol/day/infant. For female models, the brain ATP synthase flux was predicted as 4,146 - 6.123 mmol/day/infant, and the liver ATP synthase flux was predicted as 1,528 - 6,692 mmol/day/infant. No single factor could be identified that caused the sex-specific

differences in the hepatic ATP synthase flux. Although no constraints were imposed on the liver ATP demand reaction, both models predicted that the liver ATP synthase flux would increase. An analysis of the temporal changes of the ATP synthase flux showed that the muscle was predicted to experience the highest increase, which was five times higher on day 180 than on day 1. In contrast, the adipose tissue ATP synthase flux increased only by 40%, Figure 5.6.

5.4 Infant-WBM - Model application

Following the validation, three applications of the infant-WBMs are presented to show their potential use in future newborn screening research.

5.4.1 Personalization with newborn screening data

To demonstrate the infant-WBM's generalization capabilities, 10,000 models were personalized using data measurements from newborn screening. The data was already applied in this work for machine learning in newborn screening, section 3.2. For this evaluation, only unremarkable male newborn screening profiles were used. The cleaned data set consisted of 798,221 male newborns. From this, a uniformly random sampled subset of 10,000 newborn screening profiles of male newborns was drawn. Only 29 (17 amino acids and 12 acylcarnitines) of the 52 newborn screening metabolites were utilized, along with information on sex and birth weight for each newborn. A Wilcoxon rank-sum test was conducted to verify whether the birth weights of the subset (10,000 newborns) and the cleaned dataset of male newborns (798,221 newborns) originated from the same distribution, yielding a p -value of 0.27 (calculated with the Python library `scipy` [122]). The metabolite concentration data were then used to update the bounds on the uptake fluxes from the blood compartment into individual organs and all weight- and sex-specific parameters of the infant-WBMs were adjusted.

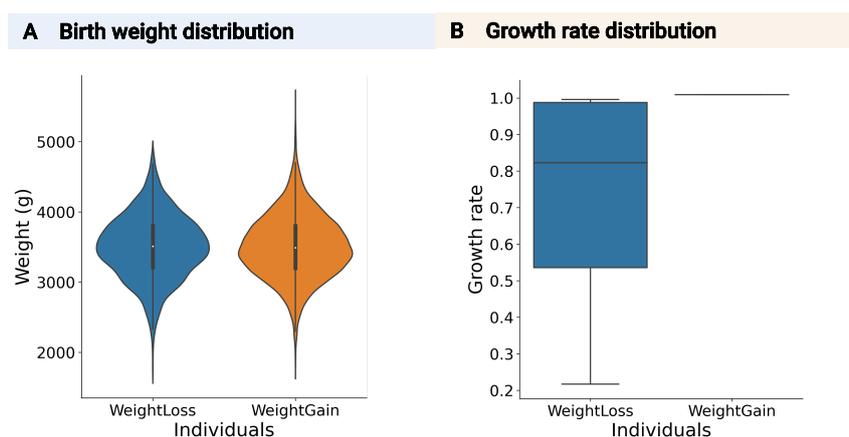


Figure 5.7: Evaluation results of 10,000 male infant-WBMs. (A) Birth weight and (B) growth rate distribution of infants with predicted weight loss (WeightLoss) and infants with predicted weight gain (WeightGain). Figure taken from Zaunseeder et al [34].

The mean measured birth weight in the subset of the 10,000 newborns was 3,508 g and ranged

between 1,770 - 5,590 g, Figure 5.7. The subset of the 10,000 individuals adequately represented the general male newborn population since the birth weight quartiles of the subset were within 170 g (3 - 5%) of the WHO reference values for boys at age 0. The 25 - 75% quartile of the subset's birth weight was 3200 - 3800 g, agreeing well with the WHO references 3027 - 3687 g, and the median was 3490 g (WHO reference 3346 g) [29], [30]. To estimate whether all these personalized models predicted reasonable *in silico* growth rates, FBA [81] was applied.

From all models in the subset, 8,736 (87.36%) models had a predicted growth rate between 1.0089 - 1.0092 (mean = 1.0091 ± 0.00008), which corresponded to a weight gain of 0.89 - 0.92% per day, Figure 5.7. Whereas 1,108 models (11.08%) had a predicted growth rate between 0.22 and 0.996 (mean = 0.74 ± 0.23), indicating a weight loss, Figure 5.7. A decrease in birth weight of up to 10% during the initial days of life is considered normal [209] and presumably attributed to fluid loss through urination [210]. However, no water-related constraints were changed in the personalized infant-WBMs. Since the data used for personalization came from unremarkable newborn screening profiles, no further information on the infants' follow-up weight, general health status, potential feeding problems, delivery mode, or the mother's age was available. This information could have helped validate the predicted weight loss since all these factors have been associated with excessive weight loss in newborns [211]. Overall, only 156 of the 10,000 personalized infant-WBMs (1.56%) were infeasible, meaning no flux solution that satisfied Eq. (5.11) could be found.

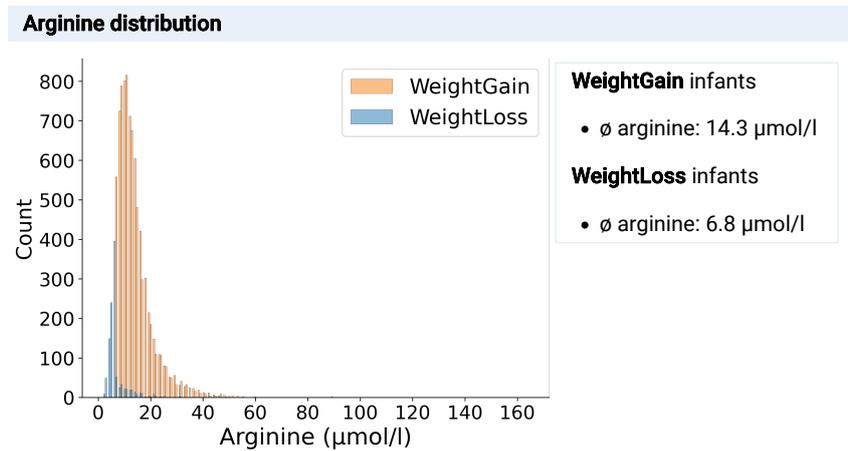


Figure 5.8: Comparison of measured arginine concentration ($\mu\text{mol/l}$) of predicted WeightGain infant models (orange) and predicted WeightLoss infant models (blue). Figure taken from Zaunseder et al [34].

The models were further investigated for potential metabolic reasons for the predicted weight loss. This comparison showed that the measured blood concentration of arginine was significantly lower in infants with predicted weight loss (mean = $6.8 \pm 4.1 \mu\text{mol/l}$) compared to infants with

predicted weight gain (mean = 14.3 ± 7.4 $\mu\text{mol/l}$), Figure 5.8. Arginine showed the highest absolute relative median difference (0.5) between the group of weight gainers and weight losers and had the lowest p -value (Wilcoxon rank sum test, $p < 0.0001$) of all compared metabolites. To test whether the arginine concentration was indeed the growth-stunting factor, the blood arginine concentration was increased *in silico* to 14 $\mu\text{mol/l}$. This increase enabled 762 (69%) of the models with predicted weight loss to grow with a growth rate of at least 1.005. The majority of these models, 99.74% (760), originally had a blood arginine concentration lower than 14 $\mu\text{mol/l}$. For the remaining 346 (31%) of the weight-loser infant-WBMs, no single measured blood metabolite could be identified to explain the predicted weight loss. Also, no dietary supplementation could achieve higher *in silico* growth rates in the predicted weight loss models, using a dedicated nutrition algorithm [212]. These results confirm literature knowledge on arginine, which is known to be an essential amino acid in newborns and infants [213], while not being an essential dietary amino acid in healthy adults [214]. The relationship between low arginine levels and growth stunting has been suggested in two studies. For children and adolescent patients with methylmalonic acidemia and propionic acidemia, low plasma l-arginine levels have been suggested as a potential cause of growth retardation [215]. Additionally, an investigation involving Danish schoolchildren revealed a connection between the intake of dietary arginine and the rate of growth [216].

5.4.2 Biomarker prediction for inherited metabolic diseases

IMD analysis is utilized in metabolic modeling to demonstrate the human metabolic model's capability to predict known biomarkers accurately. Starting with the first prediction of known IMD biomarkers using Recon1 [17], [21] this evaluation has been performed with cell-based models [18], [85], whole-body models [20], and models of organ-specific metabolism [22], [23]. Hence, the infant-WBM models' ability to predict known biomarkers was evaluated. Phenylketonuria (PKU) is an inborn error of phenylalanine (Phe) metabolism explicitly in the synthesis of tyrosine (Tyr) from Phe by the phenylalanine hydroxylase. If untreated, PKU is associated with global developmental delay and severe intellectual impairment of patients [217]. For the IMD prediction, the method *performIEMAnalysis.m* established in Thiele et al [20] was applied for male and female infant-WBMs. In addition to the known PKU biomarkers, 27 further metabolites routinely measured in the dried blood spots for newborn screening at the UKHD were evaluated. For the comparison, the relative flux change f from the wild type ('WT') model and the knock-out ('PKU') model was calculated, see also section 5.1.5. In the PKU model, the reactions tetrahydrobiopterin: oxygen oxidoreductase were deleted in all organs having the known defective gene *phenylalanine hydroxylase*.

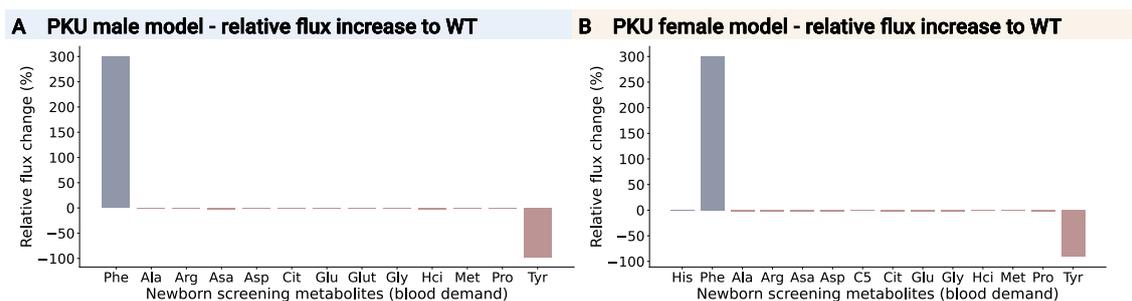


Figure 5.9: Relative change of flux through blood demand reactions of metabolites measured in newborn screening for (A) the male and (B) the female infant-WBM comparing the wild type (WT) with a phenylketonuria (PKU) knock-out model. Metabolite abbreviations can be found in Table 2.1. Figure taken from Zaunseder et al [34].

In male and female PKU infant-WBMs, the Phe flux was predicted to increase by more than 300%, and the Tyr flux was predicted to decrease by 97% in male and 91% in female PKU models. The flux through all other newborn screening metabolite reactions showed no or small relative flux changes below 3%, Figure 5.9. Hence, the known biomarkers showed the highest relative change in flux prediction. From the remaining metabolites with small relative flux changes, 12 out of 27 metabolite fluxes were predicted to change in both female and male models. One metabolite flux glutarylcarntine (Glut) was only predicted to change in the male PKU model, and two metabolite fluxes, histidine (His) and isovalerylcarnitine (C5), were only predicted to change in the female PKU model, Figure 5.9.

The infant-WBMs enable an evaluation of fluxes over different time points in the first six months of an infant's life. Hence, the biomarker prediction for IMDs was also analyzed over three different time points. This analysis was performed for the three newborn screening conditions PKU, IVA, and GA1, section 2.1.4. For each of these IMDs, the flux through their known biomarkers over the first six months was predicted for both male and female models. The predicted fluxes were analyzed qualitatively by comparing the flux through the known biomarker reactions in the wild type and IMD model, and it was noted whether an elevation or reduction of the flux was predicted, Table 5.2. For all IMDs and all biomarkers, the predictions were consistent over all three time points as well as with the change in biomarkers reported in IEMBase [218]. Further, the biomarker changes were consistent with changes predicted by the adult WBMs, Table 5.2. The adult WBMs have previously demonstrated strong predictive capabilities for IMDs [20].

Table 5.2: Qualitative flux change through biomarker reactions in blood on day 1, 90, and 180 for (A) male and (B) female infant-WBMs compared to adult whole-body models [20], when maximizing the respective biomarker reactions in the knockout (disease) and the wild type (healthy) model. A predicted flux increase is highlighted with an up-arrow \uparrow and a decrease with a down-arrow \downarrow . Metabolite abbreviations can be found in Table 2.1. PKU - phenylketonuria, IVA - isovaleric aciduria, GA1 - glutaric aciduria type 1.

Diseases	Metabolites	Day 1	Day 90	Day 180	Adult WBM
(A) MALE MODELS					
PKU	Phe	\uparrow	\uparrow	\uparrow	\uparrow
	Tyr	\downarrow	\downarrow	\downarrow	\downarrow
IVA	C5	\uparrow	\uparrow	\uparrow	\uparrow
GA1	Glut	\uparrow	\uparrow	\uparrow	\uparrow
(B) FEMALE MODELS					
PKU	Phe	\uparrow	\uparrow	\uparrow	\uparrow
	Tyr	\downarrow	\downarrow	\downarrow	\downarrow
IVA	C5	\uparrow	\uparrow	\uparrow	\uparrow
GA1	Glut	\uparrow	\uparrow	\uparrow	\uparrow

5.4.3 Dietary treatment effect prediction

The management of IMDs consists of treatment strategies such as the restriction of metabolite uptake or dietary supplementation [219]. Based on these dietary guidelines, necessary treatment and therapy can be planned. However, due to the substantial variability of IMDs, the disease management and therapy need to be adapted to the age-dependent requirements and the individual disease severity [35]. Identifying the best treatment strategy for a patient, including the amount of natural food and supplements, is part of the patient-specific diet [220]. Hence, the response of the infant-WBMs to dietary interventions for PKU was analyzed. In the IMD analysis, Table 5.2, it was shown that the infant-WBMs predicted an increase or decrease of a biomarker flux caused by the respective *in silico* gene knockout in an IMD model. For the dietary interventions, it was assessed whether these predicted changes could be attenuated through applying an *in silico* dietary treatment strategy. Therefore, the female and male PKU infant-WBMs were compared to the wild-type prediction, with and without dietary intervention.

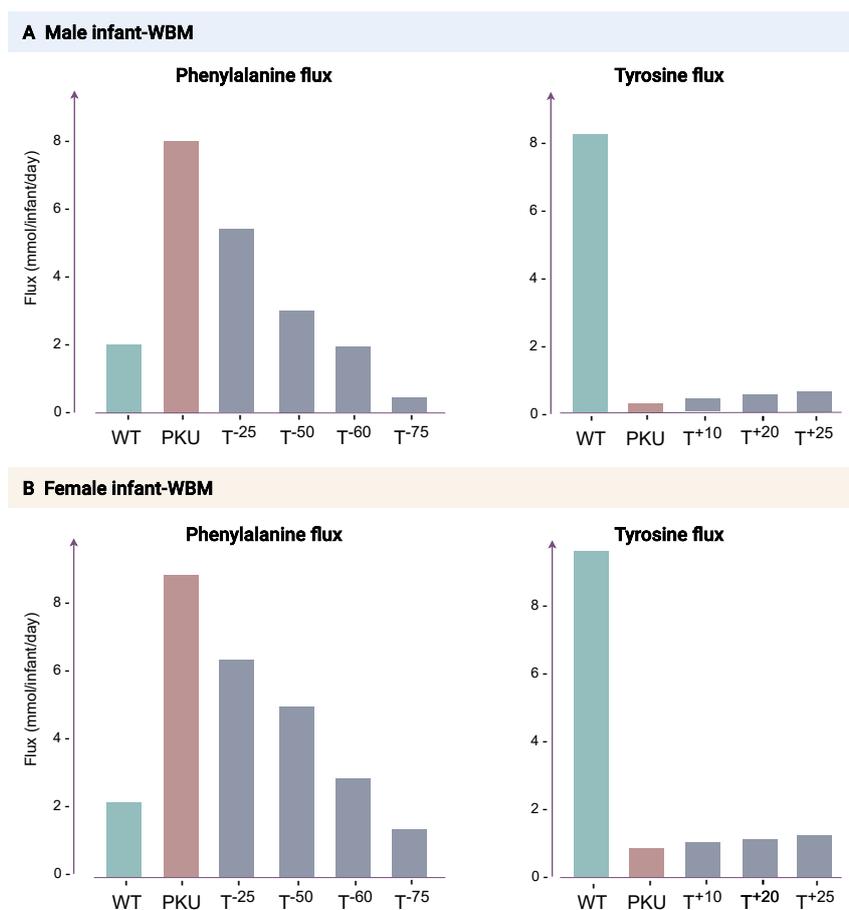


Figure 5.10: Flux prediction through blood demand reactions of biomarkers of infant-WBMs with dietary interventions. For (A) the male and (B) the female infant-WBMs, the predicted phenylalanine flux is compared in the wild type (WT) and the phenylketonuria (PKU) model, as well as in PKU models with four dietary treatment strategies (T^{-25} , T^{-50} , T^{-60} , T^{-75}), where 25%, 50%, 60% and 75% less dietary phenylalanine is given, respectively. The predicted tyrosine flux is compared in the WT and the PKU model, as well as in PKU models with three dietary treatment strategies (T^{+10} , T^{+20} , T^{+25}), where 10%, 20%, and 25% more dietary tyrosine is given, respectively. Figure taken from Zaunseeder et al [34].

For PKU patients, a highly restrictive low-Phe diet directly starting after birth is the most important dietary recommendation [219]. A low-Phe diet is achieved by restricting but not eliminating natural food to allow some Phe intake as it is an essential amino acid [219]. Additionally, an amino acid-based, Phe-free formula or amino acid-based medical food is added to provide sufficient nutrition [219]. The tolerance level of dietary Phe is very patient-specific as it depends, among others, on the type of PKU (mild or moderate PKU), the dosage, the growth, and the catabolic state during illness [220]. Furthermore, the supplementation of Tyr has been tested to increase the plasma Tyr levels found to be low in PKU patients [221].

Two dietary interventions with varying therapeutic strategies for PKU were tested for the

treatment analysis. In the first dietary intervention, the dietary Phe intake of the infant-WBM was varied to different therapeutic intensity levels, i.e., 25%, 50%, 60%, and 75% of the regular intake of the infant-WBMs. The impact of the dietary treatment on the infant-WBMs was compared in Figure 5.10 for male and female PKU models. For both models, the first dietary intervention had the clinically desired effect as a reduction of dietary Phe led to a reduced flux through the blood demand reaction of Phe in the PKU infant-WBM compared to the non-treated PKU infant-WBM, Figure 5.10. At a reduction of 60% of dietary Phe intake, the predicted Phe blood flux was comparable to the flux in the wild-type infant-WBM, while the further decrease in Phe intake lowered the predicted Phe blood flux below the wild type. Note that the results could only be compared qualitatively and not quantitatively since the IMD analysis maximizes the flux through the biomarker blood demand reaction to assess a model's capacity to produce this metabolite.

In the second dietary intervention, the *in silico* dietary intake of Tyr was increased, i.e., by 10%, 20%, and 25% of the regular intake of the infant-WBMs. This second dietary intervention also showed the clinically desired effect since it increased the predicted flux through the blood demand reaction of Tyr, Figure 5.10. Moreover, the varying therapeutic strategies showed small, gradual increases in the predicted flux through the blood demand reaction of Tyr. However, these predicted increases were still relatively small compared to the wild-type flux. Overall, the two therapeutic strategies led to similar results in the female and male infant-WBMs, Figure 5.10.

5.5 Summary and outcome

This chapter presented the development and validation of metabolic whole-body models for newborns and infants. These models simulate the metabolic systems of male and female infants by connecting metabolic pathways throughout several organs and blood compartments. They present an infant's metabolism not only for a specific age but throughout its first six months of life. During this time, the models were constrained to ensure that energy requirements and nutrition match values from reference infants. In contrast to adult WBMs, which are modeled in a resting state, the infant-WBMs were modeled in an active state. The energy requirements, especially due to the energy demand for brain development, heart function, muscular activity, and thermoregulation, were essential to simulate an active metabolism. Based on these requirements, the infant-WBMs could accurately, in accordance with WHO growth charts [30], predict the growth trajectory during the first six months. The water balance and ATP synthase evaluation demonstrated that the infants-WBMs agreed with *in vivo* measurements of infants throughout the six-month time frame and presented an emergent feature of the infant-WBM. Related to newborn screening, three applications were presented. The personalization capabilities of the models were evaluated with 10,000 personalized infant-WBMs utilizing the sex, birth weight, and 29 metabolite concentrations measured during newborn screening. Further, the infant-WBMs showed that they could correctly predict known metabolic biomarkers for PKU, IVA, and GA1 at different time points. Moreover, they proved their applicability for personalized treatment planning by predicting expected metabolic responses to simulated dietary interventions in PKU patients.

5.5.1 Modeling choices

The infant-WBMs were derived from female and male adult WBMs. The included reactions were exclusively expanded for biomass growth reactions in each organ, highlighting a crucial distinction between adult and infant metabolism. Except for these reactions, no further stoichiometric changes were placed on S^{female} and S^{male} compared to the adult models. However, the metabolic profile changes at various stages of human life [222]. For example, an examination of urine samples revealed significant variations in the activity of pathways related to amino acid metabolism between infants aged six months and one year [222]. These changes were not accounted for in the infant-WBMs and could lead to inaccuracies in the representation of the infant metabolism. Furthermore, the organ weights used to parameterize the infant-WBM were based on average measurements of real

newborns at birth [165]. After birth, each organ was assumed to grow equally according to the predicted body growth rate. A comparison with measurements of infants at age six months showed that the organ weights of the infant-WBMs agreed well for most organs but not for muscle and adipose tissue, Figure 5.4. Hence, future research could focus on applying different growth models for adipose tissue and muscle, which agree better with data measurements.

This work used the COBRA [79] approach for metabolic modeling. The placed steady-state assumption $\frac{dx}{dt} = Sv = 0$ is a strong restriction that erases the time component by assuming no changes in metabolite concentrations over time. The infant-WBMs consistently represent an accumulated flux over an entire day, making small-scale, time-dependent changes have a limited impact. Metabolite concentration changes over time could be more important when evaluating these models on a smaller time scale. These evaluations could include metabolite concentrations of a remarkably high biomarker concentration or simulations of dietary treatments for patients with IMDs.

5.5.2 Model limitations and future work

For the development of the infant-WBMs, no data measurements of an infant at different time points were available. Therefore, the essential information on infant metabolism was extracted from the literature and interpolated where necessary, resulting in model approximations and assumptions. These approximated parameters add uncertainty to the model and could influence the model's output. Moreover, the ability of the infant-WBMs to integrate personalized data and develop models that predict a reasonable growth rate was already shown using the newborn screening data. Hence, in future work, it would be very beneficial to obtain data from an infant at several time points, where all the required information, such as daily weight measurements, breast milk intake, and urine excretion, is measured. This data could be used to personalize the models, gain insights, and further validate them. For a Crohn's disease patient, fecal time series data covering several dates was already integrated to personalize metabolic whole-body models and simulate the patient-specific metabolism [223]. Using data from several time steps, also the growth prediction could be evaluated, which was difficult in the infants with predicted weight loss since no further information on the infants' follow-up weight (e.g., on day 3) and factors that are associated with weight loss in newborns [211] were available.

Furthermore, information on the exact milk composition and amount of milk intake could be used to personalize the models. The milk composition is very infant-specific, depending on the envi-

ronmental pollution [177], exogenous chemicals, such as drugs and synthetic compounds [178], the mother's nutrition [224] and milk production quantities [179]. Hence, the applied adjustments for the model feasibility of the milk diet could be justified with this variability of the diet composition. This personalization could provide valuable insights into the newborn's early-life metabolism and, thus, enable personalized interventions. A first analysis for *in silico* evaluation of dietary treatment strategies with infant-WBMs was already presented for the treatment of PKU. The infant-WBMs showed that they accurately predicted qualitative flux changes due to dietary treatment strategies with varying intensity levels. With this, the management of IMDs in infants could be improved as therapies could be tested *in silico*. Due to the extreme variability of IMDs, their management must be personalized for each patient, based on their diagnosis and phenotype [35]. In future research, infant-WBMs could be a first step toward a digital metabolic twin for infant metabolism, which could be used for personalized systematic simulations, treatment planning, and quantitative flux change prediction.

Moreover, time-dependent infant-WBMs could be very valuable in the area of drug research as the infant's immature drug metabolism is often associated with drug toxicity [225], and the pharmacokinetics and pharmacodynamics of drugs are often different in children and adults [36], [37]. For drug modeling, often PBPK models [226] are applied, accounting for the time-dependent metabolism changes. Hence, integrating these existing time-dependent modeling techniques with the infant-WBMs could allow for a detailed analysis of the infant metabolism. This could also include an analysis on a small time scale to estimate the effects of metabolic changes, nutrition, and drugs. Therefore, the development of a time-dependent infant-WBM will be presented and discussed in chapter 6.

6

Time-dependent modeling for infant-WBMs

The infant-WBMs established in the previous chapter allow a detailed evaluation of metabolic fluxes through an infant's body and individual organs. However, they are only evaluated daily and predict the flux through a reaction operating under steady-state conditions. By this, these models do not account for short-term effects, such as responses to perturbations or time-dependent variations in metabolites. Time-dependent modeling of metabolite concentrations enables the simulation and analysis of metabolite concentrations and their dynamics at different, also very small, time scales. One possibility to capture the time-dependent features of the infant metabolism is the integration of kinetic modeling such as in physiological-based pharmacokinetic (PBPK) modeling [105], [226]. This chapter will present the development of time-dependent evaluations for the metabolic infant whole-body models. First, the methodological background of PBPK modeling will be described. Then, the integration into the infant-WBM models will be presented and evaluated on IMD biomarker predictions, with a subsequent quantification of the model uncertainty. Finally, the integrated model and its potential application in newborn screening will be discussed.

