Aus der Klinik für Geburtsmedizin
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Anwendung von Machine-Learning in medizinischer Diagnostik in der Geburtshilfe
Applications of machine-learning in medical diagnostics in obstetrics

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Leon J. Schmidt

Datum der Promotion: 30.6.2024

Table of contents

List of tables

List of figures

List of abbreviations

AI	Artificial Intelligence
EHR	Electronic Health Record
GBTree	Gradient-boosted trees
HELLP - syndrome	Haemolysis, Elevated Liver Enzymes, Low Platelet - syndrome
IUGR	Intrauterine growth restriction
IVH	Intraventricular haemorrhage
MAE	Mean absolute error
ML	Machine-Learning
NEC	Necrotizing enterocolitis
NPV	Negative predictive value
PPV	Positive predictive value
RDS	Respiratory distress syndrome
RF	Random forest
SD	Standard deviation
SGA	Small for gestational age
SVM	Support vector machine

Abstract

Background

Improvements in computational capacity and new algorithmic approaches to data analysis have created enormous opportunities to improve conventional diagnostics in the hospital in recent years. Especially obstetrics, a speciality with high-dimensional data and limited performances in their conventional diagnostic approaches for many adverse outcomes in pregnancy, stands to benefit greatly from the application of machine-learning. This dissertation intends to present our own work which predicts the occurrence of adverse outcomes in preeclampsia high-risk-pregnancies and to contextualise it with the current state of research for the application of machine-learning in preeclampsia as well as other obstetric/gynecologic conditions in general.

Methods

The presented study is based on a patient collective of 1647 women which presented to the obstetric department of the Charité Universitätsmedizin Berlin between July 2010 and March 2019. We determined predictive performance of different machine-learning algorithms (Gradient boosted trees, Random Forest) for adverse outcomes commonly associated with preeclampsia and compared them to models based on laboratory and vital parameter cutoffs (blood pressure, sFlt-1/PlGF ratio and their combination with proteinuria measurements) used in the clinic. Dataset splitting was performed in a per-patient randomised fashion using a 90-10 split and evaluation was performed using a 10x10-fold cross-validation approach.

Results

Our own study showed gains in predictive performance when using machine-learning models. Accuracy for gradient boosted trees was 87 ± 3 % while blood pressure cutoffs achieved only 65 ± 4 % and a cutoff of 38 applied to the sFlt-1/PlGF-ratio yielded an accuracy of 68 ± 5 %. The positive predictive value especially improved from 33 ± 9 % for the blood-pressure-cutoffs to 82 ± 10 % for the gradient-boosted trees classifier with the “full clinical model” consisting of blood pressure, sFlt-1/PlGF ratio and proteinuria achieving 44 ± 9 % PPV. Overall we found that using machine-learning methods leads to great improvements in all assessed performance metrics with potential for further enhancement using optimization on the algorithms’ output probabilities’ cutoffs.

Conclusions

Machine-learning greatly improves the diagnostic capabilities for preeclampsia and, as shown by many other works in this dissertation, obstetrics/gynaecology and medicine in general. This could represent a starting point for further research which leads to more sophisticated diagnostic or decision-support tools.

Zusammenfassung

Einleitung

Verbesserungen in Rechenkapazitäten und neue algorithmische Ansätze der Datenanalyse haben große Möglichkeiten zur Verbesserung konventioneller Diagnostik in Krankenhäusern über die letzten Jahre kreiert. Besonders die Geburtshilfe, eine Fachrichtung mit hochdimensionalen Datensätzen und limitierter Performance der konventionellen diagnostischen Methoden für viele der adversen Events in der Schwangerschaft, kann stark von der Anwendung von Machine-Learning profitieren.

Diese Dissertation beabsichtigt unsere eigene Arbeit, welche das Auftreten adverser Events in Präeklampsie-Hochrisikoschwangerschaften vorhersagt, vorzustellen und mit dem aktuellen stand der Forschung für Machine-Learning in der Präeklampsie sowie Gynäkologie/Geburtshilfe in Kontext zu setzen.

Methoden

Die vorgestellte Studie basiert auf einer Patientinnengruppe von 1647 Frauen, die sich zwischen Juli 2010 und März 2019 in der Klinik für Geburtsmedizin der Charité Universitätsmedizin Berlin vorstellten.

Wir untersuchten die Leistung verschiedener Machine-Learning-Algorithmen (Gradient Boosted Trees, Random Forest) zur Vorhersage häufig mit Präeklampsie assoziierter adverser Events und verglichen diese mit Modellen basierend auf klinisch angewendeten Labor- und Vitalparameter-Grenzwerten (Blutdruck, sFlt-1/PIGF-Ratio und ihre Kombination mit Proteinurie-Messungen).

Der Datensatz wurde auf einer randomisierten Pro-Patient-Basis in einem 90-10-split in Trainings- und Testsatz geteilt und mittels einer 10x 10-fachen Kreuzvalidierung evaluiert.

Ergebnisse

Unsere Studie zeigte Zugewinne an prädiktiver Leistung durch Nutzung von Machine-Learning-Modellen. Genauigkeit für Gradient boosted trees war 87 ± 3 %, während Blutdruckgrenzwerte lediglich 65 ± 4 % erreichen konnten und ein Grenzwert von 38 der sFlt-1/PLGF-Ratio eine Genauigkeit von 68 ± 5 %. Insbesondere der positiv prädiktive Wert verbesserte sich von 33 ± 9 % für den Blutdruckgrenzwert auf 82 ± 10 % für den Gradient-boosted Trees-Klassifizierer, während das “vollständige” klinische Modell bestehend aus Blutdruck, sFlt-1/PlGF-Ratio und Proteinurie 44 ± 9 % erreichen konnte. Insgesamt fanden wir, dass Machine-Learning Methoden zu großen Verbesserungen in allen untersuchten Performance-Metriken führt, mit Potential zu weiteren Verbesserungen durch Optimierung von Grenzwerten auf den ausgegebenen Wahrscheinlichkeiten der Modelle.

Schlussfolgerung

Machine-Learning führt zu immensen Verbesserungen der diagnostischen Möglichkeiten für Präeklampsie und, wie durch viele weitere Arbeiten in dieser Dissertation gezeigt, Gynäkologie/Geburtshilfe und Medizin im Allgemeinen.

Dies kann einen Startpunkt für weitere Forschung repräsentieren, welche zu anspruchsvolleren Diagnostik- und Entscheidung-Support-Werkzeugen führt.

1. Introduction

1.1 Machine-Learning

Digitization can be described as one of the defining megatrends in the 21st century. Computers are part of every facet of modern existence, private life, industry, governments and healthcare. This resulted in large amounts of data being created and stored which in turn led to the task of analysing this data to obtain meaningful information from it. Before the advent of large-scale data analysis the human-computer-interaction was largely restricted to one-way-information transfer. Humans could enter and retrieve information in and from the information system but additional insight was generated solely on the human side of the interaction. To aid in this endeavour machine-learning methods and automated algorithms for data analysis were conceived.

The term machine-learning itself encompasses a wide variety of statistical methods which seek to infer patterns in data via mathematical operations. These automated processes span from relatively simple methods such as linear regression to highly complex designs such as artificial neural networks which provide the basis for modern deep learning strategies.

The first application for machine-learning was established in the 1950s using a formulated approach to the game of checkers(Samuel, 1959) with theoretical foundations for other methods such as perceptrons(Shaw, 1986) or Bayes' theorem(Joyce, 2021) even predating this achievement.