6.1 Methodological background

In drug research, time-dependent mathematical models are developed to predict how drugs behave in the human body, particularly in pharmacokinetics [227]. These models are then used to gain

knowledge on human drug behavior for research in drug discovery and development [227].

6.1.1 Pharmacokinetic modeling

Pharmacokinetic (PK) models are used to understand drugs and their distribution in an organism following an intravenous or oral dosing [228]. A PK model usually consists of several linked compartments, where a central compartment depicting the blood plasma is connected to the other compartments [228]. PK models describe the change of drug concentration C over time t depending on the drug clearance Cl , which describes the volume of plasma or blood that is cleared of the drug per unit time (l/min) and the volume of distribution V which is the space a drug distributes into [229],

$$\frac{dC}{dt} = -\frac{Cl}{V} \cdot C. \quad (6.1)$$

Although it is possible to account for demographic and physiological parameters as covariates in these models, not all prior information on the physiology can be integrated despite their known effect on the drug-concentration relationship [228].

6.1.2 Physiological-based pharmacokinetic modeling (PBPK)

Physiological-based pharmacokinetic (PBPK) models are based on the same mathematical framework and extend PK models. They enable the integration of various physiological parameters by adding compartments that correspond to different organs and tissues in the human body [228]. PBPK models can extrapolate a drug dose from one population to another if the relevant physiological parameters are available [228]. The arterial and venous blood compartments connect the different organs in the model. In the following, the term tissue T will be used for all organs and compartments of the human body considered in the PBPK models. A schematic image of the compartments and the direction of their corresponding blood flow is presented in Figure 6.1.

In a PBPK model, tissues T are connected through the blood exchange in both venous and arterial blood. The arterial blood is the oxygenated blood, which is transported to all organs except the lung, where the organs take up oxygen, Figure 6.1. The venous blood is the deoxygenated blood, which leaves the organs and is pumped back into the lungs for oxygen uptake, Figure 6.1. For each tissue T , except for the lung, the rate of concentration change over time t of a compound C_T is described by the arterial inflow concentration C_a and the tissue-specific venous outflow

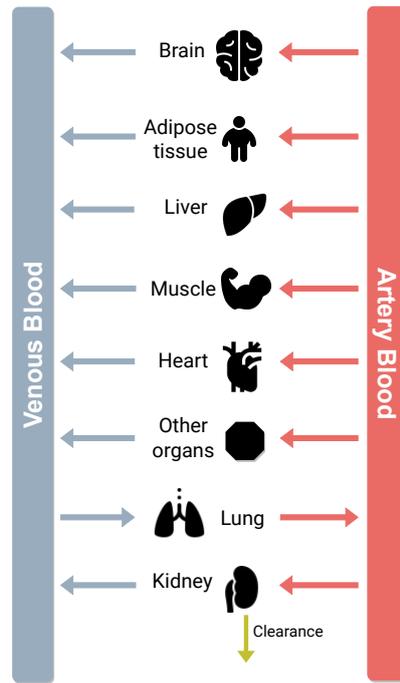


Figure 6.1: Overview of general PBPK model. Red arrows describe the blood exchange between artery blood and organs, blue arrows present the blood exchange between venous blood, and organs. The green arrow presents the renal and bile clearance.

concentration C_{v_T} ,

$$V_T \cdot \frac{dC_T}{dt} = Q_T \cdot C_a - Q_T \cdot C_{v_T}, \quad (6.2)$$

where V_T is the tissue volume (l), Q_T is the blood flow (l/h), and $C_{v_T} = \frac{C_T}{K_T/B}$ is the venous outflow concentration (mol/l) with K_T the tissue partition coefficient and B the blood-to-plasma ratio. The blood-to-plasma ratio is utilized to correct for blood when the plasma is used to determine the pharmacokinetic parameters, although the blood should be the true central compartment. In this study, the blood-to-plasma ratio is set to $B = 1$. Based on this, the generic ordinary differential equation (ODE) of the rate concentration C_T in a tissue T that is changing over time t can be written as

$$\frac{dC_T}{dt} = \frac{Q_T}{V_T} \cdot \left(C_a - \frac{C_T}{K_T} \right). \quad (6.3)$$

Different approaches, such as perfusion rate-limited kinetics and permeability rate-limited kinetics, are used to model the kinetics within the compartments. In this work, all processes are modeled with perfusion rate-limited kinetics since it often occurs for small lipophilic molecules, which can dissolve easier in lipids than in water [228]. For these molecules, the blood flow to the tissue becomes the limiting process [228]. In perfusion rate-limited kinetics, it is assumed that at steady state, the total drug concentration in a tissue T is in equilibrium with the total drug concentration

in the circulation as determined by the drug-specific tissue partition coefficient K_T [228]. The time to reach a steady state is determined by K_T , the blood flow rate Q_T , and the tissue volume V_T [228]. There are several methods to estimate the tissue partition coefficient K_T . In this work, the Poulin method will be applied [230], which assumes that the compound diffuses passively, distributing homogeneously into both tissue and plasma [228]. Based on this method, the tissue partition coefficient K_T of a tissue T can then be determined as,

$$K_T = \frac{10^l(V_T^{nl} + 0.3V_T^{ph}) + (V_T^w + 0.7V_T^{ph})f_p^u}{10^l(V_p^{nl} + 0.3V_p^{ph}) + (V_p^w + 0.7V_p^{ph})f_T^u}, \quad (6.4)$$

where l is the lipophilicity, referring to the capacity of a chemical compound to dissolve in lipids, V_T^{nl} and V_p^{nl} are the specific volume fractions of neutral lipid, V_T^{ph} and V_p^{ph} are the specific volume fractions of phospholipid, and V_T^w and V_p^w are the specific volume fractions of water in tissue T and plasma p , respectively. All these volumes can be obtained from literature [231]. Furthermore, the chemical bind in plasma or fractional unbound plasma f_p^u defines a drug's binding degree in plasma [232]. This term describes the fraction of the compound that is not bound to plasma proteins and free for interaction with receptors, metabolizing enzymes, and renal filtration. The compound-specific values for f_p^u can be obtained from experimental results in literature [233] and applied to calculate the chemical bind in tissue, f_T^u , with

$$f_T^u = \frac{1}{1 + \sqrt{\frac{1-f_p^u}{f_p^u}}}. \quad (6.5)$$

In addition to the tissue-specific models described by Eq. (6.3), the PBPK model incorporates additional equations for the arterial and venous blood compartment, as well as for the renal and bile excretion. The arterial and venous blood compartments connect all the tissue-specific ODEs. In this work, the PBPK model from Mohammad et al [105] is used for the integrated modeling, which describes the dynamics of a system with 23 ODEs. For each compound of interest i , the time-dependent concentration in a specific tissue such as adipose tissue, brain, small intestine, large intestine, heart, kidney, liver, lung, muscle, pancreas, skin, spleen, stomach, and bone is calculated. Additionally, there are auxiliary ODEs to describe the transit from bile to small and large intestine lumen as well as separate equations for renal and bile clearance. For a compound i (a metabolite or drug), the applied PBPK model can be described by the following system of equations:

Adipose tissue:

$$\frac{dC_{adipose,i}}{dt} = \frac{Q_{adipose}}{V_{adipose}} \left(C_{art,i} - \frac{C_{adipose,i}}{K_{adipose,i}} \right) \quad (6.6)$$

Artery blood

$$\frac{dC_{art,i}}{dt} = \frac{Q_{lung}C_{lung,i}}{V_{art} \cdot K_{lung,i}} - \frac{C_{art,i}Q_{art}}{V_{art}} \quad (6.7)$$

Brain

$$\frac{dC_{brain,i}}{dt} = \frac{Q_{brain}}{V_{brain}} \left(C_{art,i} - \frac{C_{brain,i}}{K_{brain,i}} \right) \quad (6.8)$$

Small intestine

$$\frac{dC_{SI,i}}{dt} = \frac{Q_{SI}}{V_{SI}} \left(C_{art,i} - \frac{C_{SI,i}}{K_{SI,i}} \right) \quad (6.9)$$

Large Intestine

$$\frac{dC_{LI,i}}{dt} = \frac{Q_{LI}}{V_{LI}} \left(C_{art,i} - \frac{C_{LI,i}}{K_{LI,i}} \right) + ka_i \cdot C_{LIL} \quad (6.10)$$

Heart

$$\frac{dC_{heart,i}}{dt} = \frac{Q_{heart}}{V_{heart}} \left(C_{art,i} - \frac{C_{heart,i}}{K_{heart,i}} \right) \quad (6.11)$$

Kidney

$$\frac{dC_{kidney,i}}{dt} = \frac{Q_{kidney}}{V_{kidney}} \left(C_{art,i} - \frac{C_{kidney,i}}{K_{kidney,i}} \right) - \frac{cl_{urine,i} \cdot C_{kidney,i}}{V_{kidney} \cdot K_{kidney,i}} \quad (6.12)$$

Liver

$$\begin{aligned} \frac{dC_{liver,i}}{dt} = & \frac{Q_{liver}}{V_{liver}} \left(C_{art,i} - \frac{C_{liver,i}}{K_{liver,i}} \right) + \frac{Q_{spleen}C_{spleen,i}}{K_{spleen,i}V_{liver}} + \frac{Q_{pancreas}C_{pancreas,i}}{K_{pancreas,i}V_{liver}} \\ & + \frac{Q_{stomach}C_{stomach,i}}{K_{stomach,i}V_{liver}} + \frac{Q_{SI}C_{SI,i}}{K_{SI,i}V_{liver}} + \frac{Q_{LI}C_{LI,i}}{K_{LI,i}V_{liver}} - \frac{cl_{bile,i}C_{liver,i}}{K_{liver,i}V_{liver}} \end{aligned} \quad (6.13)$$

Lung

$$\frac{dC_{lung,i}}{dt} = \frac{Q_{lung}}{V_{lung}} \left(C_{ven,i} - \frac{C_{lung,i}}{K_{lung,i}} \right) \quad (6.14)$$

Muscle

$$\frac{dC_{muscle,i}}{dt} = \frac{Q_{muscle}}{V_{muscle}} \left(C_{art,i} - \frac{C_{muscle,i}}{K_{muscle,i}} \right) \quad (6.15)$$

Pancreas

$$\frac{dC_{pancreas,i}}{dt} = \frac{Q_{pancreas}}{V_{pancreas}} \left(C_{art,i} - \frac{C_{pancreas,i}}{K_{pancreas,i}} \right) \quad (6.16)$$

Skin

$$\frac{dC_{skin,i}}{dt} = \frac{Q_{skin}}{V_{skin}} \left(C_{art,i} - \frac{C_{skin,i}}{K_{skin,i}} \right) \quad (6.17)$$

Spleen

$$\frac{dC_{spleen,i}}{dt} = \frac{Q_{spleen}}{V_{spleen}} \left(C_{art,i} - \frac{C_{spleen,i}}{K_{spleen,i}} \right) \quad (6.18)$$

Stomach

$$\frac{dC_{stomach,i}}{dt} = \frac{Q_{stomach}}{V_{stomach}} \left(C_{art,i} - \frac{C_{stomach,i}}{K_{stomach,i}} \right) \quad (6.19)$$

Bone

$$\frac{dC_{bone,i}}{dt} = \frac{Q_{bone}}{V_{bone}} \left(C_{art,i} - \frac{C_{bone,i}}{K_{bone,i}} \right) \quad (6.20)$$

Venous

$$\begin{aligned} \frac{dC_{ven}}{dt} = & \frac{Q_{liver}C_{liver,i}}{K_{liver,i}V_{ven}} + \frac{Q_{heart}C_{heart,i}}{K_{heart,i}V_{ven}} + \frac{Q_{brain}C_{brain,i}}{K_{brain,i}V_{ven}} + \frac{Q_{muscle}C_{muscle,i}}{K_{muscle,i}V_{ven}} \\ & + \frac{Q_{adipose}C_{adipose,i}}{K_{adipose,i}V_{ven}} + \frac{Q_{skin}C_{skin,i}}{K_{skin,i}V_{ven}} + \frac{Q_{bone}C_{bone,i}}{K_{bone,i}V_{ven}} \\ & + \frac{Q_{kidney}C_{kidney,i}}{K_{kidney,i}V_{ven}} - \frac{Q_{lung}C_{ven}}{V_{ven}} \end{aligned} \quad (6.21)$$

Transit from bile to small intestine lumen

$$\frac{dC_{transit1,i}}{dt} = \frac{cb_{bile,i} \cdot C_{liver,i}}{K_{liver,i}} - kb_{bile,i} \cdot C_{transit1,i} \quad (6.22)$$

$$\frac{dC_{transit2,i}}{dt} = kb_{bile,i} \cdot C_{transit1,i} - kb_{bile,i} \cdot C_{transit2,i} \quad (6.23)$$

$$\frac{dC_{transit3,i}}{dt} = kb_{bile,i} \cdot C_{transit2,i} - kb_{bile,i} \cdot C_{transit3,i} \quad (6.24)$$

Small intestine lumen

$$\frac{dC_{SIL,i}}{dt} = kb_{bile,i} \cdot C_{transit3,i} - kLI_i \cdot C_{SIL,i} \quad (6.25)$$

Large intestine lumen

$$\frac{dC_{LIL,i}}{dt} = kLI_i \cdot C_{SIL,i} - (ka_i + k_{fecal,i}) \cdot C_{LIL,i} \quad (6.26)$$

Urine

$$\frac{dC_{urine,i}}{dt} = \frac{cl_{urine,i} \cdot C_{kidney,i}}{K_{kidney,i}} \quad (6.27)$$

Feces

$$\frac{dC_{feces,i}}{dt} = kfecal_i \cdot C_{LIL_i} \quad (6.28)$$

Here, $C_{T,i}$ represents the concentration of the compound i in a specific tissue T , V_T is the tissue volume, Q_T the blood flow rate, and $K_{T,i}$ the calculated organ-specific tissue partition coefficient, Eq. (6.4), for the compound i . Furthermore, $cl_{urine,i}$ and $cl_{bile,i}$ present renal and bile clearance for a specific compound, which are set to zero if the compound cannot be found in the urine or feces, respectively. The compound-specific constants ka_i , $kbile_i$, $kfecal_i$, and $kLIL_i$ are the large intestine, bile, and feces coefficients. These are set to zero if the compound cannot be found in the feces. This system of equations can then be solved by a commercial ode solver such as the `ode15s.m` function in MATLAB [153], which is a quasi-constant step size implementation of the numerical differentiation formulas in terms of backward difference [234].

6.1.3 Uncertainty quantification

Mathematical models, such as infant-WBMs and PBPK models, attempt to simulate complex biological processes. These models are developed using necessary assumptions about process behaviors, parameters from the literature, and parameter approximations. However, these assumptions and simplifications introduce uncertainties into the models. Uncertainty quantification (UQ) analyzes all sources of error and uncertainty in mathematical models. These methods assess the extent to which a model should be trusted rather than determining whether the model is correct or incorrect [235]. Model uncertainties are often divided into *aleatoric* and *epistemic* uncertainties. Aleatoric uncertainties describe a model's uncertainty inherent to the model, such as the natural variability in the system itself [235]. For the infant-WBM, such a natural variability could be the infant's birth weight, which varies greatly for an infant's population [30]. Epistemic uncertainties stem from a system's lack of knowledge, either because the model itself is incorrect or the parameters used in the models are incorrect [235]. These could be parameters such as the lipophilicity l in the PBPK models, which varies from different literature sources, and it needs to be clarified which estimation is correct. UQ can help assess an uncertain parameter's impact on a model by studying the relationship between the uncertainty in the output and input of a model [236].

Monte-Carlo Simulations

To analyze parametric uncertainties in a mathematical model m computational methods, such as Monte Carlo simulation, can be applied. This non-intrusive method propagates uncertainties through a mathematical model [237]. It is based on random sampling for which an underlying probability distribution with a probability density function f is chosen. This distribution could be a continuous uniform distribution in an interval $I = [a, b]$,

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b, \\ 0, & \text{for } x < a \text{ or } x > b. \end{cases} \quad (6.29)$$

Such a uniform distribution can be applied when there is a minimum and maximum value in which the parameter arbitrarily lies. Then, N input random variables $Z \in \mathbb{R}^d$ are generated according to the chosen probability density function f . The parameters $Z^{(i)}$ with $i = 1, \dots, N$ are used as input to the model $m(Z^{(i)})$. When evaluating the model output $m(Z^{(i)})$, the impact of the variation of the parameter in the interval I on the model output can be assessed. Using these model realizations, the quantity of interest, such as the expected value, can be computed [237],

$$\mathbb{E}[m(Z)] \approx \frac{1}{N} \sum_{i=1}^N m(Z^{(i)}). \quad (6.30)$$

6.2 Development of time-dependent infant-WBMs

The development of time-dependent infant-WBMs is based on the coupling of PBPK models and infant-WBMs such that metabolite concentration predictions can be used to update the constraints on the fluxes within the infant-WBM and vice versa. Integrated time-dependent metabolic models have been developed for adult metabolism, especially for a cell-based liver model [23] and for studying the gut-brain axis in autism spectrum disorder [105]. Furthermore, PBPK models have been combined with whole-body models to investigate glucose regulation in type 1 diabetes [238] and the effect of processes and genetic variations in ethanol metabolism [107]. This section describes in detail the development of time-dependent infant-WBMs, particularly how PBPK models and infant-WBMs can be integrated and used for dynamic IMD biomarker predictions.

6.2.1 Coupling of PBPK modeling and infant-WBMs

For the coupling of PBPK models and infant-WBMs, the flux through a metabolic reaction in the infant-WBM was interpreted as the intrinsic ability of a compound to be metabolized by the relevant enzyme. Hence, the flux was used to update the initial value of the initial value problem in the PBPK model for predicting the metabolite concentration, Figure 6.2. Likewise, the predicted metabolite concentrations were used to update the bounds on the flux values in the infant-WBM. This led to an iterative scheme, where both models were evaluated one after the other, and the predictions of one model were used to update the other. Through this iterative scheme, it was possible to combine all information from physiology, pharmacokinetics, and metabolism to obtain more accurate predictions on the metabolite concentrations. The following four steps describe the coupling of both models leading to time-dependent infant-WBMs.

Step 1. PBPK adaptations for infant physiology

PBPK models can be adapted for specific physiological parameters corresponding to an individual. These parameters were extracted from the *model.IndividualParameters* variable of the infant-WBM. Similar to Toroghi et al [23], the physiological and physicochemical properties in the PBPK model were assumed to be constant over time since only short time frames of several hours were evaluated. In particular, the PBPK parameters for the body mass, organ volumes, and organ flow rates were obtained directly from the infant-WBM models; see also chapter 5. The calculation of the organ volumes was based on the organ weights reported in the infant-WBM and the respective

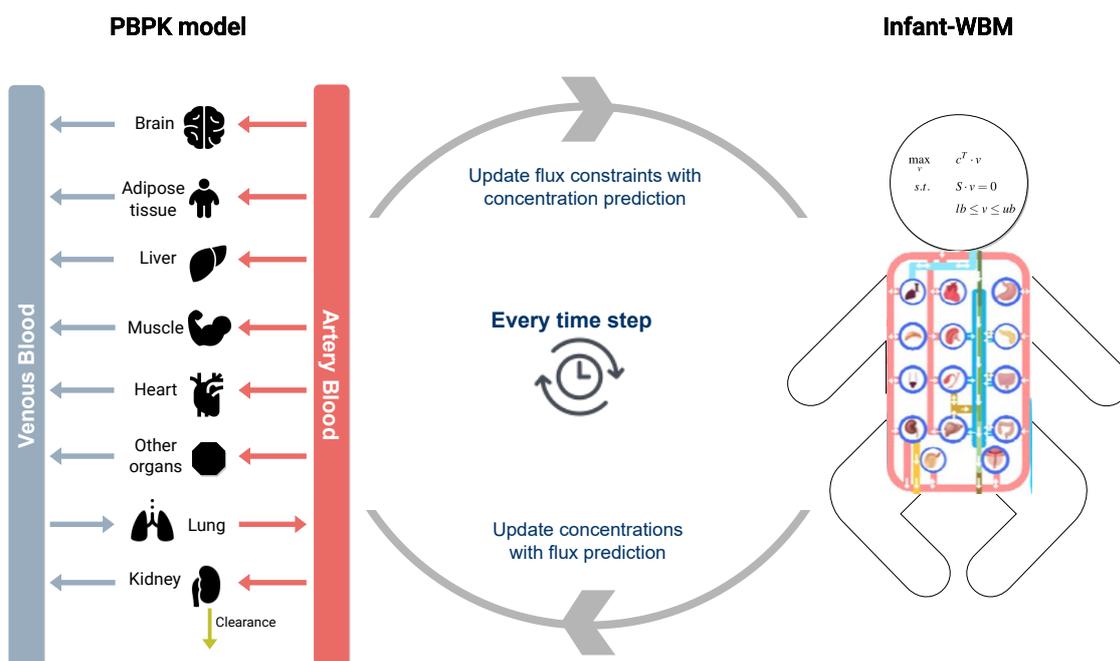


Figure 6.2: Schematic overview presenting coupling of PBPK models and infant-WBMs to account for time-dependent changes in metabolite concentrations. Icons from <https://icons8.de>.

organ density obtained from literature [239]. The venous and arterial blood volume estimations were based on the total blood volume. They resulted in 144 ml blood in both compartments for a 3.3 kg newborn at day 1 (total blood volume: 288 ml), which is comparable to the blood range of newborns at 72 hours of age (75 to 107 ml/kg) ranging between 248 - 353 ml for a 3.3 kg infant [240].

The organ flow rates were obtained from the organ flow rate parameters of the infant-WBM, which were measured in l/min/kg tissue and converted to l/h/organ. The blood flow rate for venous and arterial blood was estimated at 4.85 l/h. Hence, it is estimated to be higher than the blood flow rate ranging from 10 to 50 ml/min in infants weighing less than 5 kg [241], leading to an estimation of 0.6 - 3 l/h by multiplication with 60/1000. The body surface area was calculated based on the Mosteller method [242], using the average height of 50 cm for male and 49 cm for female newborns [243].

Step 2. Compound-specific parameters

In the second step, the compound-specific parameters were set. A compound of interest could be a drug or a biomarker, such as an amino acid or acylcarnitine. All processes were modeled with perfusion rate-limited kinetics, and the Poulin method [230] was used for calculating the tissue

partition coefficient K_t , section 6.1.2. For every compound, the molecular weight, lipophilicity l , and fractional unbound plasma f_p^u had to be determined. The molecular weight and lipophilicity l were obtained from databases such as HMDB [108]. The fractional unbound plasma f_p^u could be measured experimentally for amino acids [233] and acylcarnitines [244] evaluating the recovery rate, which was obtained by comparing results before and after spiking with known concentrations of standards in plasma [244]. Furthermore, the renal and bile clearance rate (l/h) was extracted from the literature for amino acids [245] and L-carnitine [246].

Step 3. Time-dependent model parameters

To integrate time dependency into the models, the length of the total time intervals and the time intervals calculated by the ODE solver had to be defined. Furthermore, the frequency of the update of the infant-WBM to the ODE model had to be determined. The ODE system of equations was solved with the `ode15s.m` function in MATLAB [153], [234].

Step 4. Iterative integration of PBPK and infant-WBM

After steps 1. - 3. an infant-specific PBPK model for a specific compound of interest was assembled. However, this model did not account for the intrinsic ability of a compound to be metabolized by the relevant enzymes as it did not take the metabolic system into consideration [228]. The integrated PBPK infant-WBM is able to account for this by iteratively updating the PBPK model with the flux predictions from the infant-WBM, which is similarly updated based on the predicted metabolite concentrations from the infant-WBMs. This integration is presented in the following sub-steps.

Step 4.1 Set initial values for compound concentrations in organs The initial value of the compound concentration of interest was set to a published measurement of this compound. This value was set for the concentration in every organ and compartment that had an exchange reaction from the organ to the blood compartment. Additionally, the concentration of the artery and venous blood compartment was set to the initial concentration. The renal and bile clearance ODEs were added to the PBPK model depending on the excretion ways of the compound of interest. All other tissue-specific ODEs from the full PBPK model were excluded from the compound-specific PBPK model.

Step 4.2 Compute metabolite concentration with PBPK model The PBPK model computed the metabolite concentration for a set time frame (e.g., 30 seconds, 1 minute). Hence, a dynamic concentration prediction on the PBPK model assembled in steps 1-3 was solved with the ode15s solver implemented in MATLAB [153], [234].

Step 4.3 Update bounds of infant-WBM with PBPK prediction After the ODE time interval, the predicted metabolite concentration (in mmol/l) was utilized to update the infant-WBM bounds. For this, the bounds on the fluxes in the infant-WBM were set precisely instead of arbitrary. Therefore, the metabolite concentration $m_{conc,m} \in \mathbb{R}_{\geq 0}^m$ was converted to $\mu\text{mol/l}$ by multiplying it with 1000. This metabolite concentration was multiplied by 0.9 to obtain a minimal metabolite concentration,

$$m_{min,m} = 0.9 \cdot m_{conc,m}, \quad (6.31)$$

$$m_{max,m} = m_{conc,m}. \quad (6.32)$$

Based on this, the uptake of metabolites from the blood into an organ was updated similarly to the newborn screening data integration; see also section 5.2.6. Hence, the lower bounds on the uptake flux $lb_{m,organ}$ from the blood circulation into individual organs were updated utilizing the organ-specific plasma flow rate (PFR_{organ}), Eq. (5.3),

$$lb_{m,organ} = (-1) \cdot \frac{m_{max,m}}{1000} \cdot PFR_{organ} \cdot \frac{60 \cdot 24}{1000}, \quad (6.33)$$

given in mmol/day/person. For the kidney, the lower bound $lb_{m,kidney}$ and upper bound $ub_{m,kidney}$ on its metabolite uptake reaction fluxes were calculated using the glomerular filtration rate (GFR); see also Table 5.1,

$$ub_{m,kidney} = (-1) \cdot \frac{m_{min,m}}{1000} \cdot GFR \cdot \frac{60 \cdot 24}{1000} \quad (6.34)$$

$$lb_{m,kidney} = (-1) \cdot \frac{m_{max,m}}{1000} \cdot GFR \cdot \frac{60 \cdot 24}{1000}, \quad (6.35)$$

and were given in mmol/day/person. Restricting the lower bound of the reactions restricted the uptake of the biomarker from the blood into the organ since the lower bounds of these reactions are negative. The arterial blood is the oxygenated blood transporting oxygen from the lungs to the organs. Hence, the concentration of the biomarker in the arterial blood was used to update the individual organ bounds of all organs except the lung. The lung obtains inflowing metabolites

from the venous blood. Hence, the concentration of the venous blood compartment was used to update the lower bounds of reaction fluxes to the lung.