For a significant amount of time computational power and data storage hindered the practical application of the formulated algorithms. As both increased along an exponential curve("Moore's Law: The number of transistors per microprocessor," n.d.; Schaller, 1997) applications of these algorithms grew increasingly more realistic and implementations working on real-world problems became possible. In 1997 IBM's DeepBlue(Campbell et al., 2002) managed to defeat the reigning chess world champion Gary Kasparov in a game of chess in what can be argued to be the first major public event that demonstrated the capabilities of statistical computing in a field which hitherto was considered to be an exclusive domain of human creativity. Further improvements in hardware and algorithms(Ciresan et al., 2011; Le, 2013; Oh and Jung, 2004;

Schmidhuber, 2015; Taigman et al., 2014) throughout the 2010s enabled deep learning and saw the widespread application in many different contexts. The current status of machine-learning sees its usage throughout almost every facet of data processing, especially in data-driven markets such as financial analytics, social media, image recognition or statistical modelling. (Dixon et al., 2020; Mohammed et al., 2016)

1.2 Application in Medicine

Although machine-learning has advanced to a significant degree over the past decades, adoption in medicine has seen very little improvements.(“eHealth Trend Barometer,” 2019) Despite being one of the most data-driven sectors of human life, with dozens of measurements taken during an average hospital stay, evaluation of the results depends on the treating physician, which in turn are themselves increasingly more occupied by clerical tasks and documentation.(Hill et al., 2013; Tipping et al., 2010) These high demands and the little time for the actual medical decision-making are components of a clinical diagnostic error rate of about 5%.(Singh et al., 2014; Winters et al., 2012)

Adoption of decision-support in the hospital is still only present at a very rudimentary level.(“eHealth Trend Barometer,” 2019; “Electronic Medical Record Adoption Model | HIMSS Analytics - Europe,” n.d.) Clinicians are guided in their decisions largely by established clinical standards and standard operating procedures. Created to serve as guidelines for the largest possible number of patients these necessarily can’t apply to each individual to the same extent. This scenario constitutes a prime example for the application of machine-learning techniques which, given sufficient data in both quality and quantity, can more precisely obtain patterns present in that data and serve as a basis for a more individualised healthcare system by categorising patients along a multitude of axes.

1.3 Preeclampsia

Preeclampsia constitutes a major factor in the morbidity and mortality of pregnant women and infant children with an incidence of around 3-6%(Lisonkova and Joseph, 2013; Purde et al., 2015) and high potential for severe complications such as HELLP-syndrome, Eclampsia, placental abruption or fetal necrotizing enterocolitis, respiratory distress syndrome and others.

Detection of these outcomes has always proven very difficult. With a variety of unspecific symptoms and the potential for severe courses developing quickly and no definitive diagnostic tool the recommended treatment in Germany consists of hospitalisation and generally induced delivery at the latest possible time in the pregnancy.(German Society of Gynecology and Obstetrics, 2019) Though in recent years development of new biomarkers(Verlohren et al., 2014) and scoring systems("The Fetal Medicine Foundation Risk for preeclampsia calculator," n.d.; Wright et al., 2012) have led to increases in the performance of negative rule-out, the problem of rule-in meaning prediction of actual adverse outcomes from the risk condition of preeclampsia still remains low.(Verlohren et al., 2014) This in turn leads to great distress for the women who are often unnecessarily hospitalised and a large burden for the healthcare system in terms of resource consumption and financial burden.(Stevens et al., 2017) Correctly identifying women that will develop adverse outcomes therefore represents a major challenge in pregnant womens' healthcare.

1.4 ML Application in Preeclampsia

The described situation, high dimensionality in data (ultrasound, biomarkers, standard laboratory, patient history etc.) and no single data point sufficiently predicting the target variable, calls for new statistical methods to integrate the entire available range. Machine-learning techniques lend themselves supremely to this task. Provided enough data of sufficient quality, these algorithms are theoretically capable of picking up patterns in highly complex data and therefore identifying the patients at risk. Furthermore the potential to make use of large amounts of data stored in clinical information systems and direct integration with those could provide physicians with a veritable decision-support tool.

Though significant problems such as staff-adoption and privacy concerns should be considered, the overall potential for major improvements in patient safety, in our estimation, outweighs the challenges regarding the implementation.(Henry et al., 2022)

We therefore hypothesised that the application of machine-learning techniques in preeclampsia is capable of improving the identification of pregnant women who will sustain an adverse outcome in their pregnancy.