Step 4.4 Run infant-WBM QP optimization After step 4.3, the infant-WBM was updated based on the PBPK model and could be evaluated. The whole-body biomass reaction was set to 1 since the model was evaluated in a very short time interval (seconds/minutes). In this time interval, no body growth was expected, and the model was supposed to maintain its weight. Then, the quadratic program (QP), section 5.1.3, for the infant-WBM was solved, minimizing the fluxes while adhering to all constraints to obtain a flux value v_j for each reaction j in the metabolic network.

Step 4.5 Update PBPK model with infant-WBM flux prediction From the QP solution, the predicted flux v through the exchange reaction of the biomarker was extracted. First, these fluxes were converted from mmol/person/day to mmol/volume/second. Second, the resulting value was divided by the respective organ volume, as the concentration change over time in the PBPK model is given per organ volume. By this, the converted flux \hat{v}_{organ} for each organ was obtained. Third, the organ-specific metabolite concentrations were updated in the PBPK model. Therefore, the respective flux \hat{v}_{organ} from the organ into the blood compartment,

$$\frac{dC_{organ,i}}{dt} = \frac{Q_{organ}}{V_{organ}} \left(C_{art,i} - \frac{C_{organ,i}}{K_{organ,i}} \right) - \hat{v}_{organ}, \quad (6.36)$$

was subtracted. The updated concentration was set as the initial concentration value for the subsequent PBPK evaluation. The time interval was updated by setting the starting point of the next time interval to the last point of the previous time interval, and the procedure was initiated again at step 4.1.

6.2.2 Assembly of time-dependent infant-WBMs

For model application, the time-dependent infant-WBM was applied to predict a biomarker concentration over two hours. Therefore, the biomarker for isovaleric aciduria (IVA), isovalerylcarnitine (C5), was chosen as the target compound for the time-dependent infant-WBM and its concentration was predicted in several organs. The model was assembled according to the developed four-step process.

Table 6.1: Calculated tissue partition coefficients for C5 in 16 different tissues calculated with Poulin’s method [230].

Tissue T	K_T
Adipose tissue	0.2
Artery	1
Brain	0.8559
Small intestine	0.7711
Large intestine	0.7711
Heart	0.8132
Kidney	0.8394
Liver	0.8124
Lung	0.8635
Muscle	0.8085
Pancreas	0.7119
Skin	0.7670
Spleen	0.8474
Stomach	0.8421
Bone	0.4654
Venous	1

Step 1 for C5 prediction: First, a generic male infant-WBM at age day one was created using the `create_newborn.m` function [34]. Then, this personalized model and its parameters were used to create an infant-specific PBPK model. The model was assigned a body weight of 3.3 kg, based on which the organ weights and organ-specific blood flow rates were extracted. The body surface area was then estimated as $0.2141m^2$ using the Mosteller method [242].

Step 2 for C5 prediction: In the second step, the compound-specific, in this case the C5-specific parameters, were obtained and used to parameterize the model. The molecular weight of C5 is 245.3153 mol [108]. Further, the predicted lipophilicity $l = -2$ [108] was applied. The fractional unbound plasma was determined based on experimental studies performed in literature as $f_p^u = 0.908$ for all tissues T [244]. The blood-to-plasma ratio was set to $B = 1$. Based on these parameters and the predefined organ volumes [231], the tissue partition coefficient K_T was calculated for C5 utilizing the Poulin method [230] and presented for different tissues T , Table 6.1.

Furthermore, it was assumed that C5 is only excreted through urine, not feces, since HMDB only lists the urine as excreta in the disposition section [108]. No data were available on the renal clearance of C5; hence, the clearance was estimated based on the renal clearance of free carnitine (C0). The renal clearance of C0 is 1 - 3 ml/min per adult person [246]. Taking 2 ml/min and dividing this by the average body surface area of $1.73m^2$, the renal clearance of C0 could be estimated as $1.16 \text{ ml/min}/1.73m^2$. The renal clearance of acylcarnitines is estimated to be 4 - 8 times higher than the renal clearance of C0 [247]. Hence, the renal clearance of C5 was estimated

to be 4.64, which is six-times higher than the renal clearance of C0.

Step 3 for C5 prediction: The PBPK model was evaluated for a time frame of five minutes in which the flux \hat{v}_{organ} was assumed to be constant. After five minutes, \hat{v}_{organ} was updated using a new flux prediction from the adapted infant-WBM. This step was repeated 24 times, resulting in a total C5 concentration prediction interval of two hours.

Step 4 for C5 prediction: In the case of C5, there only existed exchange reactions for the lung, heart, muscle, and kidney in the infant-WBMs. Hence, the PBPK model was reduced to these organs, the blood compartment, and the respective excretion pathways, resulting in the following reduced PBPK model for time-dependent C5 prediction:

$$\frac{dC_{heart}}{dt} = \frac{Q_{heart}}{V_{heart}} \left(C_{art} - \frac{C_{heart}}{K_{heart}} \right) - \hat{v}_{heart} \quad (6.37)$$

$$\frac{dC_{kidney}}{dt} = \frac{Q_{kidney}}{V_{kidney}} \left(C_{art} - \frac{C_{kidney}}{K_{kidney}} \right) - \frac{cl_{urine} \cdot C_{kidney}}{V_{kidney} \cdot K_{kidney}} - \hat{v}_{kidney} \quad (6.38)$$

$$\frac{dC_{lung}}{dt} = \frac{Q_{lung}}{V_{lung}} \left(C_{ven} - \frac{C_{lung}}{K_{lung}} \right) - \hat{v}_{lung} \quad (6.39)$$

$$\frac{dC_{muscle}}{dt} = \frac{Q_{muscle}}{V_{muscle}} \left(C_{art} - \frac{C_{muscle}}{K_{muscle}} \right) - \hat{v}_{muscle} \quad (6.40)$$

$$\frac{dC_{art,i}}{dt} = \frac{Q_{lung}C_{lung,i}}{V_{art} \cdot K_{lung,i}} - \frac{C_{art,i}Q_{art}}{V_{art}} \quad (6.41)$$

$$\frac{dC_{ven}}{dt} = \frac{Q_{heart}C_{heart}}{K_{heart}V_{ven}} + \frac{Q_{muscle}C_{muscle}}{K_{muscle}V_{ven}} + \frac{Q_{kidney}C_{kidney}}{K_{kidney}V_{ven}} - \frac{Q_{lung}C_{ven}}{V_{ven}} \quad (6.42)$$

$$\frac{dC_{urine}}{dt} = \frac{cl_{urine} \cdot C_{kidney}}{K_{kidney}} \quad (6.43)$$

6.2.3 Biomarker prediction with time-dependent infant-WBMs

In the time-dependent infant-WBM for C5 prediction, the initial concentration of C5 was set to 0.1 $\mu\text{mol/l}$, the healthy average of over 2 million newborns [31]. This initial concentration was set for all tissues T . The integrated analysis for the C5 prediction over a two-hour interval for the artery, heart, kidney, lung, muscle, and venous blood compartment of a healthy infant-WBM model is presented in Figure 6.3.

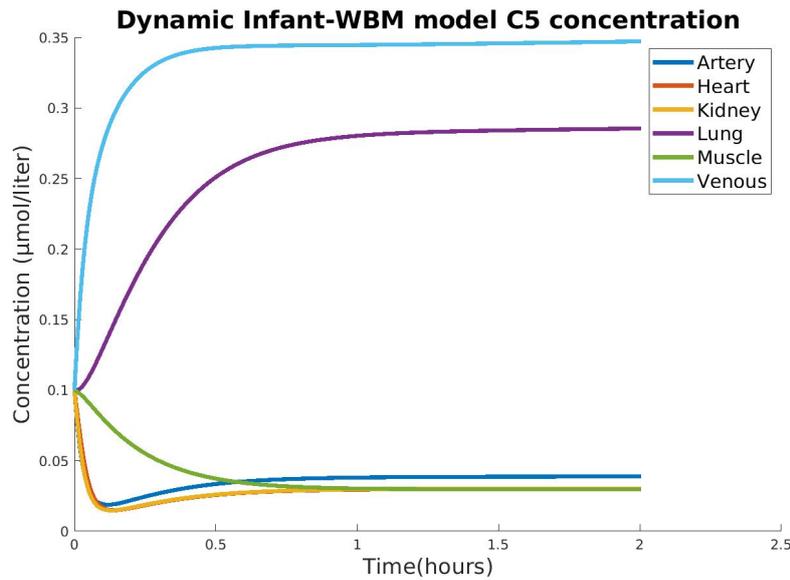


Figure 6.3: Concentration prediction of C5 from the time-dependent infant-WBM for the artery and venous blood compartment, as well as the heart, kidney, lung, muscle.

At time point 0, all tissues had a concentration of $0.1 \mu\text{mol/l}$. After two hours, the predicted venous blood compartment had the highest concentration with $0.34 \mu\text{mol/l}$. Followed by the lung with a predicted concentration of $0.29 \mu\text{mol/l}$, the artery blood compartment with a predicted concentration of $0.04 \mu\text{mol/l}$, as well as the muscle, heart, and kidney all with a predicted concentration of $0.03 \mu\text{mol/l}$. After approximately 1 hour, the predicted C5 concentration in all organs was almost constant. These predictions were difficult to verify as no organ-specific C5 measurements were available. However, all of the predicted values were beneath the 99th percentile ($0.51 \mu\text{mol/l}$) of blood C5 concentration used as cutoff value for IVA [248].

Furthermore, to investigate the applicability of time-dependent infant-WBMs, healthy and diseased infant models were compared. For this comparison, a time-dependent infant-WBM with IVA was simulated using the gene-knockout method for IMD modeling, section 5.1.5. The initial C5 concentration of the IVA model was set to $5 \mu\text{mol/l}$, the reported mean of 28 newborns with IVA [31]. The comparison of the wild type (WT) and IVA time-dependent infant-WBM for C5 over a two-hour interval for several tissues is presented in Figure 6.4. At time point 0, all compartments in the WT model had a C5 concentration of $0.1 \mu\text{mol/l}$, whereas all compartments in the IVA model had a set C5 concentration of $5 \mu\text{mol/l}$, Figure 6.4. After two hours, the predicted venous blood compartment had the highest concentration with $18 \mu\text{mol/l}$ (WT: $0.34 \mu\text{mol/l}$). Followed by the lung with a predicted concentration of $14 \mu\text{mol/l}$ (WT: $0.29 \mu\text{mol/l}$), the artery blood compartment which had a predicted concentration of $2 \mu\text{mol/l}$ (WT: $0.04 \mu\text{mol/l}$), as well as the

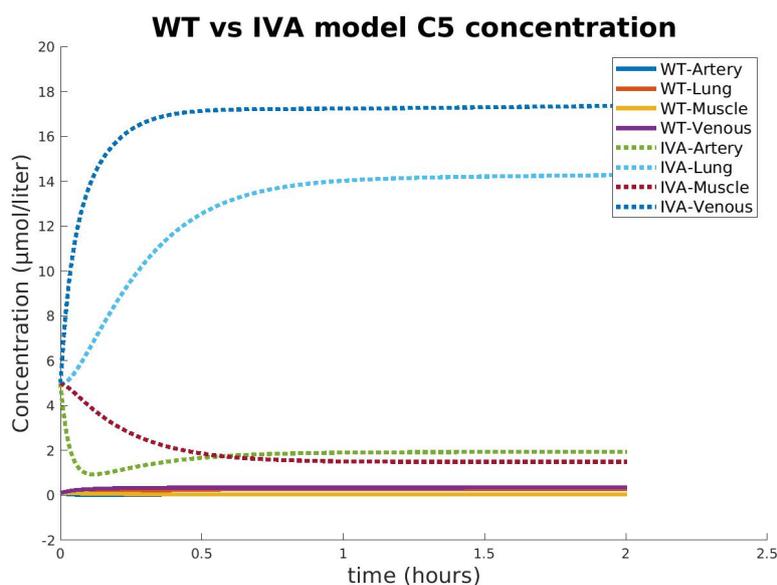


Figure 6.4: Concentration prediction of C5 from time-dependent infant-WBM for wild type (WT) and IVA models over the two hours interval for the artery, lung, muscle, and venous blood compartment.

muscle which had a predicted C5 concentration of 1.7 $\mu\text{mol/l}$ (WT: 0.03 $\mu\text{mol/l}$). Hence, the predictions of the IVA model in all compartments were higher than the 99.99th percentile (>1.37 $\mu\text{mol/l}$) of the blood C5 concentration used as cutoff value for IVA [248].

Furthermore, to show the personalization capabilities of the time-dependent infant-WBMs, measurements from newborn screening from a mild and classic IVA patient were used to personalize the infant-WBM model that was integrated into the PBPK model. For this, the initial concentration for all tissues in the mild IVA model was set to 2.6 $\mu\text{mol/l}$, the reported mean of 22 newborns with mild IVA, Supplementary Table B.2. In the classic IVA model, the initial concentration for all tissues was set to 12.6 $\mu\text{mol/l}$, the reported mean of 6 newborns with classic IVA.

The C5 concentration prediction after 2 hours was the highest in the classic IVA model, with predictions between 4 - 36 $\mu\text{mol/l}$, Figure 6.5. The C5 predictions of the mild IVA were lower, with values between 0.2 and 2 $\mu\text{mol/l}$. The classic IVA model predictions were above the 99th percentile (0.51 $\mu\text{mol/l}$) of the blood C5 concentration used as cutoff values for IVA [248]. In the mild IVA model, the predictions for the lung were above the 99.99th percentile (>1.37 $\mu\text{mol/l}$) of the blood C5 concentration. However, the prediction of the artery and muscle were beneath the 99th percentile (0.51 $\mu\text{mol/l}$) of the blood C5 concentration used as cutoff values for IVA [248]. These predictions could not be validated as no personalized data of C5 measurements of the individuals

for later time points or specific organs were available. Nevertheless, this application also showed that time-dependent infant-WBMs can be personalized using data from newborn screening.

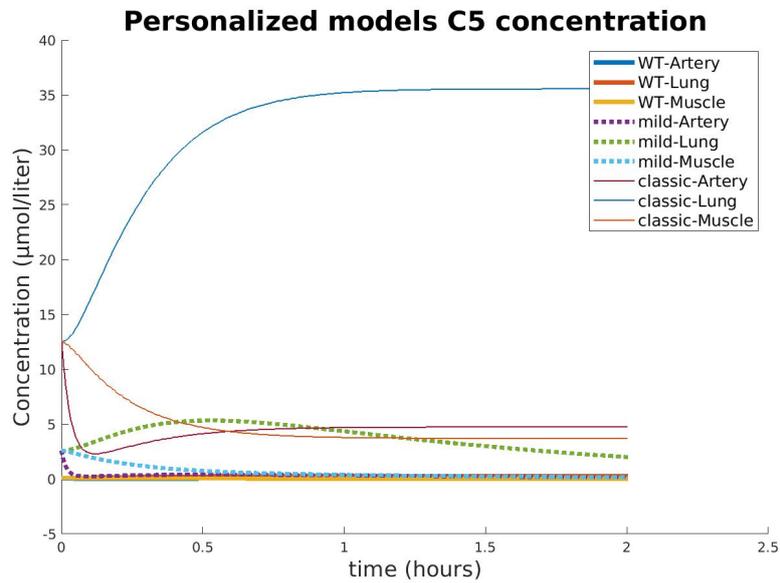


Figure 6.5: Concentration prediction of C5 from time-dependent infant-WBM for wild type (WT), mild IVA, and classic IVA models over the two hours interval for the artery, lung, and muscle blood compartment.

6.3 Uncertainty quantification for time-dependent infant-WBMs

In the previous section, time-dependent infant-WBMs were developed based on a PBPK formulation of the infant metabolism's pharmacokinetic interactions. This model was parameterized with parameters obtained from literature, which can introduce uncertainties into the models. Therefore, in this section, UQ methods will be employed to analyze the uncertainty of the time-dependent infant-WBM regarding specific parameters. By quantifying this uncertainty and its impact on the model's output, the reliability of the model when confronted with real-world data errors can be estimated.

6.3.1 Monte Carlo simulations for uncertain model parameters

In the time-dependent infant-WBMs, the parameters lipophilicity l , urinary clearance cl_{urine} , and unbound plasma flow rate f_p^u were obtained from the literature. A Monte Carlo simulation [237] with 100,000 realizations for every parameter was applied to analyze the uncertainties these parameters added to the model. From these, 100 randomly selected realizations were compared in line plots, showing the impact of the uncertain parameters over a two-hour interval. Since the metabolic flux predictions v showed a minor effect on the C5 prediction C_t , the realizations were only tested on the C5-specific PBPK modeling part for computational efficiency,

$$PBPK(l, cl_{urine}, f_p^u, C_{t-1})_t = C_t. \quad (6.44)$$

The C5 concentration in the venous compartment, representing the blood concentration, was evaluated for the prediction comparison. In the previous section, the lipophilicity was assumed to be -2 [108]. The fractional unbound plasma was determined based on experimental studies performed in literature as $f_p^u = 0.908$ for all tissues T [244]. The renal clearance of C5 was estimated as 4.64 [246], [247]. With these parameters fixed, the model predicted a C5 concentration in the venous compartment of 0.338 $\mu\text{mol/l}$ after two hours. For each Monte Carlo evaluation, the parameters that were not randomly sampled were fixed to these values.

Lipophilicity

The lipophilicity parameter refers to the capacity of C5 to dissolve in lipids. The lipophilicity was predicted as $l = -2$ using ALOGPS (<https://vcclab.org/lab/alogps/>) and $l = -3$ using ChemAxon (https://chemaxon.com/calculators-and-predictors#logp_logd) according to the HMDB [108]. Uniform sampling in the interval $I_l = [-3, -2]$ was applied to obtain 100,000 random samples for the lipophilicity parameter $Z_l^{(i)}, i = 1, \dots, 100,000$. These samples were then used as lipophilicity input parameters for the PBPK model to predict the C5 concentration over 2 hours in the venous compartment. The 100,000 C5 concentration predictions are presented in a histogram, Figure 6.6. The mean of the predicted concentration was $0.3386 \mu\text{mol/l}$, and the values were between $0.3384 - 0.3387 \mu\text{mol/l}$. Although the underlying distribution of the random

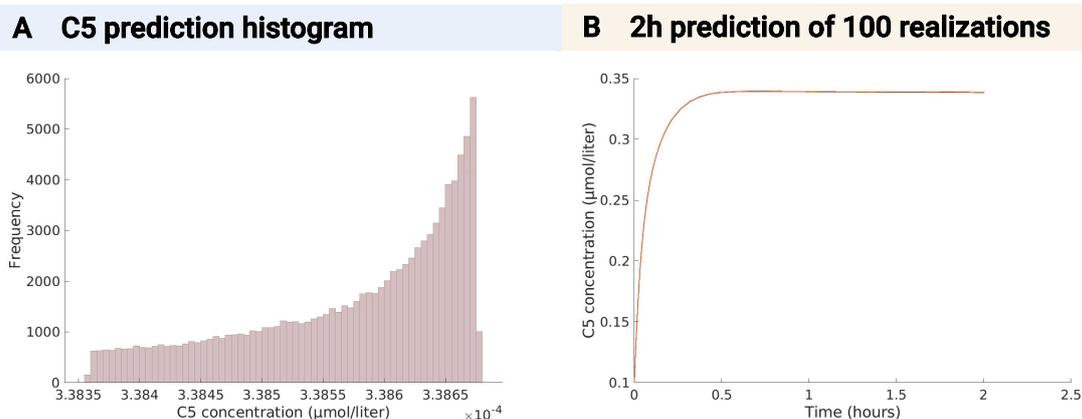


Figure 6.6: (A) Histogram of 100,000 random initializations of lipophilicity parameter evaluated in the venous blood concentration prediction after 2 hours. (B) 100 realizations of two-hour C5 predictions with randomly initialized lipophilicity parameter.

variables $Z_l^{(i)}$ is uniform, the C5 prediction was negatively skewed, $s = -0.7763$, (calculated with *skewness.m* [153]), Figure 6.6. The C5 predictions over two hours of 100 realizations showed minimal variability in the C5 prediction over time due to the variation in the lipophilicity parameter l , Figure 6.6.

Urine clearance rate

The urine clearance rate cl_{urine} describes the amount of urine excreted by an infant per time unit. The urine clearance rate of C5 is 4 - 8 times higher than the renal clearance of C0 [247], estimated as $1.16 \text{ ml/min}/1.73\text{m}^2$ [246]. This leads to interval boundaries for the renal clearance between 4.64 and 9.28, rounded to the interval $I_{cl} = [4.5, 9.5]$. Uniform sampling in this interval was applied to obtain 100,000 random samples for the renal clearance parameter, which was multiplied with the

infant-specific body surface area. The resulting value was used for the renal clearance parameter $Z_{cl_{urine}}^{(i)}$, $i = 1, \dots, 100,000$ to predict the C5 concentration over 2 hours in the venous compartment. The 100,000 C5 concentration predictions were presented in a histogram, Figure 6.7. The mean of the predicted concentration was $0.3383 \mu\text{mol/l}$, and the values were between $0.3317 - 0.3451 \mu\text{mol/l}$.

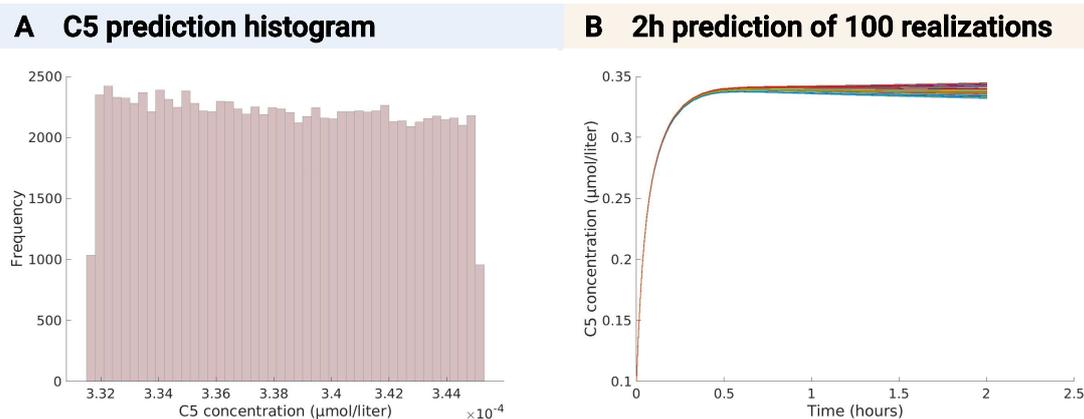


Figure 6.7: (A) Histogram of 100,000 random initializations of urine clearance rate parameter evaluated in the venous blood concentration prediction after 2 hours. (B) 100 realizations of two-hour C5 predictions with randomly initialized renal clearance rate parameter.

The histogram of the model output $PBPK(Z_{cl_{urine}}^{(i)})$ shows that the predictions are nearly uniform distributed with a slight positive skewness, $s = 0.0345$, (calculated with *skewness.m* [153]). The C5 predictions over two hours of 100 realizations showed that there was some variability due to the variation in the urine clearance rate cl_{urine} , Figure 6.7. For the first 30 min, the prediction was very similar for all 100 realizations, and with progressing time, the predictions showed more variation, Figure 6.7.

Fractional Unbound Plasma

The chemical bind in plasma or fractional unbound plasma, f_p^u , defines a drug's binding degree in plasma [232]. It describes the fraction of the compound that is not bound to plasma proteins and free for interaction with receptors, metabolizing enzymes, and renal filtration. The fractional unbound plasma f_p^u was determined based on a standard addition experiment performed in literature, as $f_p^u = 0.908$, $f_p^u = 0.933$, and $f_p^u = 1.04$ depending on the amount of compound addition [244]. Therefore, uniform sampling was applied in the interval $I_f = [0.9, 1.05]$ for 100,000 samples, $Z_{f_p^u}^{(i)}$, $i = 1, \dots, 100,000$. These samples were then used as fractional unbound plasma input parameters for the PBPK model to predict the C5 concentration over 2 hours in the venous

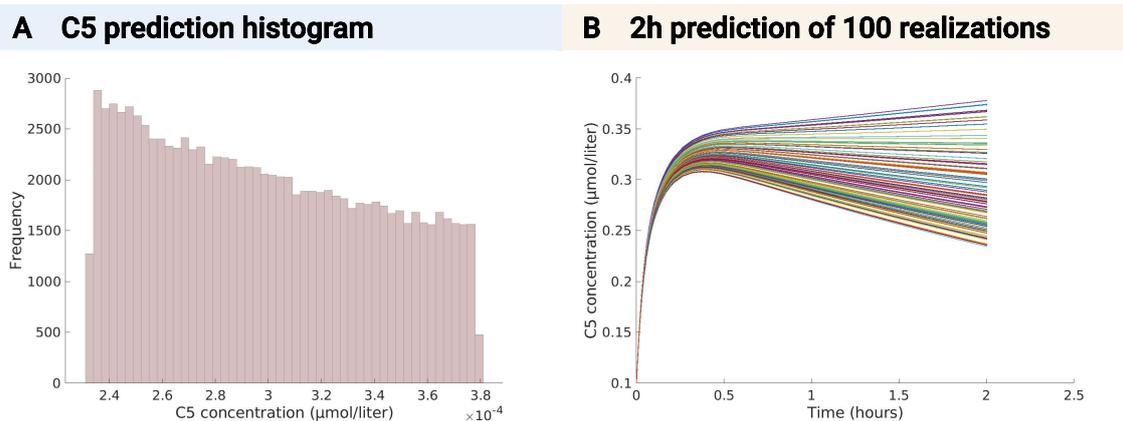


Figure 6.8: (A) Histogram of 100,000 random initializations of fractional unbound plasma parameter evaluated in the venous blood concentration prediction after 2 hours. (B) 100 realizations of two-hour C5 predictions with randomly initialized fractional unbound plasma parameter.

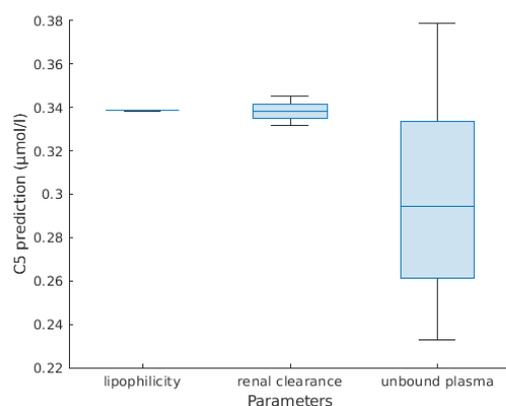


Figure 6.9: Boxplot of venous C5 prediction of PBPK model after 2 hours based on Monte Carlo Sampling for three uncertain parameters: lipophilicity, renal clearance rate, and unbound plasma rate.

compartment. The C5 concentration predictions for the 100,000 realizations are presented in a histogram, Figure 6.8. The mean of the predicted concentration was $0.2982 \mu\text{mol/l}$, and the values were between $0.2327 - 0.3789 \mu\text{mol/l}$. Unlike the underlying uniform distribution of the random variables $Z_{f_p}^{(i)}$, the model output $PBPK(Z_{f_p}^{(i)})$ shows a positive skewness $s = 0.2167$ (calculated with *skewness.m* [153]). Already starting after 10 minutes of prediction time, the 100 evaluated realizations showed variation in the concentration prediction, which increased with progressing time, Figure 6.8.

6.3.2 Evaluation of uncertain parameters

All three investigated parameters, lipophilicity l , urinary clearance cl_{urine} , and unbound plasma flow rate f_p^u , had uncertainties attached to them based on the methods and literature that were used to determine them. The model's prediction responded differently depending on the investigated parameter. Figure 6.9 shows a boxplot comparing the C5 predictions in venous blood after two hours from the UQ evaluations of all three parameters. The variation in the output due to the random sampling in the lipophilicity parameter l and urine clearance rate cl_{urine} is small compared to the impact the unbound plasma fraction f_p^u has on the prediction, Figure 6.9. The expected mean value of the C5 concentration after two hours is 0.34 $\mu\text{mol/l}$ with the UQ evaluation for the lipophilicity l , and 0.34 $\mu\text{mol/l}$ with the UQ for renal clearance cl_{urine} . However, the expected mean value of the C5 concentration after two hours is considerably lower with the UQ evaluation for f_p^u (0.30 $\mu\text{mol/l}$).

6.4 Summary and outcome

In this chapter, the development and application of time-dependent infant-WBMs were presented. These models are an extension of the infant-WBMs developed in chapter 5. They incorporate PBPK modeling, which allows them to account for the intrinsic ability of a compound to be metabolized by the relevant enzymes [228]. Using time-dependent infant-WBMs, the development of a biomarker concentration prediction over time can be analyzed. Further, the models enable a comparison of the concentration development for personalized and disease-specific models on different time scales. By taking the metabolite kinetics into account, changes in small time intervals and metabolic responses could also be modeled.

6.4.1 Modeling choices

For the integration of time dependencies, an existing PBPK model [105] developed for drug research was utilized. This PBPK model was then adapted for infant physiology based on measured data [165] and adult references. The choice of this PBPK model allowed the integration of 23 compartments and metabolic transportation pathways. For the prediction of C5, only the organs that had C5 blood exchange reactions in the infant-WBMs were integrated into the time-dependent infant-WBM. However, the chosen setup enables a flexible integration of different compartments, which can be selected depending on the investigated metabolite concentration. This allows for a compound-specific adaptation of the model, which is probably more accurate for each compound. Nevertheless, this could make model comparisons between different compounds difficult, as the PBPK part of the model is then based on different ODEs. Hence, the compound-specific inclusion and exclusion of ODEs into the model should be evaluated carefully for every use case. For the PBPK integration, it was assumed that the predicted flux through the infant-WBM did not change every time step and was only updated every five minutes for computational efficiency. When changing this evaluation interval to shorter intervals, the simulations showed no major differences in the C5 prediction. However, this assumption could be inaccurate in scenarios where a rapid change of the metabolite flux is expected in a short time interval and should be reevaluated for different applications in future studies.

The models were parameterized with multiple parameters calculated or obtained from experimental results in the literature. These parameters have uncertainties attached to them as a result of the experimental methods applied to obtain them. Hence, UQ for three of these parameters was

performed to quantify the corresponding uncertainty. The Monte Carlo method highlighted the impact of variations of lipophilicity l , urinary clearance cl_{urine} , and unbound plasma flow rate f_p^u parameters on the C5 predictions over two hours. Here, especially the unbound plasma flow rate f_p^u showed large variations in the output C5 prediction of the model after two hours. The mean C5 prediction of these realizations after two hours was $0.30 \mu\text{mol/l}$ and, therefore, considerably lower than the mean values of the other predictions ($0.34 \mu\text{mol/l}$). Hence, it should be ensured that the fractional unbound plasma rate f_p^u used in the time-dependent infant-WBM is estimated correctly, as this parameter strongly impacts the output of the model. The UQ analysis highlighted the importance of quantifying parameter uncertainties in mathematical models such as the time-dependent infant-WBMs.

6.4.2 Model limitations and future work

The time-dependent infant-WBMs enable a simulation of the dynamic behaviors of infant metabolism. Also, the field of IMD research is evolving from a 'static' view of simple pathways towards a 'dynamic' view of metabolic fluxes [35]. For IMDs, the time-dependent infant-WBMs showed their predictive capabilities by predicting the C5 concentrations in various organs for WT ('healthy') and IVA models. All values predicted by the WT model remained below the 99th percentile cutoff value ($0.51 \mu\text{mol/l}$) used in clinical practice for IVA [248]. However, the lack of organ-specific and time-dependent C5 measurements hinders quantitative model validation. In future applications, the prediction results need to be further analyzed and validated by clinical experts. Obtaining all the data necessary to personalize the time-dependent infant-WBMs from one individual would allow for an improved personalization of the models. Additional data, especially measurements from infants over several time intervals, is essential to enhance model validation in the future. This would contribute to the model's reliability and applicability. In this work, the time-dependent infant-WBMs were used to predict the biomarker of IVA and C5 over time and provide a proof of concept for dynamic biomarker evaluations in infants. Using this framework, future studies could investigate biomarkers of other inherited metabolic diseases associated with newborn screening.

Moreover, the developed framework can be used for modeling pharmacokinetics and pharmacodynamics of drugs specifically for infants as these processes are often different from adults [36], [37]. This could allow for infant-specific drug dosage determination, which enables to account for the infant's immature drug metabolism that is often associated with drug toxicity [225]. This is important since almost 50% of prescription drugs lack age-appropriate dosing guidelines [225]. Ad-

ditionally, pediatric drug clearance data is less attainable, which is probably due to the difficulties associated with conducting pediatric clinical trials [249], [250]. Hence, analyzing these drug-related metabolic processes *in silico* could be very beneficial for both researchers and patients.

In section 5.4.3, the infant-WBMs ability to predict the effect of dietary interventions on an IMD model were demonstrated. This kind of treatment analysis needs to be patient-specific since the disease therapy and dietary guidelines for IMD patients have to be adapted to age-dependent requirements and the patient's disease severity [35]. Here, personalized infant-WBMs can account for the substantial variability of the diseases in patients. The progression to time-dependent infant-WBMs enables *in silico* predictions of short-term metabolic responses of an individual's metabolism to the diet intake. By this, the models could support clinicians in personalized treatment planning and therapy for IMDs.

In conclusion, the time-dependent infant-WBMs showcased the potential of combining PBPK and infant-WBM models, opening new research directions for dynamic evaluations of infant metabolism. This capability holds significance for future infant-specific drug research and dietary treatment planning, where understanding the development of metabolite concentrations and the metabolic response to dietary interventions over time is crucial.

7

Singular value decomposition for infant-WBMs

In this chapter, first, the singular value decomposition (SVD) is introduced. Then, its application on the stoichiometric matrices of adult and infant metabolic models is demonstrated, and the biological meaning of the singular values revealed in the metabolic network structure is examined. Further, preliminary work on model reduction of metabolic networks based on the SVD is presented, which will be subsequently discussed in the final section.

7.1 Methodological background

First, the required methodological background on SVD is presented which is based on the works of Brunton et al [251], Golub et al [252], and Strang et al [253].

7.1.1 Singular value decomposition (SVD)

SVD presents a matrix factorization method that is used in various applications such as image processing or noise reduction, engineering, and kinematics. It is numerically reliable and efficient, guaranteeing its existence for any given matrix [251]. For high-dimensional data sets, SVD provides a systematic approach to determining low-dimensional approximations. No supervision or expert

knowledge is required for this, as SVD autonomously uncovers dominant patterns directly from the data. In general, SVD is performed to obtain a complex-valued matrix decomposition of a matrix $C \in \mathbb{C}^{m \times n}$. In this work, SVD will only be applied to real-valued matrices $A \in \mathbb{R}^{m \times n}$. SVD decomposes A into two orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Sigma \in \mathbb{R}^{m \times n}$. The vectors $u_i \in \mathbb{R}^m$ are the left singular vectors and $v_i \in \mathbb{R}^n$ the right singular vectors,

$$AV = \Sigma U. \quad (7.1)$$

The diagonal matrix Σ stores the singular values σ_i of A ,

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n},$$

where $p = \min(m, n)$. The singular values are arranged in decreasing order, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$. Due to the orthogonality of V , Eq. (7.1) can be transformed into the well-known formulation of the SVD using Theorem 7.1.1 stated by Golub [252].

Theorem 7.1.1 (Singular value decomposition) *If A is a real $m \times n$ matrix, then there exist orthogonal matrices*

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,m} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m,1} & u_{m,2} & \cdots & u_{m,m} \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,n} \end{bmatrix}$$

such that

$$U^T AV = \text{diag}(\sigma_1, \dots, \sigma_p) \quad \text{with} \quad p = \min(m, n),$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

Proof 7.1.2 *See Golub et al [252].*

The orthogonal matrices U and V are real quadratic matrices with orthonormal column and row vectors with respect to the scalar product. Hence, the inverse of these matrices is equal to the transposed matrix, $U^T = U^{-1}$. The rank r of matrix A represents the number of independent

columns in A . Hence, the column and row space is r -dimensional and there are r positive singular values, ordered in descending order $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ with,

$$A \cdot v_1 = \sigma_1 u_1, \dots, A \cdot v_r = \sigma_r u_r.$$

The $m - r$ columns of U are in the nullspace of A^T and the $n - r$ columns of V are in the null space of A ,

$$A \cdot v_{r+1} = \sigma_{r+1} u_{r+1}, \dots, A \cdot v_0 = \sigma_0 u_0.$$

Property 7.1.3 *If $U^T A V = \Sigma$ is the SVD of $A \in \mathbb{R}^{m \times n}$, then for $i = [1, \dots, n]$,*

$$A v_i = \sigma_i u_i \quad \text{and} \quad A^T u_i = \sigma_i v_i.$$

To see this property, compare columns in $AV = U\Sigma$ and $A^T U = V\Sigma^T$, see Golub et al [252].

7.1.2 SVD matrix approximation

For a given matrix A , the SVD provides an optimal low-rank approximation. This approximation creates a hierarchy of low-rank approximations since a rank- r approximation is obtained by keeping the first r singular values and vectors while discarding the rest. The approximation theorem of Schmidt, which was reformulated by Eckart-Young, Theorem 7.1.4, establishes the truncated SVD as the optimal low-rank approximation of the matrix A [251].

Theorem 7.1.4 (Eckart-Young) *The optimal rank- r approximation to A , in a least-squares sense, is given by the rank- r SVD truncation \tilde{A} :*

$$\underset{\substack{\tilde{A}, \\ \text{s.t. rank}(\tilde{A})=r}}{\operatorname{argmin}} (\|A - \tilde{A}\|_F) = \tilde{U} \tilde{\Sigma} \tilde{V}, \quad (7.2)$$

where \tilde{U} and \tilde{V} denote the first r leading columns of U and V , and $\tilde{\Sigma}$ contains the leading $r \times r$ sub-block of Σ , where $\|\cdot\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$ is the Frobenius norm.

Proof 7.1.5 *See Brunton et al [251].*

This theorem asserts that, as the singular values are arranged in order of importance, the sum of the first r components provides the optimal rank- r approximation of A . Additionally, $\sigma_1 u_1 v_1^T$

represents the closest rank-one matrix to A . By this, the truncated SVD basis will be denoted by $\tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{V}^T$,

$$\begin{aligned} \tilde{A} &= \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,r} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m,1} & u_{m,2} & \cdots & u_{m,r} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,r} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,r} \end{bmatrix}^T \\ &= \begin{bmatrix} \sigma_1 u_{1,1} & \sigma_2 u_{1,2} & \cdots & \sigma_r u_{1,r} \\ \sigma_1 u_{2,1} & \sigma_2 u_{2,2} & \cdots & \sigma_r u_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1 u_{m,1} & \sigma_2 u_{m,2} & \cdots & \sigma_r u_{m,r} \end{bmatrix} \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,r} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,r} \end{bmatrix}^T \\ &= \sum_{k=1}^r \sigma_k u_k v_k^T \\ &= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T, \end{aligned}$$

since, $\tilde{\Sigma}$ is a diagonal matrix. For a given rank r , this approximation is the best approximation for A in the L_2 norm. This means that high-dimensional data could be represented by a few dominant patterns given by the columns of \tilde{U} and \tilde{V} .

7.1.3 Computation of SVD

In this work, the exact SVD method *svd.m* embedded in MATLAB2020b [153] was applied for small matrices. For larger matrices such as the stoichiometric matrix of metabolic whole-body models, the reduced SVD method *svds.m* implemented in MATLAB2020b [153] was applied. This method is based on the Lanczos bidiagonalization with partial reorthogonalization algorithm (PROPACK package) [254].

7.2 SVD of whole-body models

The adult WBMs established by Thiele et al [20] and the infant-WBMs established in chapter 5, are constrained-based models. These models consist at their core of a stoichiometric matrix $S \in \mathbb{R}^{m \times n}$ which describes the metabolic reactions network of m metabolites taking part in n metabolic reactions,

$$S_{m,n} = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,n} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m,1} & s_{m,2} & \cdots & s_{m,n} \end{bmatrix}.$$

The stoichiometric matrix maps the vector of fluxes v onto the time derivatives of the metabolite concentrations x ,

$$\frac{dx}{dt} = Sv. \quad (7.3)$$

7.2.1 SVD of infant and adult whole-body models

A stoichiometric matrix S is a real matrix. According to Theorem 7.1.1 there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that S can be decomposed into three matrices, which compose the SVD,

$$S = U\Sigma V^T. \quad (7.4)$$

For the constrained-based model in Eq. (7.3), this implies that the application of V^T creates a linear combination of the fluxes v , which are then stretched by the diagonal matrix Σ . Then, U ensembles the linear combinations of the time derivatives,

$$\frac{dx}{dt} = U\Sigma V^T v. \quad (7.5)$$

For the application on WBMs, the SVD was applied to the stoichiometric matrix of the adult male WBM S^H (Harvey) [20] and the male infant-WBM S^{male} . The stoichiometry of the infant-WBM is based on the adult model, but extends this by adding growth-enabling reactions for all organs. Hence, the infant-WBMs consist of more reactions and metabolites. More specifically $S^H \in \mathbb{R}^{m^H, n^H}$ with $m^H = 57,897$ the number of metabolites and $n^H = 82,913$ the number of reactions, was decomposed to

$$S^H = U^H \Sigma^H (V^H)^T. \quad (7.6)$$

And the stoichiometric matrix of the infant-WBM $S^{male} \in \mathbb{R}^{m^{male}, n^{male}}$ with $m^{male} = 57,980$ the number of metabolites and $n^{male} = 83,149$ the number of reactions, was decomposed to

$$S^{male} = U^{male} \Sigma^{male} (V^{male})^T. \quad (7.7)$$

The singular values appear in decreasing order on the diagonal entries of Σ^H and Σ^{male} . Due to computational limitations, the full SVD for these matrices could not be calculated. The *svds.m* algorithm from MATLAB [153] was applied to obtain a r -reduced SVD containing the r largest singular values. By Theorem 7.1.4 this gives us the best rank- r approximation. Therefore, the reduced SVD with $r=100$ was computed for both matrices, S^H and S^{male} . Figure 7.1 shows the corresponding 100 largest singular values σ_i of both matrices.

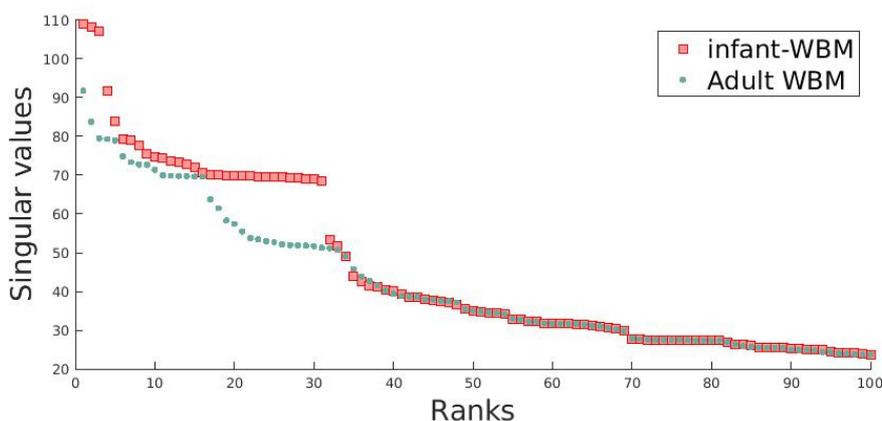


Figure 7.1: The 100 largest singular values of the adult WBM (green) and infant-WBM (red).

Interestingly, the infant-WBM had three singular values $\sigma_1, \sigma_2, \sigma_3$, which were larger than any singular value of the adult model, Figure 7.1. Moreover, from rank 37 onwards, the singular values of both matrices were nearly identical. In particular, the first 31 singular values in the infant-WBM had high values (above 68) compared to the subsequent singular values, which decreased quickly, Figure 7.1.

7.2.2 Mode-by-mode analysis

The mode-by-mode analysis of the SVD was investigated to gain further insights into the mechanisms underlying these differences. The SVD expresses S as a non-negative linear combination of $\min(m, n)$ rank-1 matrices, with the singular values providing the multipliers and the outer products of the left and right singular vectors providing the rank-1 matrices,

$$S = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i \cdot v_i^T. \quad (7.8)$$

The outer product of this equation $\langle u_i v_i^T \rangle$ is a matrix of rank 1 as it is obtained from a single vector, and each of these matrices has a norm at the order of unity $\| \cdot \| \approx 1$. Hence, the singular values directly give the relative contribution of each outer product to the reconstruction of the original matrix [79].

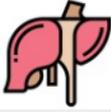
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Main Organ Involved					
Metabolites					
Glycogen	0.01	0.00	0.00	0.00	0.00
ADP	0.43	0.40	0.45	0.00	0.38
Acetyl-CoA	0.00	0.00	0.00	0.39	0.00
ATP	0.44	0.40	0.45	0.00	0.39
Carbon dioxide	0.01	0.01	0.01	0.01	0.03
Cholesterol	0.00	0.00	0.00	0.00	0.00
Coenzyme A	0.00	0.03	0.00	0.37	0.00
Elaidic acid	0.02	0.00	0.00	0.00	0.00
FAD	0.00	0.00	0.00	0.24	0.04
FADH	0.00	0.00	0.00	0.24	0.04
Hydrogen Ion	0.55	0.57	0.52	0.51	0.65
L-Glutamic acid	0.01	0.01	0.01	0.00	0.01
L-Leucine	0.01	0.01	0.01	0.00	0.01
L-Lysine	0.01	0.01	0.01	0.00	0.01
Malonyl-CoA	0.00	0.03	0.00	0.00	0.02
NAD	0.01	0.01	0.01	0.37	0.00
NADH	0.01	0.01	0.01	0.37	0.00
NADP	0.02	0.01	0.02	0.02	0.01
NADPH	0.02	0.01	0.02	0.02	0.01
Oxygen	0.01	0.01	0.01	0.00	0.01
Palmitic acid	0.02	0.00	0.00	0.00	0.00
Phosphate	0.43	0.40	0.44	0.00	0.38
Phosphatidylcholine	0.00	0.00	0.00	0.00	0.00
PE	0.00	0.00	0.00	0.00	0.00
Uridine 5'-diphosphate	0.01	0.01	0.01	0.00	0.01
Water	0.37	0.45	0.37	0.45	0.51

Figure 7.2: Modes corresponding to u_i vectors with highest singular values in mlae adult (Harvey) model. Colored according to the u_i values, where darker blue indicates a higher value.

In particular, the vectors u_i and v_i corresponding to the largest singular values were of interest. Hence, the u_i vectors corresponding to the five largest singular values were analyzed. The results are presented in Figure 7.2 for the male adult WBM and in Figure 7.3 for the male infant-WBM. In each evaluated mode the metabolites with the highest values belonged to the same organ. In the adult WBM, these were liver, kidney, muscle, and colon, Figure 7.2, whereas in the infant-WBM,

these were muscle, prostate, testis, liver, and kidney, Figure 7.3. Moreover, in both models, the metabolites corresponding to the largest u_i values were adenosine triphosphate (ATP), Hydrogen Ion, Phosphate, and water in all organs and additionally adenosine diphosphate (ADP) in the adult WBMs.

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Main Organ Involved	Muscle	Prostate	Testis	Liver	Kidney
					
Metabolites					
ATP	0.46	0.44	0.44	0.44	0.40
Carbon dioxide	0.00	0.00	0.00	0.01	0.03
Cholesterol	0.00	0.00	0.00	0.00	0.00
Coenzyme A	0.00	0.00	0.00	0.00	0.03
FAD	0.00	0.00	0.00	0.00	0.00
FADH	0.00	0.00	0.00	0.00	0.00
Hydrogen Ion	0.49	0.47	0.48	0.55	0.57
NAD	0.00	0.00	0.00	0.01	0.01
NADH	0.00	0.00	0.00	0.01	0.01
NADP	0.01	0.01	0.01	0.02	0.07
NADPH	0.01	0.01	0.01	0.02	0.07
Palmitic acid	0.00	0.00	0.00	0.02	0.00
Phosphate	0.45	0.44	0.44	0.43	0.40
Phosphatidylcholine	0.00	0.00	0.00	0.00	0.00
Phosphatidylethanolamine	0.00	0.00	0.00	0.00	0.00
Water	0.36	0.45	0.45	0.37	0.45

Figure 7.3: Modes corresponding to u_i vectors with highest singular values in male infant-WBM. Colored according to the u_i values, where darker blue indicates a higher value.

These represent the main metabolites that take part in the biomass reactions in the individual organs. For instance, the adult liver biomass reaction is presented by the following equation, using virtual metabolic human identifiers which can be found in the data base (<https://www.vmh.life/>) [180].

Liver-Biomass reaction: $20.3289 \text{ h}_2\text{o} + 26.4397 \text{ atp} + 0.3832 \text{ glu}_L + 0.3023 \text{ asp}_L + 0.03903 \text{ gtp} + 0.00017 \text{ adrn} + 2.1\text{e-}05 \text{ pa}_{\text{hs}} + 0.5441 \text{ ala}_L + 0.2838 \text{ asn}_L + 0.1332 \text{ cys}_L + 0.3586 \text{ gln}_L + 0.6217 \text{ gly} + 0.3292 \text{ ser}_L + 0.3292 \text{ thr}_L + 0.002295 \text{ arachd} + 0.4557 \text{ lys}_L + 8.5\text{e-}05 \text{ arach} + 0.2995 \text{ arg}_L + 0.1031 \text{ met}_L + 6\text{e-}06 \text{ pail}_{\text{hs}} + 0.03444 \text{ ctp} + 3\text{e-}05 \text{ pchol}_{\text{hs}} + 2.8\text{e-}05 \text{ pe}_{\text{hs}} + 0.03078 \text{ chsterol} + 1\text{e-}06 \text{ clpn}_{\text{hs}} + 0.000425 \text{ crvnc} + 0.005659 \text{ dgtp[n]} + 0.005659 \text{ dctp[n]} + 0.008488 \text{ datp[n]} + 0.02067 \text{ utp} + 2.1\text{e-}05 \text{ tag}_{\text{hs}} + 0.000425 \text{ dlnlcg} + 0.008488 \text{ dttp[n]} + 0.024225 \text{ elaid} + 0.001105 \text{ ddca} + 0.00289 \text{ ttdca} + 8.5\text{e-}05 \text{ ttdcea} + 0.030175 \text{ hdca} + 0.0051 \text{ hdcea} + 0.006205 \text{ ocdca} + 0.01054 \text{ lnlc} + 0.000255 \text{ lnlnca} + 8.5\text{e-}05 \text{ tmndnc} + 0.9399 \text{ glygn2} + 0.1327 \text{ his}_L + 0.1352 \text{ tyr}_L + 0.2778 \text{ ile}_L + 0.5186 \text{ leu}_L + 0.03431 \text{ trp}_L + 0.1863 \text{ phe}_L$

+ 0.2967 pro_L + 9e-06 ps_hs + 6e-06 sphmyln_hs + 0.3649 val_L + 1.6e-05 xolest2_hs + 0.000255 CE2510 + 0.00034 clpnd + 0.000255 ptdca + 8.5e-05 hpdca \rightarrow **26.4191 h** + **26.4191 adp** + **26.4191 pi** + 0.1431 ppi

The evaluation of the vectors v_i corresponding to the largest singular values further underscored the significance of the biomass reaction, emphasizing its important role in both models. This evaluation revealed that the highest values of the first v_i vectors corresponded to the biomass reactions of the organs identified in Figure 7.2 and Figure 7.3 for both models. In addition to the biomass reaction, the analysis highlighted further reactions. For the adult Harvey model, the Long-Chain-Fatty-Acid—Coa Ligase was associated with a high value in the first vectors v_i , whereas for the infant-WBM, the values corresponding to reaction Folylpolyglutamyl Synthetase had high values in the first vectors v_i .