2. Methods

2.1 Dataset & target variable

The basis for our analysis was a dataset of 1647 women across a set of 2472 samples gathered by a dedicated team between July 2010 and March 2019 at the department of obstetrics, Charité University Medicine Berlin.
Exclusion criteria for participation in our study was age under 18 years, gestational age of less than 20 weeks or unavailable measurements for sFlt, PlGF and missing outcome reporting. Important inclusion criteria were symptoms or measurements that would suggest an imminent preeclampsia or were highly associated with it. First, abnormally high blood pressure measurement of above 140 mmHg systolic or 90 mmHg diastolic. Second, proteinuria as characterised by urine dipsticks recording a value above ‘++’ in tests at least 6 hrs apart or a 24h urine protein measurement of above 300 mg. Third, abnormal ultrasound findings such as intrauterine growth restriction (IUGR, <10th percentile), pathological values for uterine artery, umbilical artery or foetal medial cerebral artery. Fourth, presence of preeclampsia-related symptoms such as epigastric pain, headaches, visual disorders, increasing presence of edema or weight gain above a normal level. The fifth inclusion criteria was abnormal readings in specific laboratory values such as thrombocyte count or ALT/AST elevations as signs for liver damage.(Schmidt et al., 2022)

The patient data was organised into visits, each of which compromised the data gathered at one presentation to the clinic. In case of multiple measurements gathered at one visit we chose the measurement performed last. Association with adverse outcomes and thus labelling the data in a manner fit for supervised machine-learning applications was performed one month after birth since preeclampsia can occur after delivery.

A significant portion of the study population (1122 patients, 68.12%), representing an earlier version of the database, was the basis for another study conducted in our working group by Dröge et al. in 2021.(Dröge et al., 2021)

Features were all information accessible via our hospital’s clinical information system such as biographical information/medical history, measurements of vital signs, symptoms associated with preeclampsia or adverse outcomes, laboratory results and ultrasound findings, resulting in a total of 114 features. For a full list please refer to Schmidt et al.(Schmidt et al., 2022)

The target variable for our analysis was the occurrence of an adverse outcome in either the mother or the child at any point in the future after the presentation to our clinic.
Foetal adverse outcomes included IUGR, SGA, premature birth (<=34th weeks of gestation) due to preeclampsia, respiratory distress syndrome (RDS), necrotizing enterocolitis (NEC), intraventricular haemorrhage, placental abruption or death.
Combined with these were the maternal adverse outcomes of disseminated intravascular coagulopathy (DIC), pleural effusion or lung edema, cerebral hematoma, renal failure, HELLP syndrome, eclampsia and death via a logical OR.

Data gathering relied on manual data entry using our hospital's clinical information system (SAP Hana, SAP) and ultrasound records program (Viewpoint, GE Healthcare) as data sources.(Schmidt et al., 2022) If a patient's records necessitated it, we relied on written records.

2.2 Preprocessing

We did not perform interpolation on missing data by inserting newly generated values. All missing data was signified with a special indicator variable.
The exception to this was the treatment of biographical information or data pertaining to a patient’s medical history. If this data was missing from subsequent entries but appeared in a prior visit, we carried it forward to all following entries of that particular pregnancy. This did not apply to features which could be considered highly variable between visits such as laboratory values, ultrasound data and vital parameters.(Schmidt et al., 2022)
Categorical features such as ethnicity were replaced by indicator variables for each of the features options. We decided not to choose one option as a so-called baseline in order to have all information directly accessible in case of subsequent analysis.