7.2.3 Preliminary work on model reduction with SVD

Model reduction aims to lower the computational complexity of high-dimensional complex mathematical models, which pose challenges in numerical simulations. A reduced model is obtained by reducing the associated state space or decreasing the degrees of freedom. In this section, preliminary work on model reduction of the constrained growth optimization model using SVD is presented. The flux balance analysis (FBA), Eq. (5.2), applied for growth optimization in the infant-WBM was defined as

$$\begin{aligned} \max_v \quad & c^T v \\ \text{s.t.} \quad & S \cdot v = 0 \\ & lb \leq v \leq ub, \end{aligned}$$

where lb and ub are the lower and upper bounds on the fluxes v based on specific constraints, and c is a vector of weights indicating which and how much each reaction v contributes to the objective [81]. FBA is an n dimensional optimization problem. Model reduction aims at reducing the computational complexity of this problem while capturing the most important features of the model.

The approximated SVD revealed a small number of singular values with high values and a large number of singular values with low values. Therefore, the SVD was used to obtain a reduced model approximating the original problem. For the model reduction the rank- r approximated SVD was

computed, with $\hat{U} \in \mathbb{R}^{m \times r}$, the first r columns of U , $\hat{V} \in \mathbb{R}^{n \times r}$ the first r columns of V , and $\hat{\Sigma} \in \mathbb{R}^{r \times r}$ containing the r largest singular values,

$$S \approx \hat{U} \hat{\Sigma} \hat{V}^T.$$

The basis vectors in \hat{V} were then used to reduce the optimization problem to dimension r with $y \in \mathbb{R}^r$ by projecting the problem into an r -dimensional space. Therefore, a two-step process was followed. First, the steady-state condition needed to be satisfied in the new problem. Therefore, a quadratic optimization problem was formulated to obtain the optimal vectors $\Omega = [\omega_1, \dots, \omega_n]$ with $\omega_i \in \mathbb{R}^r$,

$$\begin{aligned} \min_{\omega_i \in \mathbb{R}^r} \quad & \|\omega_i\|^2 \\ \text{s.t.} \quad & S\omega_i = -S\hat{v}_i. \end{aligned}$$

By this, $S\omega_i = 0$ with $w_i = \hat{v}_i + \omega_i$ should always fulfill the steady-state condition. This system was solved applying the *lsqlin.m* function available in MATLAB 2023b [255]. In the second step, the computed matrix $W = [w_1, \dots, w_n]^T$ with $w_i \in \mathbb{R}^r$ was used to solve the problem:

$$\begin{aligned} \max_{y \in \mathbb{R}^r} \quad & c^T W y \\ \text{s.t.} \quad & S W y = 0 \\ & lb \leq W y \leq ub, \end{aligned}$$

for $y \in \mathbb{R}^r$. This system was solved by applying the Optimization Toolbox available in MATLAB 2023b [256]. To determine the length r of y , different values for r were evaluated and compared to the real flux vector v with an error analysis. For the error analysis, the absolute error \mathcal{E}_v and relative error \mathcal{R}_v of v were compared and presented in Table 7.1,

$$\mathcal{E}_v = \|(W y) - v\|, \quad \mathcal{R}_v = \frac{\|(W y) - v\|}{\|v\|}.$$

Both, the absolute and the relative error decrease with increasing r , Table 7.1. However, both errors are still very large, indicating that the reduced model does not adequately capture the features of the original model to compute an appropriate approximation of v .

Table 7.1: Absolute error \mathcal{E}_v and relative error \mathcal{R}_v of flux v optimization for different rank r values and solution after 1,000 steps of the ODE solver.

Ranks r	Absolute error \mathcal{E}_v	Relative error \mathcal{R}_v
10	23,810	0.9941
30	23,543	0.9829
50	23,480	0.9803
100	23,251	0.9708

7.3 Summary and outcome

This chapter applied SVD to analyze and compare the stoichiometric matrices of the male infant-WBM, S^{male} , and the male adult model S^H . For high-dimensional data sets, SVD provides a data-based systematic approach to determine low-dimensional approximations. No supervision or expert knowledge is required for this, as SVD autonomously uncovers dominant patterns directly from the data.

When evaluating both models' first 100 singular values, a similar pattern could be detected, Figure 7.1. Both models exhibited 31 large first singular values. The subsequent singular values rapidly decreased to smaller values. These results suggest that a few factors in both stoichiometric matrices represent dominant low-dimensional patterns within the matrix [251]. The columns u_i and v_i corresponding to the largest singular values were evaluated in the mode-by-mode analysis. This evaluation showed that each u_1, \dots, u_5 corresponded to one organ. This pattern was consistent for both the infant-WBM and the adult WBM. However, the models identified different organs associated with the largest modes. Further, the largest entries in u_i corresponded to the primary metabolites participating in the biomass reaction of these individual organs, Figure 7.2 and Figure 7.3. The biomass reactions couple the biomass maintenance and growth of the individual organs with the whole-body biomass growth reaction. Moreover, the evaluation of v_i showed that the largest entries in these vectors were associated with the biomass growth and biomass maintenance reactions of the previously identified organs. Therefore, the SVD reduction and subsequent mode-by-mode analysis have demonstrated their ability to discern biologically relevant dominant patterns from the stoichiometric matrix directly.

The SVD approximation transforms a matrix from its original high-dimensional space into a low-dimensional space, in which the dominant patterns lie [251]. For a given rank r , this approximation is the best approximation for A with the L_2 norm, Theorem 7.1.4. This concept proves highly beneficial across applications involving high-dimensional data represented in large matrices, such as image compression. In such scenarios, essential attributes of a matrix can be represented through low-dimensional patterns within the dataset [251]. This work demonstrated preliminary efforts on model reduction for the FBA growth problem for infant-WBMs. Here, the main goal was to decrease the size of the problem by projecting the solution into an r -dimensional space using the truncated matrix \hat{V} obtained from SVD. However, the reduced model did not predict a sufficiently accurate flux solution, resulting in large absolute and relative errors. Even when increasing r , both errors only decreased marginally. Hence, the proposed model reduction can not sufficiently

approximate the FBA flux prediction. However, this work provides a starting point in model reduction for metabolic networks and corresponding optimization problems and should be extended in future research. Nevertheless, a critical consideration lies in the uncertain impact of model reduction on the biological accuracy of a metabolic model. Determining how much qualitative information is lost through the model reduction and how this affects the subsequent evaluation will be essential. Therefore, future studies should clarify this aspect, enabling well-informed decisions regarding model order reduction for infant-WBMs.

8

Discussion and future works

In this dissertation, mathematical and data-based models have been developed to improve newborn screening and understand the complex infant metabolism associated with inherited metabolic diseases (IMDs). Therefore, five explicit research goals were defined in the introduction, chapter 1, to develop new approaches and tackle the current challenges in newborn screening research.

1. Develop data-based models for IVA and GA1 to reduce false-positive screening results
2. Develop explainable AI models for newborn screening
3. Develop metabolic whole-body models of infants
4. Develop time-dependent whole-body models of infants
5. Perform uncertainty quantification for whole-body models of infants

The first two goals mainly involved developing data-based models to improve the diagnostic accuracy of newborn screening for isovaleric aciduria (IVA) and glutaric aciduria type 1 (GA1). The three subsequent goals involved developing metabolic whole-body models for infants. This chapter will discuss the main findings, limitations, and implications for future research and provide an overall conclusion of this work.

8.1 Main findings and limitations

The main findings and limitations of this thesis are summarized individually for each of the five defined research goals in the following.

1. Data-based models for IVA and GA1 reduce false-positive newborn screening results

This work examined the development of data-based models to overcome current challenges regarding rising numbers of false-positive newborn screening results for IVA and GA1, and this aims at the main research goal 1 of this thesis. First, a systematic literature review on machine learning-based newborn screening applications was performed, published in a peer-reviewed journal [33] and summarized in chapter 2. It showed that, due to the low prevalence of the diseases and the associated low number of true positives, handling the large data imbalance is key for developing data-based approaches to improve newborn screening. In chapter 3, the development and application of machine learning methods for IVA and GA1 were presented. The results for IVA were published in a peer-reviewed journal [31]. A digital-tier strategy was proposed for both diseases to improve the classification results in newborn screening for IVA and GA1. This strategy simulates the scenario when machine learning classification methods are applied as an additional step after traditional newborn screening, focusing only on the suspected diagnosis newborn screening profiles. The algorithms were trained on over two million newborn screening profiles screened between 2002 - 2021. For both diseases, the developed method demonstrated a remarkable decrease in false-positive results, ranging from 70% to 76%, while maintaining a sensitivity of 100% on the training and validation data sets. Furthermore, this work presents the first instance of applying machine learning classification methods to distinguish between mild and classic IVA patients and obtained a similar reduction of false positives for both groups.

The proposed methods were further validated on an additional test data set containing newborn screening profiles screened at UKHD from 2022 and 2023. The validation results showed that the machine learning methods reduced false-positive newborn screening profiles for both diseases. However, one mild IVA patient was not classified correctly, whereas all other GA1 and IVA patients were classified correctly. Due to the scarcity of positive screening outcomes and a high prevalence of false-positive classifications, the algorithms could not be tested on a more extensive data set. Hence, the low number of true positives in the data set limited the work, and the proof of concept showed

how crucial in-depth model validation is since the methods must accurately detect all newborns affected by a newborn screening condition. Therefore, in the future, the proposed methods should undergo additional validation on independent data sets from recent years. Furthermore, how these data-driven methods could be legally and technically integrated into a clinical decision-support system needs to be evaluated.

2. Explainable AI methods identify metabolic patterns in newborn screening

In chapter 4, explainable artificial intelligence (XAI) methods were implemented to interpret machine learning-based newborn screening, which aimed at main research goal 2 of this thesis. The results were presented and subsequently published at an IEEE conference [32]. The interpretability of a machine learning algorithm's actions is not just theoretical but a crucial factor for adopting algorithms in the real world, especially in system-critical areas such as the medical domain. In this work, the explainable artificial intelligence methods SHAP [129] and LIME [142] were applied. These methods enabled a model-agnostic evaluation and, by this, a simple comparison of explanations between different classification algorithms. Both methods were shown to be suitable for application in newborn screening. To enhance the benefits of these methods, the results on group levels were evaluated by distinguishing the swarm plots of the SHAP values of different groups and introducing a cumulative LIME score. The evaluation of the SHAP values and cumulative LIME method confirmed biological knowledge, such as highlighting the known biomarker of newborn screening for IVA, as a significant feature for the full data set. Furthermore, the birth year was identified as an influential feature, which coincided with the increasing number of false-positive IVA screening results in recent years. This increase is due to the increased use of the antibiotic pivmecillinam for urinary tract infections in pregnant women [24]. However, both methods highlighted further metabolite concentrations, which were important for the correct prediction but are not known to be associated with IVA. The features highlighted by the cumulated LIME method varied depending on whether the XGBoost or NN classification method was applied. This discrepancy needs to be further analyzed by biochemical experts.

3. Mathematical whole-body models for infant metabolism predict body growth, biomarkers, and treatment strategies for inherited metabolic diseases

In this work, the first whole-body models for male and female infant metabolism were presented in chapter 5, which aimed at main research goal 3 of this thesis. These models have been published in the peer-reviewed journal *Cell Metabolism* [34]. The developed models represent the infant metabolism through a life span of six months modeled in an active state. During this period, the models were constrained to ensure that energy requirements and nutrition matched reference values from the literature. The energy requirements, mainly due to the energy demand for brain development, heart function, muscular activity, and thermoregulation, were essential to simulate an active metabolism. Based on these requirements, the infant-WBMs could accurately, in accordance with growth standards from the WHO [30], predict the growth trajectory during the first six months of life. The water balance and ATP synthase evaluation demonstrated that the infant-WBMs agreed with *in vivo* measurements of infants throughout the six-month time frame and presents an emergent feature of the infant-WBM. Related to newborn screening, three further applications were presented. The models' personalization capabilities were evaluated using 10,000 personalized infant WBMs utilizing the sex, birth weight, and 29 metabolite concentrations measured during newborn screening. Moreover, the infant-WBMs showed they could correctly predict known metabolic biomarkers for three IMDs at different time points. Furthermore, they proved their applicability for personalized treatment planning by predicting expected metabolic responses to simulated dietary interventions in IMD patients.

The most significant limitation of the developed model is the absence of personalized data from individual newborns and infants. Except for the sex, weight, and 29 metabolite concentrations, no further physiological information from one specific newborn was available to parameterize the models. Hence, the models were adapted using reported means, measurements from different infants, and relative values calculated from the reference adult models. Due to this lack of personalized data, the growth rate, metabolic fluxes, and biomarker predictions could only be evaluated qualitatively. Furthermore, the milk composition and amount of milk intake used for the infant-WBM nutrition were based on different literature sources [29], [180]. Hence, the milk composition in the infant-WBMs was adjusted. These adjustments for the model feasibility of the milk diet could be justified by the natural variability of the diet composition in newborns and infants. Still, they should be further validated in future studies.

Moreover, the infant-WBMs originated from both female and male adult WBMs. The included reactions were exclusively expanded for biomass growth reactions in each organ, highlighting a crucial distinction between adult and infant metabolism. No further stoichiometric changes were placed except for these reactions compared to the adult models. However, there should be more changes in the infantile stoichiometry since the metabolic profile changes at various stages of human life [222].

4. Model advancement with PBPK modeling enables the creation of time-dependent infant-WBMs

In chapter 6, time-dependent infant-WBMs were developed to incorporate dynamic metabolic changes in the infant-WBMs and achieved main research goal 4 of this thesis. This development showed how infant-WBMs operating in a steady state can be coupled with existing physiological-based pharmacokinetic (PBPK) modeling to enable time-dependent evaluations of infant metabolism. These models allowed for comparing the time-dependent concentration trends between personalized and disease-specific models across various time scales. For the integration of time dependencies, an existing PBPK model [105] developed for drug research was utilized. Based on measured data and adult references, this PBPK model was then adapted for infant physiology. The choice of this PBPK model allowed the integration of 23 compartments and metabolic transportation pathways. The setup enabled adapting the model for specific compounds, such as biomarkers of IMDs. However, this could limit model comparisons between different compounds, as the PBPK part of the model is based on different ordinary differential equations (ODEs). Hence, the compound-specific inclusion and exclusion of ODEs into the model should be evaluated carefully for future compounds of interest. For the PBPK integration, it was assumed that the predicted flux through the infant-WBM did not change every time step. For computational efficiency, it was only updated every five minutes of the infant's life. This assumption was sufficient in the investigated biomarker prediction. However, it could be inaccurate for predictions where a rapid change of the metabolite flux is expected in a short time interval and should be reevaluated when applied to different scenarios in future studies.

5. Uncertainty quantification for time-dependent infant-WBMs estimate parameter uncertainties

Furthermore, in chapter 6, uncertainty quantification (UQ) was performed to assess the developed time-dependent infant-WBMs and fulfill main research goal 5 of this thesis. The time-dependent infant-WBMs were parameterized with multiple parameters derived from calculations or extracted from experimental findings in existing literature [108], [246]. These parameters inherently carry uncertainties due to the experimental methodologies employed in their determination. Hence, UQ for three of these parameters was performed to quantify the corresponding uncertainty. The Monte Carlo method [237] was used to highlight the impact of variations of three model parameters with uncertain parameter ranges in literature. These parameters included the lipophilicity l , urinary clearance cl_{urine} , and unbound plasma flow rate f_p^u parameters on the C5 predictions over two hours. Here, especially the unbound plasma flow rate f_p^u showed large variations in the C5 prediction of the model after two hours. Hence, it should be ensured that the fractional unbound plasma rate f_p^u used in the time-dependent infant-WBM is estimated correctly, as this parameter strongly impacts the output of the model. In this application, the Monte Carlo method was sufficient for assessing the parameter uncertainty; however, in future UQ evaluations, more advanced techniques such as Quasi-Monte Carlo methods could be applied [257]. Overall, the UQ analysis highlights the importance of quantifying parameter uncertainties in mathematical models such as the time-dependent infant-WBMs.

6. Singular value decomposition reveals dominant patterns for infant-WBMs

Moreover, in chapter 7, the stoichiometric matrix of the infant-WBM was decomposed with a singular value decomposition to investigate the metabolic network structure of the infant model and compare it to the adult model. The largest singular values revealed a similar pattern in both models. These results suggested that a few factors in both stoichiometric matrices represent dominant low-dimensional patterns within the matrix [251]. Furthermore, for both models, the largest modes corresponded to different organs, and the largest entries to the primary metabolites participated in the biomass reaction of these individual organs. Therefore, the SVD reduction and subsequent mode-by-mode analysis showed their ability to discern biologically relevant dominant patterns from the stoichiometric matrix directly. Additionally, this work demonstrated preliminary

efforts on model reduction for the FBA growth problem for infant-WBMs based on SVD. Although the reduced model did not predict a sufficiently accurate flux solution, resulting in large absolute and relative errors, the work provides a starting point for research in model reduction for metabolic networks and should be extended in future research.

8.2 Implications and future work

This work developed new approaches for mathematical and data-based modeling, opening new avenues for future research in newborn screening. This section will discuss the implications and potential starting points for future work on data-based modeling in newborn screening and metabolic whole-body modeling for infants.

8.2.1 Data-based modeling in newborn screening

The proposed digital-tier strategy integrates machine learning methods into newborn screening diagnostics for IVA and GA1 and presents a methodology for how data-based methods can improve specificity in newborn screening. Based on the developed methodology, promising avenues for future research open up. For instance, other newborn screening conditions with high false-positive rates can be evaluated, potentially leading to significant advancements in the field. Moreover, a reduction of the false-positive rates for both investigated diseases, IVA and GA1, would reduce the burden of false-positive classifications for newborns and their families, ultimately mitigating the risks associated with over-treatment. This reduction yields additional positive effects by decreasing the demand for human and material resources. It would streamline the reporting and tracking workload for newborn screening laboratories. Importantly, the costs associated with hardware and software for implementing a digital-tier strategy are relatively low, offering a cost-effective solution. Therefore, this approach should be assessed alongside traditional newborn screening in daily practice to evaluate its potential cost-effectiveness. By this, the methods provide a first step to an AI-enhanced clinical-decision support system in newborn screening.

For the successful integration of machine learning methods into clinical routine there needs to be a high acceptance rate for the algorithms by the practitioner. In critical areas such as the medical domain, there is a high demand for medical professionals to understand how and why AI methods work and what their decisions are based on. Therefore, an algorithm's lack of interpretability could hinder and completely prevent its usage. This underscores the urgency and necessity of

this research. To overcome the challenges of non-interpretable algorithms, this work showed the first integration of XAI methods for machine learning-based newborn screening. This lays the basis for a more frequent application of XAI methods, which could lead to higher acceptance of machine learning methods in newborn screening and be valuable for future data-based research in newborn screening. Highlighting the most important metabolic features in an algorithm's decision can enhance its traceability and uncover valuable metabolic patterns. Patterns revealed by this approach might serve as starting points for future biochemical research in newborn screening diseases.

8.2.2 Metabolic whole-body modeling for infants

The developed metabolic models for infant metabolism (infant-WBMs) already demonstrated promising results in accurately representing infant metabolism and predicting metabolic functions. In future research, several avenues exist to build upon this research and further enhance and advance these models. In the adult WBMs, microbiome data were integrated, and a similar integration can be pursued for the infant-WBMs. This has the potential to provide a more accurate representation of infant metabolism, given the crucial role of gut microbiome interactions in infant development [258]. Additionally, by integrating the microbiome, the mode of birth could be considered, which is known to impact the newborn's microbiome. Vaginally delivered newborns have a microbiome more similar to the mother's vaginal microbiome. In contrast, the microbiome of C-section delivery newborns shows a more similar microbiome found in the environment and the mother's skin [259].

Future studies that collect personalized data, including various physiological parameters, urinary excretion, data from different time points, and information on nutrition intake and composition, can integrate this data into the developed models to simulate personalized infant metabolism. The models presented in this work can be seen as the first step to metabolic digital twins. The developed models present an infant up to age six months. However, based on the framework developed in this work, future researchers could extend the model to represent metabolism in children and adolescents at all life stages, from 6 months to 18 years, when retrieving the necessary data and literature information. In IMD research, understanding a patient's metabolism and the impact of a condition on an individual's metabolic processes is very valuable. The infant-WBMs demonstrated the ability to predict known metabolic biomarkers at various time points accurately. The precision observed in biomarker predictions suggests that the infant-WBMs can also accurately

predict biomarkers for other IMDs. However, achieving this would require an accurate representation of the impaired metabolism in the metabolic model and thorough clinical validation of the findings. Furthermore, personalized infant-WBMs could contribute to newborn screening for IVA. They could be employed to model the effects of antibiotics administered to pregnant women for treating urinary tract infections [24] and pivalate-containing creams [260] on infant metabolism. This approach could assist in distinguishing between true- and false-positive IVA screening results.

Moreover, in newborn screening, treatment strategies primarily involve dietary interventions, such as restricting the uptake of metabolites or providing supplements. These nutritional guidelines are meticulously planned for each patient because, due to the variability of IMDs, disease management and therapy must be personalized based on the patient's diagnosis [35]. The presented *in silico* evaluation of dietary treatment strategies with infant-WBMs marked the first simulation of dietary treatment for an IMD using metabolic models. In this context, the infant-WBMs accurately predicted qualitative flux changes resulting from dietary treatment strategies of varying intensity levels. This suggests that *in silico* testing can be used to manage IMDs in infants, allowing therapies to be assessed virtually. Future work can build upon the developed model to conduct *in silico* treatment evaluations for various IMDs and quantitatively compare these to data from treated IMD patients. In the era of precision medicine, this ability to accurately predict an infant's metabolic response to different dietary interventions holds significant potential for developing personalized nutritional strategies [38].

Furthermore, the developed time-dependent infant-WBM pave the way for research on metabolically modeling the pharmacokinetics of drugs tailored explicitly to infants, given that these processes often differ from those in adults. This application could facilitate the determination of infant-specific drug dosages, accounting for the immature drug metabolism frequently associated with drug toxicity [225]. Hence, the presented framework, enabling *in silico* analyses of these drug-related metabolic processes, is highly beneficial for researchers and infantile patients.

8.3 Conclusion

Overall, the developed mathematical and data-based modeling approaches can improve and support newborn screening and IMD research by enabling systematic and data-based analyses. The presented data-based methods demonstrate their applicability in newborn screening and are highly efficient in improving newborn screening by reducing false-positive screening results. The devel-

oped infant-WBMs present the first step towards a digital metabolic twin for infants. They enable personalized systematic simulations for newborns and infants with IMDs and open new avenues in IMD research. In the future, research in the field of clinical treatment planning for IMD patients and infant-specific drug modeling could benefit from the developed models as they enable precise *in silico* simulations of infant metabolism.



List of Abbreviations

Abbreviation	Description
17p	17-hydroxyprogesterone
AI	Artificial intelligence
ANOVA	Analysis of variance
ADP	Adenosine diphosphat
ATP	Adenosine triphosphat
C5	Isovalerylcarnitine
C8	Octanoylcarnitine
C10	Decanoylcarnitine
COBRA	Constraint-based reconstruction and analysis
CV	Cross-validation
FBA	Flux balance analysis
FN	False negatives
FP	False positives
GA1	Glutaric aciduria 1
GDPR	General data protection regulation
GEM	Genome-scale metabolic model
Hci	Homocitrulline

His	Histidine
HMDB	Human metabolome database
HMR	Human metabolic reaction series
IMD	Inherited metabolic disease
IVA	Isovaleric aciduria
IVD	Isovaleryl-CoA dehydrogenase
LDA	Linear discriminant analysis
LIME	Local interpretable model-agnostic explanations
LP	Linear program
LR	Logistic regression
ML	Machine learning
MMA	Methylmalonic aciduria
mmol	Millimol
MS/MS	Tandem mass spectrometry
NBS	Newborn screening
NN	Neural network
ODE	Ordinary differential equation
OMIM	Online mendelian inheritance in man
PBPK	Physiology-based pharmacokinetic
PDE	Partial differential equation
Phe	Phenylalanine
PK	Pharmacokinetic
PKU	Phenylketonuria
PPV	Positive predictive value
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PSCM	Physiologically and stoichiometrically constrained modeling
QP	Quadratic program
RF	Random forest
RR	Ridge logistic regression
SHAP	Shapley additive explanations
SVD	Singular value decomposition
SVM	Support vector machine
SWS	Gestational age

B

Supplementary tables

In this supplementary material, three additional tables are presented, which are associated with chapter 2 and chapter 3. Table B.1 depicts a summary of all studies analyzed in the systematic literature in chapter 2 on machine learning-based newborn screening and was presented in the published work Zaunseder et al [33]. The Tables B.2 and B.3 describe the mean metabolite concentrations recorded in newborn screening for IVA and GA1, chapter 3. Table B.2 was presented in the published work Zaunseder et al [31].

Table B.1: Summary of all reviewed studies on machine learning-based newborn screening. Diseases with * are biochemical variations nowadays known as non-diseases. Abbreviations: CAH - Congenital adrenal hyperplasia, CH - Congenital hypothyroidism, CIT1 - Citrullinemia type I, ASSD CIT2 - Citrullinemia type II, CPT1D- Carnitine palmitoyltransferase I deficiency, DT - Decision Tree, GA1 - Glutaric aciduria type I, IBDD - Isobutyryl-CoA dehydrogenase deficiency, IVA - Isovaleric aciduria, KNN - K-Nearest Neighbors, LCHADD - Long-chain hydroxyacyl-CoA deficiency, LDA - Linear Discriminant Analysis, LR - Logistic Regression Analysis, LVQ - Learned Vector Quantization, MADD - Multiple acyl-CoA dehydrogenase deficiency, MCADD - Medium-chain acyl-CoA dehydrogenase deficiency, 3-MCCD - 3-methylcrotonyl-CoA carboxylase deficiency, MET - Hypermethioninemia, MMA - Methylmalonic aciduria, MSUD - Maple syrup urine disease, NN - Neural Network, OTCD - Ornithine transcarbamylase deficiency, PA - Propionic aciduria, PCA - Principal Component Analysis, PKU - phenylketonuria, PTPSD - 6-Pyruvoyl-tetrahydrobiopterin synthetase deficiency, RF - Random Forest, RR - Logistic ridge regression, SCADD - Short-chain acyl-CoA dehydrogenase deficiency, SVM - Support vector machine, VLCADD - Very long-chain acyl-CoA dehydrogenase deficiency

Disease	Data	Feature	Feature	ML classification
	Imbalance	Construction	Selection	
PKU	Random sampling		Information gain	DT, LR [76]

MCADD, PKU	Random sampling		Information gain, Relief based	LDA, DT, KNN, LR, NN, SVM [71]
3-MCCD*, MCADD, PKU	Random sampling		Diagnostic flag	DT, LR [62]
3-MCCD*, PKU, GA1, MMA, PA, MCADD, LCHADD	Random sampling		Discriminatory threshold	KNN, LR, Naive Bayes, NN, SVM [12]
MCADD	Informed sampling	Arithmetic ratio	χ^2	Rule learner [61]
MMA			Pearson coefficient	SVM [69]
MMA	Random sampling		Pearson coefficient	SVM [66]
MCADD	Oversampling	Arithmetic ratio	Variable set optimization	DT, LR, RR [64]
PKU			Fisher score	SVM [70]
3-MCCD*, PKU, MET		Arithmetic ratio	Fisher score, Variable set optimization	SVM [65]
CIT1, CIT2, CPT1D, GA1, IBDD, IVA, MADD, MET, MMA, MSUD, PA, PKU, PTPSD, SCADD*, VLCADD	Random sampling, Oversampling, Informed sampling		χ^2 , ANOVA, Mutual information, L1-norm, Tree-based	Bagging, Boosting, DT, KNN, LDA, LR, RF, SVM [13]
MMA	Oversampling			RF [74]
SCADD*, MCADD, VLCADD		Arithmetic ratio	Discriminatory threshold	LR [72]
CH			χ^2 , Information gain, Expert consultation	Bagging, Boosting, DT, NN, SVM [68]
GA1, MMA, OTCD, VLCADD	Second tier			RF [14]
PKU		Arithmetic ratio	Pearson coefficient, LVQ	LR [73]
CAH	Informed sampling	PCA		DT [63]

Table B.2: Newborn screening data analysis showing mean values and standard deviation of 48 measured metabolites in $\mu\text{mol/l}$ for the groups of normal and false positive newborns as well as newborns with mild and classic IVA. The metabolite abbreviations can be found in Table 2.1.

Metabolites	Normal	False Positive	Mild IVA	Classic IVA
	($\mu\text{mol/l}$)	($\mu\text{mol/l}$)	($\mu\text{mol/l}$)	($\mu\text{mol/l}$)
Newborns (no.)	2,105,959	103	22	6
17p	9.2±7.22	8.0±4.2	8.8±4.67	14.9±8.72
TSH	2.6±1.75	2.7±1.86	2.9±2.09	0.7±0.47
BIO	0.4±0.09	0.5±0.1	0.5±0.06	0.5±0.09
3HMG	0.0±0.02	0.0±0.01	0.0±0.01	0.0±0.01
Ala	237.1±204.09	268.7±194.71	236.6±116.16	282.5±86.74
Arg	14.7±9.02	14.8±10.97	15.5±6.95	21.2±11.1
Asa	0.6±1.16	2.0±2.34	0.3±0.07	0.2±0.08
Asp	52.8±23.23	67.4±60.31	55.7±17.91	58.5±12.63
C0	21.9±11.35	20.8±15.98	21.1±8.49	24.5±12.39
C10	0.1±0.05	0.1±0.06	0.1±0.06	0.1±0.03
C10:1	0.1±0.05	0.1±0.06	0.1±0.07	0.1±0.02
C12	0.1±0.06	0.1±0.04	0.1±0.08	0.1±0.08
C14	0.2±0.08	0.2±0.09	0.3±0.11	0.3±0.06
C14:1	0.1±0.06	0.1±0.05	0.2±0.09	0.1±0.02
C14OH	0.1±0.03	0.0±0.03	0.1±0.06	0.1±0.04
C16	3.4±1.59	3.3±1.66	3.3±1.8	5.4±1.97
C16:1	0.2±0.08	0.1±0.1	0.2±0.07	0.2±0.04
C16:1OH	0.1±0.02	0.0±0.03	0.1±0.04	0.1±0.08
C16OH	0.0±0.02	0.0±0.03	0.0±0.03	0.0±0.01
C18	1.0±0.33	0.9±0.3	1.0±0.36	1.3±0.37
C18:1	1.1±0.6	1.2±0.7	1.0±0.61	2.0±0.99
C18:1OH	0.0±0.03	0.0±0.03	0.0±0.02	0.1±0.07
C18:2	0.1±0.1	0.2±0.1	0.1±0.06	0.3±0.26
C18OH	0.0±0.02	0.0±0.01	0.0±0.01	0.0±0.01
C2	26.4±12.72	22.8±12.53	24.1±9.56	30.5±11.97
C3	2.1±1.1	2.1±1.3	2.0±0.94	2.6±1.12
C4	0.2±0.14	0.3±0.33	0.2±0.13	0.3±0.13
C5	0.1±0.07	2.6±2.06	2.6±1.16	12.6±5.22
C5:1	0.0±0.02	0.0±0.02	0.0±0.02	0.0±0.02
C6	0.1±0.04	0.1±0.04	0.1±0.03	0.0±0.01
C8	0.1±0.05	0.1±0.06	0.1±0.02	0.1±0.05
C8:1	0.1±0.07	0.1±0.07	0.1±0.08	0.1±0.04
Cit	12.0±6.59	15.0±9.18	12.0±2.99	17.5±12.3
Glu	411.8±103.84	411.8±156.43	415.6±99.99	372.7±63.81
Glut	0.1±0.08	0.1±0.08	0.1±0.09	0.0±0.01
Gly	474.8±166.87	468.8±293.0	494.5±181.09	386.8±68.31
Hci	1.8±1.05	2.0±2.78	1.8±0.58	2.7±2.85
His	448.0±392.35	952.2±770.84	346.7±170.6	230.0±77.2
Leu+Ile	137.8±47.4	145.6±70.45	148.8±62.9	210.3±73.74
MeGlut	0.1±0.04	0.0±0.04	0.1±0.04	0.0±0.01
Met	17.5±7.96	20.6±9.28	17.0±10.84	23.0±8.54
Orn	76.0±78.38	83.5±70.38	75.6±74.08	33.7±63.66
Phe	46.6±13.06	55.5±15.5	45.5±14.01	66.3±29.08
Pro	904.4±440.26	681.6±713.35	985.1±360.82	1203.8±505.32
Thr	119.4±61.79	89.9±109.52	124.0±43.31	125.2±24.0
Trp	78.5±236.31	102.8±34.42	63.6±13.89	51.7±9.78
Tyr	81.3±37.95	96.8±40.64	70.8±34.75	170.2±147.11
Val	102.5±43.83	115.0±64.51	111.3±57.15	194.3±87.61

Table B.3: Newborn screening data analysis showing mean values and standard deviation of 48 measured metabolites in $\mu\text{mol/l}$ for the groups of normal and false positive newborns as well as newborns with GA1. The metabolite abbreviations can be found in Table 2.1.

Metabolites	Normal ($\mu\text{mol/l}$)	False Positive ($\mu\text{mol/l}$)	GA1 ($\mu\text{mol/l}$)
Newborns (no.)	1,025,459	485	9
17p	7 \pm 5.23	10 \pm 5.99	8 \pm 3.96
TSH	2 \pm 1.74	3 \pm 1.97	4 \pm 3.34
BIO	0 \pm 0.08	0 \pm 0.08	0 \pm 0.11
GALT	12 \pm 2.84	12 \pm 2.96	12 \pm 2.82
3HMG	0 \pm 0.01	0 \pm 0.01	0 \pm 0.01
Ala	158 \pm 50.93	212 \pm 166.09	210 \pm 45.29
Arg	15 \pm 9.4	23 \pm 41.63	23 \pm 13.44
Asa	0 \pm 1.58	1 \pm 2.18	1 \pm 1.84
Asp	52 \pm 23.32	64 \pm 77.04	59 \pm 14.88
C0	18 \pm 7.24	24 \pm 11.68	21 \pm 13.13
C10	0 \pm 0.03	0 \pm 0.11	0 \pm 0.03
C10:1	0 \pm 0.06	0 \pm 0.11	0 \pm 0.06
C12	0 \pm 0.06	0 \pm 0.13	0 \pm 0.1
C14	0 \pm 0.07	0 \pm 0.12	0 \pm 0.06
C14:1	0 \pm 0.06	0 \pm 0.14	0 \pm 0.06
C14OH	0 \pm 0.03	0 \pm 0.07	0 \pm 0.04
C16	3 \pm 0.93	4 \pm 1.43	3 \pm 0.65
C16:1	0 \pm 0.07	0 \pm 0.1	0 \pm 0.08
C16:1OH	0 \pm 0.02	0 \pm 0.03	0 \pm 0.04
C16OH	0 \pm 0.02	0 \pm 0.03	0 \pm 0.03
C18	1 \pm 0.28	1 \pm 0.42	1 \pm 0.44
C18:1	1 \pm 0.27	1 \pm 0.46	1 \pm 0.41
C18:1OH	0 \pm 0.02	0 \pm 0.02	0 \pm 0.02
C18:2	0 \pm 0.07	0 \pm 0.08	0 \pm 0.08
C18OH	0 \pm 0.01	0 \pm 0.01	0 \pm 0.01
C2	23 \pm 8.25	34 \pm 12.89	26 \pm 9.49
C3	1 \pm 0.83	2 \pm 1.47	2 \pm 1.07
C4	0 \pm 0.11	0 \pm 0.19	0 \pm 0.14
C5	0 \pm 0.05	0 \pm 0.09	0 \pm 0.08
C5:1	0 \pm 0.02	0 \pm 0.02	0 \pm 0.02
C6	0 \pm 0.04	0 \pm 0.06	0 \pm 0.05
C8	0 \pm 0.05	0 \pm 0.09	0 \pm 0.07
C8:1	0 \pm 0.05	0 \pm 0.1	0 \pm 0.06
Cit	12 \pm 5.59	16 \pm 7.16	17 \pm 5.6
Glu	399 \pm 105.29	423 \pm 152.61	523 \pm 107.94
Gly	450 \pm 164.27	527 \pm 209.12	578 \pm 190.9
Hci	1 \pm 0.78	2 \pm 1.77	2 \pm 1.41
His	569 \pm 475.69	811 \pm 703.58	650 \pm 354.27
Leu+Ile	124 \pm 32.85	143 \pm 78.94	143 \pm 35.58
MeGlut	0 \pm 0.04	0 \pm 0.14	0 \pm 0.06
Met	15 \pm 4.49	20 \pm 11.77	19 \pm 8.25
Orn	104 \pm 58.17	130 \pm 96.51	126 \pm 54.47
Phe	43 \pm 9.84	53 \pm 23.91	50 \pm 8.87
Pro	789 \pm 449.95	905 \pm 956.95	1023 \pm 493.43
Thr	105 \pm 66.18	104 \pm 97.95	122 \pm 65.72
Trp	76 \pm 29.91	92 \pm 45.35	81 \pm 19.78
Tyr	84 \pm 34.82	113 \pm 58.22	79 \pm 27.82
Glut	0 \pm 0.06	0 \pm 0.11	2 \pm 1.55
Val	90 \pm 26.38	99 \pm 49.89	98 \pm 32.83

Bibliography

- [1] J. Loeber, D. Platis, R. Zetterström, *et al.*, “Neonatal screening in Europe revisited: An ISNS perspective on the current state and developments since 2010,” *Int J Neonatal Screen*, vol. 7, no. 1, p. 15, 2021. DOI: /10.3390/ijns7010015.
- [2] G. Gramer, G. F. Hoffmann, and U. Nennstiel-Ratzel, *Das erweiterte Neugeborenencreening: Erfolge und neue Herausforderungen*, 1. Aufl. 2015. Wiesbaden: Springer, 2015. DOI: /10.1007/978-3-658-10493-1.
- [3] U. Spiekerkoetter and H. Krude, “Target diseases for neonatal screening in Germany—Challenges for treatment and long-term care,” *Deutsches Arzteblatt international*, vol. 119, 2022. DOI: /10.3238/arztebl.m2022.0075.
- [4] U. Mütze, L. Henze, F. Gleich, *et al.*, “Newborn screening and disease variants predict neurological outcome in isovaleric aciduria,” *Journal of Inherited Metabolic Disease*, vol. 44, no. 4, pp. 857–870, 2021. DOI: /10.1002/jimd.12364.
- [5] U. Mütze, S. Garbade, G. Gramer, *et al.*, “Long-term Outcomes of Individuals With Metabolic Diseases Identified Through Newborn Screening,” *Pediatrics*, vol. 146, no. 5, 2020. DOI: /10.1542/peds.2020-0444.
- [6] N. Wu, J. Phang, J. Park, *et al.*, “Deep neural networks improve radiologists’ performance in breast cancer screening,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1184–1194, 2020. DOI: /10.1109/TMI.2019.2945514.
- [7] S. McKinney, M. Sieniek, and V. Godbole, “International evaluation of an AI system for breast cancer screening,” *Nature*, vol. 577, pp. 89–94, 2020. DOI: /10.1038/s41586-019-1799-6.
- [8] M. Li, X. Fu, and D. Li, “Diabetes prediction based on XGBoost algorithm,” *IOP Conference Series: Materials Science and Engineering*, vol. 768, no. 7, 2020. DOI: /10.1088/1757-899x/768/7/072093.
- [9] D. Ardila, A. Kiraly, and S. Bharadwaj, “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” *Nature Medicine*, vol. 25, pp. 954–961, 2019. DOI: /10.1038/s41591-019-0447-x.

- [10] E. Svoboda, “Artificial intelligence is improving the detection of lung cancer,” *Nature*, vol. 587, S20–S22, 2020. DOI: /10.1038/d41586-020-03157-9.
- [11] G. Peng, Y. Tang, N. Gandotra, *et al.*, “Ethnic variability in newborn metabolic screening markers associated with false-positive outcomes,” *J Inherit Metab Dis*, vol. 43, no. 5, pp. 934–943, 2020. DOI: /10.1002/jimd.12236.
- [12] C. Baumgartner and D. Baumgartner, “Biomarker discovery, disease classification, and similarity query processing on high-throughput MS/MS data of inborn errors of metabolism,” *Journal of biomolecular screening*, vol. 11, 2006. DOI: /10.1177/1087057105280518.
- [13] B. Lin, J. Yin, Q. Shu, *et al.*, “Integration of machine learning techniques as auxiliary diagnosis of inherited metabolic disorders: Promising experience with newborn screening data,” *Collaborative computing: Networking, applications and worksharing*, vol. 292, pp. 334–349, 2019. DOI: /10.1007/978-3-030-30146-0_23.
- [14] G. Peng, Y. Tang, T. Cowan, G. Enns, H. Zhao, and C. Scharfe, “Reducing false-positive results in newborn screening using machine learning,” *Int J Neonatal Screen*, vol. 6, no. 1, 2020. DOI: /10.3390/ijns6010016.
- [15] A. Singh, S. Sengupta, and V. Lakshminarayanan, “Explainable deep learning models in medical image analysis,” *Journal of Imaging*, vol. 6, no. 6, 2020. DOI: /10.3390/jimaging6060052.
- [16] A. M. Antoniadis, Y. Du, Y. Guendouz, *et al.*, “Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review,” *Applied Sciences*, vol. 11, no. 11, 2021. DOI: /10.3390/app11115088.
- [17] N. C. Duarte, S. A. Becker, N. Jamshidi, *et al.*, “Global reconstruction of the human metabolic network based on genomic and bibliomic data,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 6, pp. 1777–1782, 2007. DOI: /10.1073/pnas.0610772104.
- [18] E. Brunk, S. Sahoo, D. Zielinski, *et al.*, “Recon3D enables a three-dimensional view of gene variation in human metabolism,” *Nature Biotechnology*, vol. 36, no. 3, pp. 272–281, 2018. DOI: /10.1038/nbt.4072.
- [19] J. L. Robinson, P. Kocabaş, H. Wang, *et al.*, “An atlas of human metabolism,” *Science Signaling*, vol. 13, no. 624, 2020. DOI: /10.1126/scisignal.aaz1482.
- [20] I. Thiele, S. Sahoo, A. Heinken, *et al.*, “Personalized whole-body models integrate metabolism, physiology, and the gut microbiome,” *Mol Syst Biol.*, vol. 16, no. 5, 2020. DOI: 10.15252/msb.20198982.

- [21] T. Shlomi, M. N. Cabili, and E. Ruppin, “Predicting metabolic biomarkers of human inborn errors of metabolism,” *Molecular Systems Biology*, vol. 5, no. 1, p. 263, 2009. DOI: /10.1038/msb.2009.22.
- [22] R. Pagliarini and D. di Bernardo, “A genome-scale modeling approach to study inborn errors of liver metabolism: Toward an in silico patient,” *Journal of Computational Biology*, vol. 20, no. 5, pp. 383–397, 2013. DOI: /10.1089/cmb.2012.0276.
- [23] M. K. Toroghi, W. R. Cluett, and R. Mahadevan, “A multi-scale model of the whole human body based on dynamic parsimonious flux balance analysis,” *IFAC-PapersOnLine*, vol. 49, no. 7, pp. 937–942, 2016. DOI: /10.1016/j.ifacol.2016.07.319.
- [24] S. Murko, A. D. Aseman, F. Reinhardt, *et al.*, “Neonatal screening for isovaleric aciduria: Reducing the increasingly high false-positive rate in germany,” *JIMD Reports*, vol. 64, no. 1, pp. 114–120, 2023. DOI: /10.1002/jmd2.12345.
- [25] R. Ensenauer, J. Vockley, J. Willard, *et al.*, “A common mutation is associated with a mild, potentially asymptomatic phenotype in patients with isovaleric acidemia diagnosed by newborn screening,” *The American Journal of Human Genetics*, vol. 75, no. 6, pp. 1136–1142, 2004. DOI: /10.1086/426318.
- [26] S. Moorthie, L. Cameron, G. Sagoo, J. Bonham, and H. Burton, “Systematic review and meta-analysis to estimate the birth prevalence of five inherited metabolic diseases,” *Journal of Inherited Metabolic Disease*, vol. 37, no. 6, pp. 889–898, 2014. DOI: /10.1007/s10545-014-9729-0.
- [27] F. Cabitza and G. Banfi, “Machine learning in laboratory medicine: Waiting for the flood?” *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 56, no. 4, pp. 516–524, 2018. DOI: /10.1515/cc1m-2017-0287.
- [28] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, “What clinicians want: Contextualizing explainable machine learning for clinical end use,” vol. 106, pp. 359–380, 2019. [Online]. Available: <https://proceedings.mlr.press/v106/tonekaboni19a.html>.
- [29] A. Nilsson, A. Mardinoglu, and J. Nielsen, “Predicting growth of the healthy infant using a genome scale metabolic model,” *npj Systems Biology and Applications*, vol. 3, no. 3, 2017. DOI: /10.1038/s41540-017-0004-5.
- [30] World Health Organization, *WHO child growth standards weight-for-age*, Accessed: 2022-11-23, 2023. [Online]. Available: <https://www.who.int/tools/child-growth-standards/standards/weight-for-age>.

- [31] E. Zaunseder, U. Mütze, S. F. Garbade, *et al.*, “Machine learning methods improve specificity in newborn screening for isovaleric aciduria,” *Metabolites*, vol. 13, no. 2, 2023. DOI: /10.3390/metabo13020304.
- [32] E. Zaunseder, U. Mütze, S. F. Garbade, S. Haupt, S. Kölker, and V. Heuveline, “Deep learning and explainable artificial intelligence for improving specificity and detecting metabolic patterns in newborn screening,” in *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2023, pp. 1566–1571. DOI: /10.1109/SSCI52147.2023.10371991.
- [33] E. Zaunseder, S. Haupt, U. Mütze, S. Garbade, S. Kölker, and V. Heuveline, “Opportunities and challenges in machine learning-based newborn screening—a systematic literature review,” *JIMD Reports*, 2022. DOI: /10.1002/jmd2.12285.
- [34] E. Zaunseder, U. Mütze, J. G. Okun, *et al.*, “Personalized metabolic whole-body models for newborns and infants predict growth and biomarkers of inherited metabolic diseases,” *Cell Metabolism*, vol. 0, 2024. DOI: /10.1016/j.cmet.2024.05.006.
- [35] B. Lanpher, N. Brunetti-Pierri, and B. Lee, “Inborn errors of metabolism: The flux from mendelian to complex diseases,” *Nature Reviews Genetics*, vol. 7, no. 6, pp. 449–459, 2006. DOI: /10.1038/nrg1880.
- [36] G. L. Kearns, S. M. Abdel-Rahman, S. W. Alander, D. L. Blowey, J. S. Leeder, and R. E. Kauffman, “Developmental pharmacology—drug disposition, action, and therapy in infants and children,” *New England Journal of Medicine*, vol. 349, no. 12, pp. 1157–1167, 2003. DOI: /10.1056/nejmra035092.
- [37] I. Mahmood, “Dosing in children: A critical review of the pharmacokinetic allometric scaling and modelling approaches in paediatric drug development and clinical settings,” *Clinical pharmacokinetics*, vol. 53, pp. 327–346, 2014. DOI: /10.1007/s40262-014-0134-5.
- [38] C. A. Argmann, S. M. Houten, J. Zhu, and E. E. Schadt, “A next generation multiscale view of inborn errors of metabolism,” *Cell Metabolism*, vol. 23, no. 1, pp. 13–26, 2016. DOI: /10.1016/j.cmet.2015.11.012.
- [39] R. Guthrie, “Blood Screening for Phenylketonuria,” *JAMA*, vol. 178, no. 8, pp. 863–863, 1961. DOI: /10.1001/jama.1961.03040470079019.
- [40] N. Dikow, B. Ditzen, S. Kölker, G. F. Hoffmann, and C. P. Schaaf, *Medizinische Genetik*, vol. 34, no. 1, pp. 13–20, 2022. DOI: doi:10.1515/medgen-2022-2113.

- [41] B. Therrell, C. Padilla, J. Loeber, *et al.*, “Current status of newborn screening worldwide: 2015,” *Seminars in Perinatology*, vol. 39, no. 3, pp. 171–187, 2015. DOI: /10.1053/j.semperi.2015.03.002.
- [42] J. Wilson and G. Jungner, “Principles and practice of screening for disease,” *WHO Public Health Papers*, vol. 34, 1968. [Online]. Available: <https://apps.who.int/iris/handle/10665/37650>.
- [43] P. Monostori, G. Klinke, S. Richter, *et al.*, “Simultaneous determination of 3-hydroxypropionic acid, methylmalonic acid and methylcitric acid in dried blood spots: Second-tier LC-MS/MS assay for newborn screening of propionic acidemia, methylmalonic acidemias and combined remethylation disorders,” *PLOS ONE*, vol. 12, no. 9, pp. 1–16, 2017. DOI: /10.1371/journal.pone.0184897.
- [44] C. Fernández-Lainez, J. Aguilar-Lemus, M. Vela-Amieva, and I. Ibarra, “Tandem mass spectrometry newborn screening for inborn errors of intermediary metabolism: Abnormal profile interpretation,” *Current medicinal chemistry*, vol. 19, 2012. DOI: /10.2174/092986712803251539.
- [45] L. Gómez-Romero, K. López-Reyes, and E. Hernández-Lemus, “The large scale structure of human metabolism reveals resilience via extensive signaling crosstalk,” *Frontiers in Physiology*, vol. 11, 2020. DOI: /10.3389/fphys.2020.588012.
- [46] F. Horn, *Biochemie des Menschen: das Lehrbuch für das Medizinstudium*. Thieme, 2009. [Online]. Available: https://books.google.de/books?id=U6_e1MW12_UC.
- [47] B. Tong and A. Barbul, “Cellular and physiological effects of arginine,” *Mini reviews in medicinal chemistry*, vol. 4, pp. 823–32, Nov. 2004. DOI: /10.2174/1389557043403305.
- [48] S. Li, D. Gao, and Y. Jiang, “Function, detection and alteration of acylcarnitine metabolism in hepatocellular carcinoma,” *Metabolites*, vol. 9, 2019. DOI: /10.3390/metabo9020036.
- [49] L. L. Jones, D. A. McDonald, and P. R. Borum, “Acylcarnitines: Role in brain,” *Progress in Lipid Research*, vol. 49, no. 1, pp. 61–75, 2010. DOI: /10.1016/j.plipres.2009.08.004.
- [50] A. Khan, B. Zahid, S. Khan, and S. Ahmad, “Isovaleric acidemia: A rare case of an inborn error of metabolism,” *Cureus*, vol. 12, 2020. DOI: /10.7759/cureus.7150.
- [51] A. Pinto, A. Daly, S. Evans, *et al.*, “Dietary practices in isovaleric acidemia: A european survey,” *Molecular Genetics and Metabolism Reports*, pp. 16–22, 2017. DOI: /10.1016/j.ymgmr.2017.02.001.