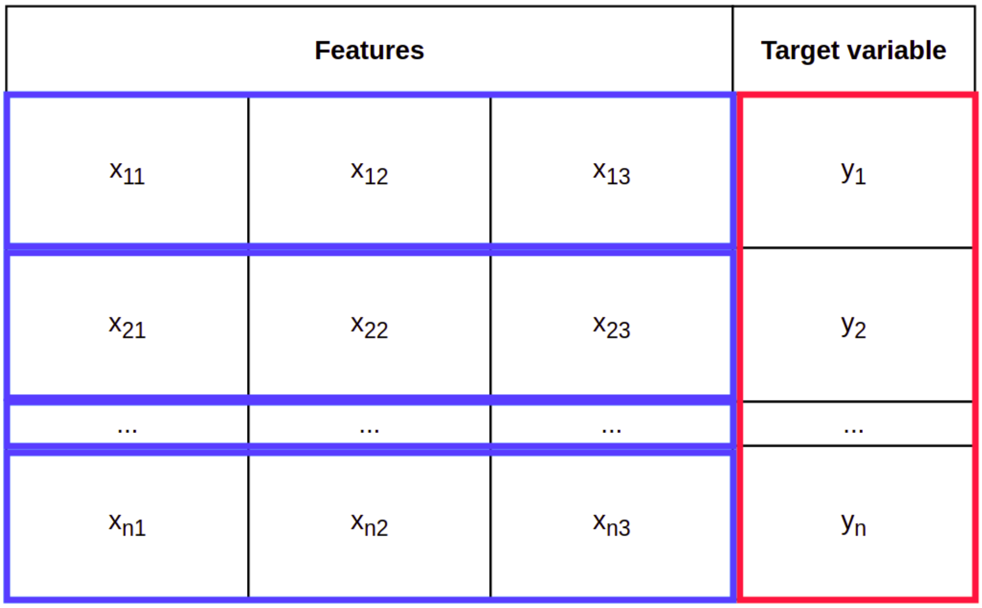
For highly specific features (sFlt, PlGF, ultrasound findings) we added additional information by placing them along known distributions in the population at large.(Ciobanu et al., 2019; Verlohren et al., 2014)

2.3 Algorithm explanation

Machine-learning is the basic application of algorithms to infer the relationship between a set of n observations or feature vectors $X = \{x_1, x_2, ..., x_n\}$ and an associated response $Y = \{y_1, y_2, ..., y_n\}$ or dependent variable. Each observation represents a set of values in a number of categories. For a number of features m the vector $x_1 = \{x_{11}, x_{12}, ..., x_{1m}\}$ represents a single observation with a defined value for each feature $x_{11}, ..., x_{1m}$.(Hastie et al., 2009)

Figure 1:

Example of a supervised-learning dataset.



Features			Target variable
x_{11}	x_{12}	x_{13}	y_1
x_{21}	x_{22}	x_{23}	y_2
...	...	...	...
x_{n1}	x_{n2}	x_{n3}	y_n

Each feature vector is associated with a target variable.
Own representation: LJ Schmidt

In order to find a suitable algorithm for a specific research question the methods can be categorised along two axes - training methodology and prediction target which will influence the choice of algorithms to use.(Burkov, 2019)

Training methodology is itself divided into four categories.

First supervised learning in where each observation $x_n = \{x_{n1}, x_{n2},..., x_{nm}\}$, is associated with a response variable y_n.

Second is unsupervised learning, where no response variable is present and solely the set of observations is analysed for structural patterns in the data.

Third is semi-supervised learning, a mixture between the first two categories where only some of the observations given in the dataset are mapped to a corresponding value for the dependent variable y.

Figure 2:

Example of a semi-supervised learning dataset

Features			Target variable
x_{11}	x_{12}	x_{13}	
x_{21}	x_{22}	x_{23}	y_2
x_{31}	x_{32}	x_{33}	
x_{41}	x_{42}	x_{43}	y_4

Own representation: LJ Schmidt

The fourth methodological approach is known as reinforcement learning, which iteratively operates in a trial-and-error approach in order to maximise a given goal metric.

The nature of our dataset, a patient collective represented by high-dimensional feature-vectors mapped directly onto the target metric of sustaining an adverse outcome, lends itself excellently to a supervised-learning approach.

For the specific algorithm selection we have to consider the second axis, the nature of the prediction target variable. Supervised learning can be broadly summed in two specific sets of problems, categorization, where each observation is mapped to a categorical variable, and regression, where the target variable represents a continuous variable and the desired prediction is a value on that particular continuous scale. Our target, prediction of an adverse outcome at any point in the woman’s future, represents a binary target variable, which classifies our problem as categorical.

Due to inherent characteristics of our dataset, especially the prevalence of missing values, the need for a more interpretable approach in machine-learning in healthcare-related fields and their generally high-performance in high-dimensional classification problems we ultimately decided to implement decision-tree-based approaches.