- [52] G. Gramer, M. Lindner, R. Santer, U. Spiekerkötter, and S. Tucci, *Leitlinie 027-021 Konfirmationsdiagnostik bei Verdacht auf angeborene Stoffwechselkrankheiten aus dem Neugeborenen-Screening*, Accessed: 01-08-2022. [Online]. Available: https://www.awmf.org/uploads/tx_szleitlinien/027-0211_S1_Konfirmationsdiagnostik-Stoffwechselkrankheiten-Neugeborenen-Screening_2020-05.pdf.
- [53] S. Kölker, E. Christensen, J. V. Leonard, *et al.*, “Diagnosis and management of glutaric aciduria type I – revised recommendations,” *Journal of Inherited Metabolic Disease*, vol. 34, no. 3, p. 677, 2011. DOI: /10.1007/s10545-011-9289-5.
- [54] K. Strauss, K. Williams, V. Carson, *et al.*, “Glutaric acidemia type 1: Treatment and outcome of 168 patients over three decades,” *Molecular Genetics and Metabolism*, vol. 131, 2020. DOI: /10.1016/j.ymgme.2020.09.007.
- [55] I. Baric, L. Wagner, P. Feyh, M. Liesert, W. Buckel, and G. Hoffmann, “Sensitivity and specificity of free and total glutaric acid and 3-hydroxyglutaric acid measurements by stable-isotope dilution assays for the diagnosis of glutaric aciduria type I,” *Journal of inherited metabolic disease*, vol. 22, no. 8, pp. 867–882, 1999. DOI: /10.1023/a:1005683222187.
- [56] G. F. Hoffmann, F. K. Trefz, P. G. Barth, *et al.*, “Glutaryl-coenzyme a dehydrogenase deficiency: A distinct encephalopathy,” *Pediatrics*, vol. 88, no. 6, pp. 1194–1203, 1991. DOI: /10.1542/peds.88.6.1194.
- [57] S. Kölker, E. Christensen, J. Leonard, *et al.*, “Guideline for the diagnosis and management of glutaryl-coa dehydrogenase deficiency (glutaric aciduria type I),” *Journal of inherited metabolic disease*, vol. 30, pp. 5–22, 2007. DOI: /10.1007/s10545-006-0451-4.
- [58] N. Boy, C. Mühlhausen, E. M. Maier, *et al.*, “Proposed recommendations for diagnosing and managing individuals with glutaric aciduria type I: Second revision,” *Journal of inherited metabolic disease*, vol. 40, pp. 75–101, 2017. DOI: /10.1007/s10545-016-9999-9.
- [59] N. Boy, C. Mühlhausen, E. M. Maier, *et al.*, “Recommendations for diagnosing and managing individuals with glutaric aciduria type 1: Third revision,” *Journal of Inherited Metabolic Disease*, vol. 46, no. 3, pp. 482–519, 2023. DOI: /10.1002/jimd.12566.
- [60] J. B. Hennermann, S. Roloff, J. Gellermann, A. Grüters, and J. Klein, “False-positive newborn screening mimicking glutaric aciduria type I in infants with renal insufficiency,” *Journal of Inherited Metabolic Disease*, vol. 32, no. S1, pp. 355–359, 2009. DOI: /10.1007/s10545-009-9017-6.

- [61] S. Ho, Z. Lukacs, G. Hoffmann, M. Lindner, and T. Wetter, "Feature Construction Can Improve Diagnostic Criteria for High-Dimensional Metabolic Data in Newborn Screening for Medium-Chain Acyl-CoA Dehydrogenase Deficiency," *Clinical Chemistry*, vol. 53, no. 7, pp. 1330–1337, 2007. DOI: /10.1373/clinchem.2006.081802.
- [62] C. Baumgartner, C. Böhm, and D. Baumgartner, "Modelling of classification rules on metabolic patterns including machine learning and expert knowledge," *Journal of Biomedical Informatics*, vol. 38, no. 2, pp. 89–98, 2005. DOI: /10.1016/j.jbi.2004.08.009.
- [63] M. Lasarev, E. Bialk, D. Allen, and P. Held, "Application of Principal Component Analysis to Newborn Screening for Congenital Adrenal Hyperplasia," *The Journal of Clinical Endocrinology and Metabolism*, vol. 105, no. 8, 2020. DOI: /10.1210/clinem/dgaa371.
- [64] T. Van den Bulcke, P. Vanden Broucke, V. Van Hoof, *et al.*, "Data mining methods for classification of Medium-Chain Acyl-CoA Dehydrogenase Deficiency (MCADD) using non-derivatized tandem MS neonatal screening data," *J. of Biomedical Informatics*, vol. 44, no. 2, pp. 319–325, 2011. DOI: /10.1016/j.jbi.2010.12.001.
- [65] W. Chen, S. Hsieh, K. Hsu, *et al.*, "Web-based newborn screening system for metabolic diseases: Machine learning versus clinicians," *Journal of Medical Internet Research*, vol. 15, no. 5, 2013. DOI: /10.2196/jmir.2495.
- [66] S. Hsieh, Y. Chien, C. Shen, *et al.*, "Newborn screening system based on adaptive feature selection and support vector machines," *2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering*, pp. 344–347, 2009. DOI: /10.1109/BIBE.2009.72.
- [67] V. Kumar and S. Minz, "Feature selection: A literature review," *Smart Computing Review*, vol. 4, no. 3, pp. 211–229, 2014. DOI: /10.6029/smartcr.2014.03.007.
- [68] S. Zarin Mousavi, M. Mohammadi Zanjireh, and M. Oghbaie, "Applying computational classification methods to diagnose congenital hypothyroidism: A comparative study," *Informatics in Medicine Unlocked*, vol. 18, 2020. DOI: /10.1016/j.imu.2019.100281.
- [69] S. Hsieh, S. Hsieh, Y. Chien, Z. Wang, Y. Weng, and F. Lai, "A newborn screening system based on service-oriented architecture embedded support vector machine," *Proceedings of the 2008 IEEE International Symposium on Service-Oriented System Engineering*, pp. 196–201, 2008. DOI: /10.1109/SOSE.2008.58.
- [70] W. Chen, H. Chen, Y. Tseng, *et al.*, "Newborn screening for phenylketonuria: Machine learning vs clinicians," *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 798–803, 2012. DOI: /10.1109/ASONAM.2012.145.

- [71] C. Baumgartner, C. Böhm, D. Baumgartner, *et al.*, “Supervised machine learning techniques for the classification of metabolic disorders in newborns,” *Bioinformatics*, vol. 20, no. 17, pp. 2985–2996, 2004. DOI: /10.1093/bioinformatics/bth343.
- [72] B. Wang, Q. Zhang, A. Gao, *et al.*, “New ratios for performance improvement for identifying acyl-coa dehydrogenase deficiencies in expanded newborn screening: A retrospective study,” *Frontiers in Genetics*, vol. 10, pp. 1–12, 2019. DOI: /10.3389/fgene.2019.00811.
- [73] Z. Zhu, J. Gu, G. Genchev, *et al.*, “Improving the diagnosis of phenylketonuria by using a machine learning-based screening model of neonatal mrm data,” *Frontiers in Molecular Biosciences*, vol. 7, 2020. DOI: /10.3389/fmolb.2020.00115.
- [74] G. Peng, P. Shen, N. Gandotra, *et al.*, “Combining newborn metabolic and DNA analysis for second-tier testing of methylmalonic acidemia,” *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, vol. 21, no. 4, pp. 896–903, 2019. DOI: /10.1038/s41436-018-0272-5.
- [75] D. Carvalho, E. Pereira, and J. Cardoso, “Machine learning interpretability: A survey on methods and metrics,” *Electronics*, vol. 8, no. 8, 2019. DOI: /10.3390/electronics8080832.
- [76] C. Baumgartner, D. Baumgartner, and C. Böhm, “Classification on high dimensional metabolic data: Phenylketonuria as an example,” *Proceedings of the IASTED International Conference on Biomedical Engineering*, 2004. [Online]. Available: <https://www.dbs.ifi.lmu.de/~boehm/publications/pku.pdf>.
- [77] S. Ho, “Optimierung diagnostischer Kriterien des Neugeborenen Screenings durch Anwendung von Data Mining Techniken,” *Inaugural Dissertation, Heidelberg University*, 2007.
- [78] B. M. Enuh, B. Nural Yaman, C. Tarzi, P. Aytar Çelik, M. B. Mutlu, and C. Angione, “Whole-genome sequencing and genome-scale metabolic modeling of *chromohalobacter canadensis* 85b to explore its salt tolerance and biotechnological use,” *MicrobiologyOpen*, vol. 11, no. 5, 2022. DOI: /10.1002/mbo3.1328.
- [79] B. O. Palsson, *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press, 2015. [Online]. Available: <https://www.bibsonomy.org/bibtex/268a8ccb55769a4b7833344c50785657c/karthikraman>.
- [80] N. Swainston, K. Smallbone, H. Hefzi, *et al.*, “Recon 2.2: From reconstruction to model of human metabolism,” *Metabolomics*, vol. 12, 2016. DOI: /10.1007/s11306-016-1051-4.
- [81] J. Orth, I. Thiele, and B. Palsson, “What is flux balance analysis?” *Nat Biotechnol*, vol. 28, no. 3, pp. 245–248, 2010. DOI: /10.1038/nbt.1614.

- [82] C. Kaas, Y. Fan, D. Weilguny, C. Kristensen, H. Kildegaard, and M. Andersen, "Toward genome-scale models of the chinese hamster ovary cells: Incentives, status and perspectives," *Pharmaceutical Bioprocessing*, vol. 2, pp. 437–448, 2014. DOI: /10.4155/PBP.14.54.
- [83] K. Baghalian, M.-R. Hajirezaei, and F. Schreiber, "Plant metabolic modeling: Achieving new insight into metabolism and metabolic engineering," *The Plant Cell*, vol. 26, no. 10, pp. 3847–3866, 2014. DOI: /10.1105/tpc.114.130328.
- [84] H. Ma, A. Sorokin, A. Mazein, *et al.*, "The Edinburgh human metabolic network reconstruction and its functional analysis," *Molecular systems biology*, vol. 3, p. 135, 2007. DOI: /10.1038/msb4100177.
- [85] I. Thiele, N. Swainston, R. Fleming, *et al.*, "A community-driven global reconstruction of human metabolism," *Nature biotechnology*, vol. 31, pp. 419–425, 2013. DOI: /10.1038/nbt.2488.
- [86] A. Mardinoglu, R. Agren, C. Kampf, *et al.*, "Integration of clinical data with a genome-scale metabolic model of the human adipocyte," *Mol Syst Biol.*, vol. 9, no. 649, 2013. DOI: /10.1038/msb.2013.5.
- [87] A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, M. Uhlen, and J. Nielsen, "Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease," *Nature communications*, vol. 5, no. 1, p. 3083, 2014. DOI: /10.1038/ncomms4083.
- [88] M. K. Aurich and I. Thiele, "Computational modeling of human metabolism and its application to systems biomedicine," *Systems medicine*, pp. 253–281, 2016. DOI: /10.1007/978-1-4939-3283-2_12.
- [89] S. Opdam, A. Richelle, B. Kellman, S. Li, D. C. Zielinski, and N. E. Lewis, "A systematic evaluation of methods for tailoring genome-scale metabolic models," *Cell systems*, vol. 4, no. 3, pp. 318–329, 2017. DOI: /10.1016/j.cels.2017.01.010.
- [90] K. Yizhak, T. Benyamini, W. Liebermeister, E. Ruppin, and T. Shlomi, "Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model," *Bioinformatics*, vol. 26, no. 12, pp. i255–i260, 2010. DOI: /10.1093/bioinformatics/btq183.
- [91] I. Thiele, A. Heinken, and R. M. Fleming, "A systems biology approach to studying the role of microbes in human health," *Current Opinion in Biotechnology*, vol. 24, no. 1, pp. 4–12, 2013, Analytical biotechnology. DOI: /10.1016/j.copbio.2012.10.001.

- [92] C. Gille, C. Bölling, A. Hoppe, *et al.*, “Hepatonet1: A comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology,” *Molecular systems biology*, vol. 6, no. 1, p. 411, 2010. DOI: /10.1038/msb.2010.62.
- [93] A. Bordbar, A. M. Feist, R. Usaita-Black, J. Woodcock, B. O. Palsson, and I. Famili, “A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology,” *BMC systems biology*, vol. 5, pp. 1–17, 2011. DOI: /10.1186/1752-0509-5-180.
- [94] A. Bordbar, N. Jamshidi, and B. O. Palsson, “iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states,” *BMC systems biology*, vol. 5, no. 1, pp. 1–12, 2011. DOI: /10.1186/1752-0509-5-110.
- [95] S. Sahoo and I. Thiele, “Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells,” *Human molecular genetics*, vol. 22, no. 13, pp. 2705–2722, 2013. DOI: /10.1093/hmg/ddt119.
- [96] M. W. Covert and B. O. Palsson, “Constraints-based models: Regulation of gene expression reduces the steady-state solution space,” *Journal of Theoretical Biology*, vol. 221, no. 3, pp. 309–325, 2003. DOI: /10.1006/jtbi.2003.3071.
- [97] C. Frezza, L. Zheng, O. Folger, *et al.*, “Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase,” *Nature*, vol. 477, no. 7363, pp. 225–228, 2011. DOI: /10.1038/nature10363.
- [98] K. Yizhak, E. Gaude, S. Le Dévédec, *et al.*, “Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer,” *eLife*, vol. 3, C. V. Dang, Ed., 2014. DOI: /10.7554/eLife.03641.
- [99] A. Heinken and I. Thiele, “Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework,” *Gut Microbes*, vol. 6, no. 2, pp. 120–130, 2015. DOI: /10.1080/19490976.2015.1023494.
- [100] R. Mahadevan and C. Schilling, “The effects of alternate optimal solutions in constraint-based genome-scale metabolic models,” *Metabolic Engineering*, vol. 5, no. 4, pp. 264–276, 2003. DOI: /10.1016/j.ymben.2003.09.002.
- [101] C. Angione, “Human systems biology and metabolic modelling: A review—from disease metabolism to precision medicine,” *BioMed Research International*, vol. 2019, 2019. DOI: /10.1155/2019/8304260.

- [102] A. Bordbar, J. T. Yurkovich, G. Paglia, O. Rolfsson, Ó. E. Sigurjónsson, and B. O. Palsson, “Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics,” *Scientific reports*, vol. 7, no. 46249, 2017. DOI: /10.1038/srep46249.
- [103] A. A. Mannan, Y. Toya, K. Shimizu, J. McFadden, A. M. Kierzek, and A. Rocco, “Integrating kinetic model of *E. coli* with genome scale metabolic fluxes overcomes its open system problem and reveals bistability in central metabolism,” *PloS one*, vol. 10, no. 10, 2015. DOI: /10.1371/journal.pone.0139507.
- [104] A. Zelezniak, S. Sheridan, and K. R. Patil, “Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes,” *PLoS computational biology*, vol. 10, no. 4, 2014. DOI: /10.1371/journal.pcbi.1003572.
- [105] F. K. Mohammad, M. V. Palukuri, S. Shivakumar, R. Rengaswamy, and S. Sahoo, “A computational framework for studying gut-brain axis in autism spectrum disorder,” *Frontiers in Physiology*, vol. 13, 2022. DOI: /10.3389/fphys.2022.760753.
- [106] M. B. Guebila and I. Thiele, “Dynamic flux balance analysis of whole-body metabolism for type 1 diabetes,” *Nature Computational Science*, vol. 1, no. 5, pp. 348–361, 2021. DOI: /10.1038/s43588-021-00074-3.
- [107] L. Zhu, W. Pei, I. Thiele, and R. Mahadevan, “Integration of a physiologically-based pharmacokinetic model with a whole-body, organ-resolved genome-scale model for characterization of ethanol and acetaldehyde metabolism,” *PLOS Computational Biology*, vol. 17, e1009110, 2021. DOI: /10.1371/journal.pcbi.1009110.
- [108] D. S. Wishart, T. Jewison, A. C. Guo, *et al.*, “HMDB 3.0—The Human Metabolome Database in 2013,” *Nucleic Acids Research*, vol. 41, no. D1, pp. D801–D807, 2012. DOI: /10.1093/nar/gks1065.
- [109] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021. DOI: /10.1007/s12525-021-00475-2.
- [110] E. R. Girden, *ANOVA: Repeated measures*. Sage, 1992.
- [111] GeeksforGeeks, *ML linear discriminant analysis*, Accessed: 2024-02-05. [Online]. Available: <https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/>.
- [112] G. Chen, *Math 253: Mathematical methods for data visualization; linear discriminant analysis (lda)*, Accessed: 2024-03-18. [Online]. Available: <https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec11lda.pdf>.

- [113] DATAtab e.U., *Datatab: Online statistics calculator; logistische regression*, Accessed: 2024-02-05. [Online]. Available: <https://datatab.de/tutorial/logistische-regression>.
- [114] A. Addagatla, *Maximum likelihood estimation in logistic regression*, Accessed: 2024-02-05. [Online]. Available: <https://arunaddagatla.medium.com/maximum-likelihood-estimation-in-logistic-regression-f86ff1627b67>.
- [115] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, pp. 503–528, 1989. [Online]. Available: <https://api.semanticscholar.org/CorpusID:5681609>.
- [116] S. Le Cessie and J. Van Houwelingen, "Ridge estimators in logistic regression," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 41, no. 1, pp. 191–201, 1992. DOI: /10.2307/2347628.
- [117] H. Šinkovec, G. Heinze, R. Blagus, and A. Geroldinger, "To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets," *BMC Medical Research Methodology*, vol. 21, 2021. DOI: /10.1186/s12874-021-01374-y.
- [118] GeeksforGeeks, *Support vector machine (svm) algorithm*, Accessed: 2023-11-23. [Online]. Available: <https://www.geeksforgeeks.org/support-vector-machine-algorithm/>.
- [119] P. Flach, *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge University Press, 2012, pp. 215–286.
- [120] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. DOI: /10.5555/1593511.
- [121] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: <https://www.jmlr.org/papers/v12/pedregosa11a.html>.
- [122] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: /10.1038/s41592-019-0686-2.
- [123] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review," *GESTS International Transactions on Computer Science and Engineering*, vol. 30, pp. 25–36, 2005. [Online]. Available: <https://api.semanticscholar.org/CorpusID:14354369>.
- [124] R. Amirkhan, M. Hoogendoorn, M. E. Numans, and L. Moons, "Using recurrent neural networks to predict colorectal cancer among patients," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–8. DOI: /10.1109/SSCI.2017.8280826.

- [125] Q. Zheng, H. Delingette, and N. Ayache, “Explainable cardiac pathology classification on cine mri with motion characterization by semi-supervised learning of apparent flow,” *Medical Image Analysis*, vol. 56, pp. 80–95, 2019. DOI: /10.1016/j.media.2019.06.001.
- [126] A. B. Tosun, F. Pullara, M. J. Becich, D. L. Taylor, J. L. Fine, and S. C. Chennubhotla, “Explainable AI (xAI) for anatomic pathology,” *Advances in Anatomic Pathology*, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:219706238>.
- [127] Xgboost developers, *Xgboost*, Accessed: 2023-03-12. [Online]. Available: <https://xgboost.readthedocs.io>.
- [128] A. Paszke, S. Gross, F. Massa, *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, 2019.
- [129] S. Lundberg and S. Lee, “A unified approach to interpreting model predictions,” *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, pp. 4768–4777, 2017. [Online]. Available: <https://dl.acm.org/doi/10.5555/3295222.3295230>.
- [130] N. Kokhlikyan, V. Miglani, M. Martin, *et al.*, “Captum: A unified and generic model interpretability library for pytorch,” *CoRR*, 2020. [Online]. Available: <https://arxiv.org/abs/2009.07896>.
- [131] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16, San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794. DOI: /10.1145/2939672.2939785.
- [132] Analytics Vidhya Blog, *Introduction to xgboost algorithm in machine learning*, Accessed: 2023-10-25. [Online]. Available: <https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/>.
- [133] R. Shwartz-Ziv and A. Armon, “Tabular data: Deep learning is not all you need,” *Information Fusion*, vol. 81, pp. 84–90, 2022. DOI: /10.1016/j.inffus.2021.11.011.
- [134] G. Bebis and M. Georgiopoulos, “Feed-forward neural networks,” *IEEE Potentials*, vol. 13, no. 4, pp. 27–31, 1994. DOI: /10.1109/45.329294.
- [135] A. Mao, M. Mohri, and Y. Zhong, *Cross-entropy loss functions: Theoretical analysis and applications*, 2023. DOI: /10.48550/arXiv.2304.07288.

- [136] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv*, 2016. DOI: /10.48550/arXiv.1609.04747.
- [137] T. Tieleman and G. Hinton, “Rmsprop gradient optimization,” 2014. [Online]. Available: http://www.cs.toronto.edu/tijmen/csc321/slides/lecture%5C_slides%5C_lec6.pdf.
- [138] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. DOI: /10.48550/arXiv.1412.6980.
- [139] T. Yu and H. Zhu, *Hyper-parameter optimization: A review of algorithms and applications*, 2020. DOI: /10.48550/arXiv.2003.05689.
- [140] A. Das and P. Rad, “Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey,” 2020. [Online]. Available: <http://arxiv.org/abs/2006.11371>.
- [141] K. Young, G. Booth, B. Simpson, R. Dutton, and S. Shrapnel, “Deep neural network or dermatologist?” In *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support*, Springer International Publishing, 2019, pp. 48–55. DOI: /10.48550/arXiv.1908.06612.
- [142] M. Ribeiro, S. Singh, and C. Guestrin, “”Why Should I Trust You?”: Explaining the predictions of any classifier,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, pp. 1135–1144, 2016. DOI: /10.1145/2939672.2939778.
- [143] M. S. Kovalev, L. V. Utkin, and E. M. Kasimov, “SurvLIME: A method for explaining machine learning survival models,” *Knowledge-Based Systems*, vol. 203, 2020. DOI: /10.1016/j.knosys.2020.106164.
- [144] A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on explainable artificial intelligence (xai),” *IEEE Access*, vol. 6, pp. 52 138–52 160, 2018. DOI: /10.1109/ACCESS.2018.2870052.
- [145] A. V. Dorogush, V. Ershov, and A. Gulin, “Catboost: Gradient boosting with categorical features support,” *CoRR*, 2018. [Online]. Available: <http://arxiv.org/abs/1810.11363>.
- [146] G. Ke, Q. Meng, T. Finley, *et al.*, “Lightgbm: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 2017, pp. 3146–3154. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>.

- [147] C. Frye, C. Rowat, and I. Feige, “Asymmetric shapley values: Incorporating causal knowledge into model-agnostic explainability,” in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/hash/0d770c496aa3da6d2c3f2bd19e7b9d6b-Abstract.html>.
- [148] M. Sundararajan and A. Najmi, “The many shapley values for model explanation,” in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 119, 2020, pp. 9269–9278. DOI: /10.48550/arXiv.1908.08474.
- [149] D. Janzing, L. Minorics, and P. Blöbaum, “Feature relevance quantification in explainable AI: A causal problem,” in *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*, ser. Proceedings of Machine Learning Research, vol. 108, PMLR, 2020, pp. 2907–2916. [Online]. Available: <http://proceedings.mlr.press/v108/janzing20a.html>.
- [150] K. Aas, M. Jullum, and A. Løland, “Explaining individual predictions when features are dependent: More accurate approximations to shapley values,” *Artificial Intelligence*, vol. 298, 2021. DOI: /10.1016/j.artint.2021.103502.
- [151] E. Zaunseder, U. Mütze, J. G. Okun, *et al.*, “Personalised metabolic whole-body models for newborns and infants predict growth and biomarkers of inherited metabolic diseases,” *bioRxiv*, 2023. DOI: /10.1101/2023.10.20.563364.
- [152] IBM, “V12. 1: User’s manual for cplex,” *International Business Machines Corporation*, vol. 46, no. 53, p. 157, 2009. [Online]. Available: <https://www.ibm.com/docs/en/icos/20.1.0?topic=cplex-users-manual>.
- [153] The MathWorks Inc., *Matlab version: 9.9.0.2037887 (r2020b) update 8*, Natick, Massachusetts, United States, 2020. [Online]. Available: <https://www.mathworks.com>.
- [154] G. B. Dantzig and M. N. Thapa, *Linear programming 1: introduction*. Berlin, Heidelberg: Springer-Verlag, 1997.
- [155] T. Magnanti and J. Orlin, *Chapter 2: Solving linear programs*, Accessed: 26-01-2024. [Online]. Available: <https://web.mit.edu/15.053/www/AMP-Chapter-02.pdf>.
- [156] H.-G. Holzhütter, “The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks,” *European journal of biochemistry*, vol. 271, no. 14, pp. 2905–2922, 2004. DOI: /10.1111/j.1432-1033.2004.04213.x.