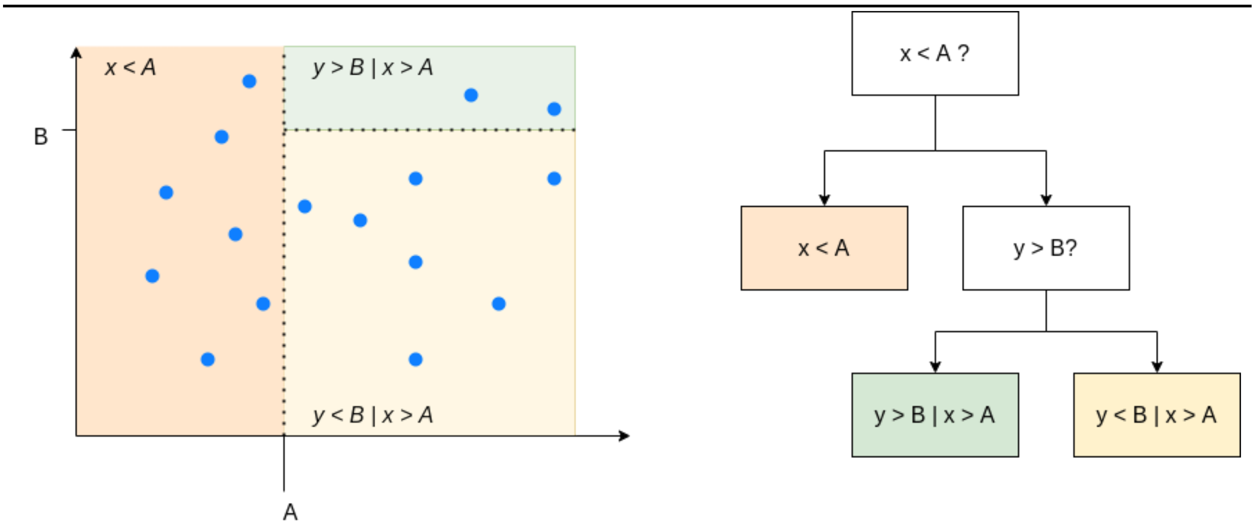
Considering an especially high risk for overfitting the dataset, we furthermore chose to enhance the decision trees by introducing the concept of boosting also known as the gradient-boosted trees algorithm (GBTree)(Breiman, 1996; Chen and Guestrin, 2016; Friedman, 2002, 2001; Mason et al., 1999) and bagging(Breiman, 1996), known as a random forest classifier (RF)(Breiman, 2001; Cutler et al., 2012; Tin Kam Ho, 1995).

2.3.1 Decision trees

A decision tree is a simple model that iteratively analyses the given dataset and splits it into distinct, non-overlapping regions similar to a Bayesian approach (see Figure 3 for an illustration). At the beginning the algorithm surveys all features present in the dataset and applies a number of cutoffs on them. It then selects the feature-cutoff combination which supplies the highest purity in the resulting subsets as expressed by, in our case, the subset’s entropy. Entropy is given by $E = - \sum_{b=1}^{B} p_{bm} \, log(p_{bm})$, with p_{bm}being the probability of an element in subset b belonging to group m, with B being the total number of subsets and m the total number of features. It is evident that this value decreases as node purity increases.(Hastie et al., 2009; James et al., 2021)

Figure 3:

Illustration of dataset-splitting by a decision tree algorithm



The dataset is being split by the decision tree at two cutoffs along a two-dimensional space (A for the x-value and B for the y-value). The split for x > A is split in two distinct regions (y > B and y < B), which represents the decision tree's second level. The tree's leaf nodes represent the dataset's different regions as illustrated by their colour.

Based on James, G., Witten, D., Hastie, T. & Tibshirani, R. Tree-Based Methods. in An Introduction to Statistical Learning 327–365 (Springer US, 2021), own Illustration

This procedure is then iteratively repeated on the resulting subdivisions in the dataset until a predefined number of nodes or subsets is achieved.

Both algorithms used by our group build on this simple principle by constructing a series of trees with differing target metrics and weights in case of GBTree or an ensemble of trees by randomly choosing features to build the trees from.