- [157] K. Iizuka, “A novel approach to dynamic flux balance analysis that accounts for the dynamic transfer of information by internal metabolites,” *University of York, PhD thesis*, 2016. [Online]. Available: <https://etheses.whiterose.ac.uk/21661/1/correctedThesisIizuka.pdf>.
- [158] N. Lewis, K. Hixson, T. Conrad, *et al.*, “Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale mode,” *Molecular systems biology*, vol. 6, p. 390, 2010. DOI: /10.1038/msb.2010.47.
- [159] L. Heirendt, S. Arreckx, T. Pfau, *et al.*, “Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0,” *Nature protocols*, vol. 14, no. 3, pp. 639–702, 2019. DOI: /10.1038/s41596-018-0098-2.
- [160] A. M. Feist and B. O. Palsson, “The biomass objective function,” *Current opinion in microbiology*, vol. 13, no. 3, 2010. DOI: doi:10.1016/j.mib.2010.03.003.
- [161] H. Malhi, A. N. Irani, S. Gagandeep, and S. Gupta, “Isolation of human progenitor liver epithelial cells with extensive replication capacity and differentiation into mature hepatocytes,” *Journal of cell science*, vol. 115, no. 13, pp. 2679–2688, 2002. DOI: /10.1242/jcs.115.13.2679.
- [162] S. E. Senyo, M. L. Steinhauser, C. L. Pizzimenti, *et al.*, “Mammalian heart renewal by pre-existing cardiomyocytes,” *Nature*, vol. 493, no. 7432, pp. 433–436, 2013. DOI: /10.1038/2Fnature11682.
- [163] Y. Li and R. A. Wingert, “Regenerative medicine for the kidney: Stem cell prospects & challenges,” *Clinical and translational medicine*, vol. 2, no. 1, pp. 1–16, 2013. DOI: /10.1186/2001-1326-2-11.
- [164] I. Thiele and B. O. Palsson, “A protocol for generating a high-quality genome-scale metabolic reconstruction,” *Nat Protoc*, vol. 5, pp. 93–121, 2010. DOI: /10.1038/nprot.2009.203.
- [165] S. Björkman, “Prediction of drug disposition in infants and children by means of physiologically based pharmacokinetic (PBPK) modelling: Theophylline and midazolam as model drugs,” *British Journal of Clinical Pharmacology*, vol. 59, no. 6, pp. 691–704, 2005. DOI: /10.1111/j.1365-2125.2004.02225.x.
- [166] Nationwide children’s, *Physical exam of the newborn*, Accessed: 2022-10-21, 2022. [Online]. Available: <https://www.nationwidechildrens.org/conditions/health-library/physical-exam-of-the-newborn>.

- [167] F. J. Walther, B. Siassi, and P. Y. K. Wu, "Echocardiographic measurement of left ventricular stroke volume in newborn infants: A correlative study with pulsed doppler and m-mode echocardiography," *Journal of Clinical Ultrasound*, vol. 14, no. 1, pp. 37–41, 1986. DOI: /10.1002/jcu.1870140108.
- [168] J. King and D. Lowery, "Physiology, cardiac output," 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:81480255>.
- [169] I. Saxena, V. Shivankur, and M. Kumar, "Urinary protein creatinine ratio in normal zero to three-day- old indian neonates," *J Clin of Diagn Res*, vol. 10, no. 03, 2016. DOI: /10.7860/JCDR/2016/18651.7442.
- [170] The Royal Children's Hospital Melbourne, *Neonatal intravenous fluid management*, Accessed: 2022-10-17, 2022. [Online]. Available: https://www.rch.org.au/rchcpg/hospital_clinical_guideline_index/Neonatal_Intravenous_Fluid_Management/.
- [171] M. Sulemanji and K. Vakili, "Neonatal renal physiology," *Seminars in Pediatric Surgery*, vol. 22, no. 4, pp. 195–198, 2013, Neonatal Physiology. DOI: /10.1053/j.sempedsurg.2013.10.008.
- [172] T. Yasuda, T. Tomita, D. G. McLone, and M. Donovan, "Measurement of cerebrospinal fluid output through external ventricular drainage in one hundred infants and children: Correlation with cerebrospinal fluid production.," *Pediatr Neurosurg*, vol. 36, pp. 22–28, 2002. DOI: /10.1159/000048344.
- [173] B. M. Popkin, K. E. D'Anci, and I. H. Rosenberg, "Water, hydration, and health," *Nutrition Reviews*, vol. 68, no. 8, pp. 439–458, 2010. DOI: /10.1111/j.1753-4887.2010.00304.x.
- [174] N. F. Butte, J. K. Moon, W. W. Wong, J. M. Hopkinson, and E. Smith, "Energy requirements from infancy to adulthood," *The American Journal of Clinical Nutrition*, vol. 62, no. 5, 1047S–1052S, 1995. DOI: /10.1093/ajcn/62.5.1047S.
- [175] World Health Organization, *Exclusive breastfeeding for six months best for babies everywhere*, Accessed: 2023-08-15, 2023. [Online]. Available: <https://www.who.int/news/item/15-01-2011-exclusive-breastfeeding-for-six-months-best-for-babies-everywhere>.
- [176] S. Bakshi, V. K. Paswan, S. P. Yadav, *et al.*, "A comprehensive review on infant formula: Nutritional and functional constituents, recent trends in processing and its impact on infants' gut microbiota," *Frontiers in Nutrition*, vol. 10, 2023. DOI: /10.3389/fnut.2023.1194679.

- [177] M. Pajewska-Szmyt, E. Sinkiewicz-Darol, and R. Gadzała-Kopciuch, “The impact of environmental pollution on the quality of mother’s milk,” *Environmental Science and Pollution Research*, vol. 26, pp. 1–23, 2019. DOI: /10.1007/s11356-019-04141-1.
- [178] S. Thomas, J. Gauglitz, A. Tripathi, *et al.*, “An untargeted metabolomics analysis of exogenous chemicals in human milk and transfer to the infant,” *Clinical and Translational Science*, vol. 15, 2022. DOI: /10.1111/cts.13393.
- [179] O. Ballard and A. L. Morrow, “Human milk composition: Nutrients and bioactive factors,” *Pediatric clinics of North America*, vol. 60, no. 1, pp. 49–74, 2013. DOI: /10.1016/j.pcl.2012.10.002.
- [180] A. Noronha, J. Modamio Chamarro, Y. Jarosz, *et al.*, “The virtual metabolic human database: Integrating human and gut microbiome metabolism with nutrition and disease,” *Nucleic acids research*, vol. 47, 2018. DOI: /10.1093/nar/gky992.
- [181] E. Romero-Velarde, D. Delgado-Franco, M. García-Gutiérrez, *et al.*, “The importance of lactose in the human diet: Outcomes of a mexican consensus meeting,” *Nutrients*, vol. 11, no. 11, p. 2737, 2019. DOI: /10.3390/nu11112737.
- [182] K. S. Yeong and Y. D. Yong, “Components of human breast milk: From macronutrient to microbiome and microrna,” *Clin Exp Pediatr*, vol. 63, no. 8, pp. 301–309, 2020. DOI: /10.3345/cep.2020.00059.
- [183] K. Overgaard Poulsen, F. Meng, E. Lanfranchi, *et al.*, “Dynamic changes in the human milk metabolome over 25 weeks of lactation,” *Frontiers in Nutrition*, vol. 9, 2022. DOI: /10.3389/fnut.2022.917659.
- [184] D. B. Jelliffe and E. F. P. Jelliffe, “The volume and composition of human milk in poorly nourished communities. a review,” *The American journal of clinical nutrition*, vol. 31, no. 3, pp. 492–515, 1978. [Online]. Available: <https://api.semanticscholar.org/CorpusID:4505837>.
- [185] P. D. Hill, J. C. Aldag, R. T. Chatterton, and M. Zinaman, “Comparison of milk output between mothers of preterm and term infants: The first 6 weeks after birth,” *Journal of Human Lactation*, vol. 21, no. 1, 2005. DOI: /10.1177/0890334404272407.
- [186] Medela, *How much breast milk does a baby need?* Accessed: 2023-06-29, 2023. [Online]. Available: <https://www.medela.us/breastfeeding/articles/how-much-breast-milk-does-a-baby-need>.

- [187] E. K. Eckart, J. D. Peck, E. O. Kharbanda, E. M. Nagel, D. A. Fields, and E. W. Demerath, "Infant sex differences in human milk intake and composition from 1- to 3-month post-delivery in a healthy united states cohort," *Annals of Human Biology*, vol. 48, no. 6, pp. 455–465, 2021. DOI: /10.1080/03014460.2021.1998620.
- [188] C. W. Kuzawa, H. T. Chugani, L. I. Grossman, *et al.*, "Metabolic costs and evolutionary implications of human brain development," *Proceedings of the National Academy of Sciences*, vol. 111, no. 36, pp. 13 010–13 015, 2014. DOI: /10.1073/pnas.1323099111.
- [189] M. Gurven and R. Walker, "Energetic demand of multiple dependents and the evolution of slow human growth," *Proceedings of the Royal Society B: Biological Sciences*, vol. 273, no. 1588, 2006. DOI: /10.1098/rspb.2005.3380.
- [190] J. M. Berg, A. Held, J. L. Tymoczko, *et al.*, *Stryer Biochemie*. Springer Berlin Heidelberg, 2012. [Online]. Available: <https://books.google.de/books?id=PWxJLgEACAAJ>.
- [191] J. M. Hawdon and P. J. Thureen, "5 - thermal regulation and effects on nutrient substrate metabolism," in *Neonatal Nutrition and Metabolism*, 2nd ed. Cambridge University Press, 2006, pp. 58–66. DOI: /10.1017/CB09780511544712.006.
- [192] M. Aylott, "The neonatal energy triangle. Part 1: Metabolic adaptation.," *Paediatric nursing*, vol. 18, no. 6, 2006. DOI: /10.7748/paed.18.6.38.s30.
- [193] S. Hendley, *Newborns don't shiver from the cold, so it could be something sinister*, Accessed: 2023-10-05, 2023. [Online]. Available: <https://www.kidspot.com.au/baby/newborn/newborn-care/newborns-dont-shiver-from-the-cold-so-it-could-be-something-sinister/news-story/f0f132defd757271d4ff5cd42600b4c7>.
- [194] H. Asakura, "Fetal and neonatal thermoregulation," *Journal of Nippon Medical School*, vol. 71, pp. 360–70, 2005. DOI: /10.1272/jnms.71.360.
- [195] M. S. Popson, M. Dimri, and J. Borger, *Biochemistry, Heat and Calories*. StatPearls Publishing LLC, 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK538294/>.
- [196] I. Luginbuehl, B. Bissonnette, and P. J. Davis, "Chapter 6 - thermoregulation: Physiology and perioperative disturbances," in *Smith's Anesthesia for Infants and Children (Eighth Edition)*, 8th, Mosby, 2011, pp. 157–178. DOI: /10.1016/B978-0-323-06612-9.00006-7.
- [197] C. A. Gleason and S. E. Juul, *Avery's Diseases of the Newborn, Tenth Edition*. Elsevier, 2017. DOI: /10.1016/C2013-0-00320-9.

- [198] M. I. Rowe, G. Weinberg, and W. Andrews, "Reduction of neonatal heat loss by an insulated head cover," *Journal of Pediatric Surgery*, vol. 18, no. 6, pp. 909–913, 1983. DOI: /10.1016/S0022-3468(83)80045-1.
- [199] B. K. Siesjö, *Brain energy metabolism*. John Wiley & Sons, 1978. DOI: /10.1002/ana.410050318.
- [200] H. Wang, B. Wang, K. P. Normoyle, *et al.*, "Brain temperature and its fundamental properties: A review for clinical neuroscientists," *Frontiers in Neuroscience*, vol. 8, 2014. DOI: /10.3389/fnins.2014.00307.
- [201] A. Pierro, V. Carnielli, R. M. Filler, L. Kicak, J. Smith, and T. F. Heim, "Partition of energy metabolism in the surgical newborn," *Journal of pediatric surgery*, vol. 26, no. 5, pp. 581–586, 1991. DOI: /10.1016/0022-3468(91)90712-3.
- [202] N. F. Butte, W. W. Wong, J. M. Hopkinson, C. J. Heinz, N. R. Mehta, and E. Smith, "Energy requirements derived from total energy expenditure and energy deposition during the first 2 y of life," *The American journal of clinical nutrition*, vol. 72, no. 6, pp. 1558–1569, 2000. DOI: /10.1093/ajcn/72.6.1558.
- [203] R. Rising, D. Duro, M. Cedillo, S. Valois, and F. Lifshitz, "Daily metabolic rate in healthy infants," *The Journal of pediatrics*, vol. 143, no. 2, pp. 180–185, 2003. DOI: /10.1067/s0022-3476(03)00362-7.
- [204] J. C. Wells and P. S. Davies, "Estimation of the energy cost of physical activity in infancy," *Archives of disease in childhood*, vol. 78, no. 2, pp. 131–136, 1998. DOI: /10.1136/adc.78.2.131.
- [205] Children's Hospital Los Angeles, *Chapter 1: Nutritional need of infants*, Accessed: 2024-01-12, 2024. [Online]. Available: https://www.chla.org/sites/default/files/migratedChapter1_NutritionalNeeds.pdf.
- [206] J. K. Patel and A. S. Rouster, *Infant Nutrition Requirements and Options*. StatPearls Publishing, 2023. DOI: <https://www.ncbi.nlm.nih.gov/books/NBK560758/>.
- [207] Y. Yamada, X. Zhang, M. E. T. Henderson, *et al.*, "Variation in human water turnover associated with environmental and lifestyle factors," *Science*, vol. 378, no. 6622, pp. 909–915, 2022. DOI: /10.1126/science.abm8668.
- [208] B. Bakker, "A negative iodine balance is found in healthy neonates compared with neonates with thyroid agenesis," *Journal of Endocrinology*, vol. 161, pp. 115–120, 1999. DOI: /10.1677/joe.0.1610115.

- [209] D. DiTomasso and M. Cloud, "Systematic review of expected weight changes after birth for full-term, breastfed newborns," *Journal of Obstetric, Gynecologic and Neonatal Nursing*, vol. 48, no. 6, pp. 593–603, 2019. DOI: /10.1016/j.jogn.2019.09.004.
- [210] A. Sarmiento-Aguilar, D. Horta-Carpinteyro, and A. Prian-Gaudiano, "Percentage of birth weight loss as a reference for the well-being of the exclusively breastfed newborn.," *Boletín medico del Hospital Infantil de Mexico*, vol. 79, no. 6, pp. 341–349, 2022. DOI: /10.24875/BMHIM.22000032.
- [211] Y. Miyoshi, H. Suenaga, M. Aoki, and S. Tanaka, "Determinants of excessive weight loss in breastfed full-term newborns at a baby-friendly hospital: A retrospective cohort study," *International Breastfeeding Journal*, vol. 15, 2020. DOI: /10.1186/s13006-020-00263-2.
- [212] B. R. Weston and I. Thiele, "A nutrition algorithm to optimize feed and medium composition using genome-scale metabolic models," *Metabolic Engineering*, vol. 76, pp. 167–178, 2023. DOI: /10.1016/j.ymben.2023.01.010.
- [213] E. S. Köhler, S. Sankaranarayanan, C. J. van Ginneken, *et al.*, "The human neonatal small intestine has the potential for arginine synthesis; developmental changes in the expression of arginine-synthesizing and -catabolizing enzymes," *BMC Dev Biol*, vol. 8, no. 107, 2008. DOI: /10.1186/1471-213X-8-107.
- [214] S. M. Morris, "Arginine metabolism: Boundaries of our knowledge," *The Journal of Nutrition*, vol. 137, no. 6, 1602S–1609S, 2007. DOI: /10.1093/jn/137.6.1602S.
- [215] F. Molema, F. Gleich, P. Burgard, *et al.*, "Decreased plasma l-arginine levels in organic acidurias (mma and pa) and decreased plasma branched-chain amino acid levels in urea cycle disorders as a potential cause of growth retardation: Options for treatment," *Molecular Genetics and Metabolism*, vol. 126, no. 4, pp. 397–405, 2019. DOI: /10.1016/j.ymgme.2019.02.003.
- [216] A. J. van Vught, P. C. Dagnelie, I. C. Arts, *et al.*, "Dietary arginine and linear growth: The copenhagen school child intervention study," *The British journal of nutrition*, vol. 109, no. 6, pp. 1031–1039, 2013. DOI: /10.1017/S0007114512002942.
- [217] R. Williams, C. Mamotte, and J. Burnett, "Phenylketonuria: An inborn error of phenylalanine metabolism," *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*, vol. 29, pp. 31–41, 2008. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2423317/>.

- [218] C. R. Ferreira, C. D. van Karnebeek, J. Vockley, and N. Blau, "A proposed nosology of inborn errors of metabolism," *Genetics in medicine*, vol. 21, no. 1, pp. 102–106, 2019. DOI: /10.1038/s41436-018-0022-8.
- [219] E. MacLeod and D. Ney, "Nutritional management of phenylketonuria," *Annales Nestle [English ed.]*, vol. 68, pp. 58–69, 2010. DOI: /10.1159/000312813.
- [220] A. MacDonald, A. van Wegberg, K. Ahring, *et al.*, "Pku dietary handbook to accompany pku guidelines," *ORPHANET J RARE DIS*, vol. 15, no. 1, p. 171, 2020. DOI: 10.1186/s13023-020-01391-y.
- [221] L. Kalsner, F. Rohr, K. Strauss, M. Korson, and H. Levy, "Tyrosine supplementation in phenylketonuria: Diurnal blood tyrosine levels and presumptive brain influx of tyrosine and other large neutral amino acids," *The Journal of pediatrics*, vol. 139, no. 3, pp. 421–427, 2001. DOI: /10.1067/mpd.2001.117576.
- [222] C.-Y. Chiu, K.-W. Yeh, G. Lin, *et al.*, "Metabolomics reveals dynamic metabolic changes associated with age in early childhood," *PloS one*, vol. 11, 2016. DOI: /10.1371/journal.pone.0149823.
- [223] A. Basile, A. Heinken, J. Hertel, *et al.*, "Longitudinal flux balance analyses of a patient with episodic colonic inflammation reveals microbiome metabolic dynamics," *Gut microbes*, vol. 15, 2023. DOI: /10.1080/19490976.2023.2226921.
- [224] S. M. Innis, "Impact of maternal diet on human milk composition and neurological development of infants," *The American journal of clinical nutrition*, vol. 99, no. 3, 734S–741S, 2014. DOI: /10.3945/ajcn.113.072595.
- [225] A. Thakur, M. M. Parvez, J. S. Leeder, and B. Prasad, "Enzyme kinetics in drug metabolism: Fundamentals and applications," in *Ontogeny of Drug-Metabolizing Enzymes*. New York, NY: Springer US, 2021, pp. 551–593. DOI: /10.1007/978-1-0716-1554-6_18.
- [226] F. Y. Bois, "Physiologically based modelling and prediction of drug interactions," *Basic & clinical pharmacology & toxicology*, vol. 106, no. 3, pp. 154–161, 2010. DOI: /10.1111/j.1742-7843.2009.00488.x.
- [227] X. Zhuang and C. Lu, "PBPK modeling and simulation in drug research and development," *Acta Pharm Sin B*, vol. 6, no. 5, pp. 430–440, 2016. DOI: /10.1016/j.apsb.2016.04.004.
- [228] H. Jones and K. Rowland-Yeo, "Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development," *CPT: pharmacometrics & systems pharmacology*, vol. 2, no. 8, 2013. DOI: /10.1038/psp.2013.41.

- [229] J. Gabrielsson and D. Weiner, *Pharmacokinetic and pharmacodynamic data analysis: concepts and applications*. CRC press, 2001.
- [230] P. Poulin and F.-P. Theil, “A priori prediction of tissue:plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery,” *Journal of Pharmaceutical Sciences*, vol. 89, no. 1, pp. 16–35, 2000. DOI: /10.1002/(SICI)1520-6017(200001)89:1<16::AID-JPS3>3.0.CO;2-E.
- [231] M. K. Toroghi, W. R. Cluett, and R. Mahadevan, “A personalized multiscale modeling framework for dose selection in precision medicine,” *Industrial & Engineering Chemistry Research*, vol. 59, no. 21, pp. 9819–9829, 2020. DOI: /10.1021/acs.iecr.0c01070.
- [232] REVIVE, *Plasma protein binding*, Accessed: 2023-11-20. [Online]. Available: <https://revive.gardp.org/resource/plasma-protein-binding/?cf=encyclopaedia>.
- [233] R. Mc Menamy, C. Lund, and J. Oncley, “Unbound amino acid concentrations in human blood plasmas,” *J Clin Invest*, vol. 36, no. 12, 1957. DOI: doi:10.1172/JCI103568.
- [234] L. F. Shampine and M. W. Reichelt, “The matlab ode suite,” *SIAM Journal on Scientific Computing*, vol. 18, pp. 1–22, 1997. DOI: /10.1137/S1064827594276424.
- [235] T. Sullivan, *Introduction to uncertainty quantification*. 2015. [Online]. Available: <http://dx.doi.org/10.1007/978-3-319-23395-6>.
- [236] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, “Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index,” *Comput. Phys. Commun.*, vol. 181, pp. 259–270, 2010. [Online]. Available: <https://api.semanticscholar.org/CorpusID:206995386>.
- [237] M. M. Rajput and J. Kusch, *Uncertainty quantification dictionary monte carlo method*, Accessed: 29-01-2024. [Online]. Available: https://dictionary.helmholtz-ug.de/content/monte%5C_carlo%5C_method.html.
- [238] M. Ben Guebila and I. Thiele, “Model-based dietary optimization for late-stage, levodopa-treated, parkinson’s disease patients,” *npj Systems Biology and Applications*, vol. 2, p. 16 013, 2016. DOI: /10.1038/npjjsba.2016.13.
- [239] IT’IS Foundation, *Density*, Accessed: 2023-10-16. [Online]. Available: <https://itis.swiss/virtual-population/tissue-properties/database/density/>.
- [240] R. Usher, M. Shepard, and J. Lind, “The blood volume of the newborn infant and placental transfusion,” *Acta Paediatrica*, vol. 52, no. 5, pp. 497–512, 1963. DOI: /10.1111/j.1651-2227.1963.tb03809.x.

- [241] C. H. Cramer and P. D. Brophy, "Chapter 201 - technical aspects of pediatric continuous renal replacement therapy," in *Critical Care Nephrology (Third Edition)*, C. Ronco, R. Bellomo, J. A. Kellum, and Z. Ricci, Eds., Third Edition, Philadelphia: Elsevier, 2019, pp. 1210–1214. DOI: /10.1016/B978-0-323-44942-7.00201-6.
- [242] R. Mosteller, "Simplified calculation of body-surface area," *New England Journal of Medicine*, vol. 317, no. 17, pp. 1098–1098, 1987. DOI: /10.1056/NEJM198710223171717.
- [243] MedicalNewsToday, *What is the average length for a baby?* Accessed: 2024-02-06. [Online]. Available: <https://www.medicalnewstoday.com/articles/324728#average-lengths>.
- [244] M. Peng, L. Liu, M. Jiang, *et al.*, "Measurement of free carnitine and acylcarnitines in plasma by HILIC-ESI-MS/MS without derivatization," *Journal of Chromatography B*, vol. 932, pp. 12–18, 2013. DOI: /10.1016/j.jchromb.2013.05.028.
- [245] K. Tada, H. Hirono, and T. Arakawa, "Endogenous renal clearance rates of free amino acids in prolinuric and hartnup patients," *Tohoku Journal of Experimental Medicine*, vol. 93, no. 1, pp. 57–61, 1967. DOI: /10.1620/tjem.93.57.
- [246] C. J. Rebouche, "Kinetics, pharmacokinetics, and regulation of l-carnitine and acetyl-l-carnitine metabolism," *Annals of the New York Academy of Sciences*, vol. 1033, no. 1, pp. 30–41, 2004. DOI: /10.1196/annals.1320.003.
- [247] S. Reuter and A. Evans, "Carnitine and acylcarnitines," *Clin Pharmacokinet*, vol. 51, pp. 553–572, 2012. DOI: /10.1007/BF03261931.
- [248] R. Ensenauer, R. Fingerhut, E. M. Maier, *et al.*, "Newborn Screening for Isovaleric Acidemia Using Tandem Mass Spectrometry: Data from 1.6 Million Newborns," *Clinical Chemistry*, vol. 57, no. 4, pp. 623–626, 2011. DOI: /10.1373/clinchem.2010.151134.
- [249] I. Mahmood, "A comparison of different methods for the first-in-pediatric dose selection," *Journal of Clinical and Translational Research*, vol. 8, no. 5, p. 369, 2022.
- [250] T.-M. Hu, "A general biphasic bodyweight model for scaling basal metabolic rate, glomerular filtration rate, and drug clearance from birth to adulthood," *The AAPS Journal*, vol. 24, no. 3, p. 67, 2022. DOI: /10.1208/s12248-022-00716-y.
- [251] S. L. Brunton and J. N. Kutz, "Singular value decomposition (svd)," in *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019, pp. 3–46. DOI: /10.1017/9781108380690.002.
- [252] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Third Edition. The Johns Hopkins University Press, 1996.

- [253] G. Strang, *Linear Algebra and Learning from Data*. Wellesley-Cambridge Press, 2019. [Online]. Available: https://books.google.de/books?id=L0Y%5C_wQEACAAJ.
- [254] R. M. Larsen, “Lanczos bidiagonalization with partial reorthogonalization,” *DAIMI Report Series*, vol. 27, no. 537, 1998. DOI: /10.7146/dpb.v27i537.7070.
- [255] The MathWorks Inc., *Matlab version: 23.2.0.2485118 (r2023b)*, Natick, Massachusetts, United States, 2023. [Online]. Available: <https://www.mathworks.com>.
- [256] The MathWorks Inc., *Optimization toolbox version: R2023b*, Natick, Massachusetts, United States, 2023. [Online]. Available: <https://www.mathworks.com>.
- [257] C. Lemieux, *Monte Carlo and Quasi-Monte Carlo Sampling*. New York, NY, USA: Springer, 2009.
- [258] T. Mathews, S. S. Hayer, D. Dinkel, *et al.*, “Maternal-child microbiome and impact on growth and neurodevelopment in infants and children: A scoping review,” *Biological Research For Nursing*, vol. 25, no. 3, pp. 454–468, 2023. DOI: /10.1177/10998004221151179.
- [259] D. Rios-Covián, P. Langella, and R. Martin, “From short- to long-term effects of c-section delivery on microbiome establishment and host health,” *Microorganisms*, vol. 9, p. 2122, 2021. DOI: /10.3390/microorganisms9102122.
- [260] J. Bonham, R. Carling, M. Lindner, *et al.*, “Raising awareness of false positive newborn screening results arising from pivalate-containing creams and antibiotics in europe when screening for isovaleric acidaemia,” *International Journal of Neonatal Screening*, vol. 4, p. 8, 2018. DOI: /10.3390/ijns4010008.