2.3.2 Random forests

Random forests use two strategies to improve upon the basic concept of decision trees since small shifts in the data can result in relatively large changes to the decision tree. The term for this - changes in the dataset predicting large changes in the algorithm - is

called variance. To decrease variance random forests use bagging, that means creating multiple decision trees from random subsamples of the dataset and random feature selection, that means at each split only a subset of features is allowed to be analysed for entropy reduction. Bagging reduces variance by “averaging” the features’ distributions throughout random subsets while the random feature sampling provides a necessary decorrelation of features. That means that in case of one or more very strong predictors present, these would not appear in all trees of the random forest thus making it more robust to changes in these strong predictors in other datasets.
Predictions for classification are made by majority vote of all created decision trees. (Hastie et al., 2009; James et al., 2021)

2.3.3 Gradient boosted trees

Gradient boosting also relies on building a number of decision trees but approaches it in an iterative fashion. At each step i in the algorithm it fits the dataset and prediction (X, y_i) to a shallow decision tree, usually of depth 3 - 8 and then adds this tree to previously created trees with a penalty parameter or learning rate α via the formula $f_{i+1} = \sum_{n=1}^{i-1} \alpha f_n + \alpha f_i$.(James et al., 2021) The learning rate governs the gradual increase in ability for classification thus allowing each new tree to only contribute a small part to the overall decision. The label for the next tree is then calculated by comparing the prediction $\widehat{y}_i$ to the response of the i -th step $y_{i+1} = y_i - \widehat{y}_i$. This residual can be interpreted as the error the new tree made which needs to be corrected by subsequent trees and provides the target for the next tree. (Hastie et al., 2009; James et al., 2021)

2.4 Hyperparameter tuning

Tuning of hyperparameters was performed using a 10x10-fold cross-validation approach which was also used in gathering the final results. Though random search algorithms provide a veritable alternative, we chose to perform a complete grid-search over the defined hyperparameters.(Higgins, 2020; Refaeilzadeh et al., 2009)

2.5 Dataset splitting

Because of the relatively small dataset we chose to rely on a train-test approach with a 90-10 % split for training, testing and validation respectively.

We also chose to only move a patient's medical record as a whole, so each pregnancy was either part of the training set or the test set but not both.

2.6 Clinical decision making models

In order to properly compare the machine-learning models to clinical decision-making we tried to imitate it by also establishing models that were based on defined cutoff on certain parameters (blood pressure, proteinuria, sFlt-PlGF-ratio).

The first model was a basic blood-pressure model, which predicted an adverse outcome in the future if either the systolic blood pressure was above 140 mmHg or the diastolic blood pressure was above 90 mmHg.(World Health Organization, n.d.)

The second model was based on known cutoffs on the sFlt-PlGF-ratio which is known to be highly predictive for adverse outcomes, especially for the rule-out of such events.(Dröge et al., 2021; Verlohren et al., 2014)

Third we combined the first two models via a logical OR and additionally added proteinuria measurement via dipsticks reading above '++' or a 24h urine sampling containing more than 300mg of protein.

2.7 Evaluation metrics

We based our statistical evaluation on basic confusion-matrix-derived metrics which are a very commonly used tool for ML-algorithm evaluation. We focused on sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, area under the receiver operating characteristics curve (ROCAUC) and the F1-Score, which combines sensitivity and positive predictive value in a single metric via the harmonic mean: $F1 = 2 * \frac{sensitivity * PPV}{sensitivity + PPV}$.

2.8 Cutoffs on probability score

The native statistical classifiers were initially set up to produce a probability score using a naive cutoff for positive classification, labelling any output of ≥ 0.5 as positive in terms of likelihood of sustaining an adverse outcome in the future.

We also investigated whether optimising this cutoff for any of the "composite" metrics would result in significant performance gains for the other metrics assessed.

For each fold and individual run on the test data we examined the output of the machine-learning classifiers and applied all possible cutoffs c with two-point decimal precision and $0 \leq c \leq 1$. We recorded the cutoffs that resulted in the highest value for the individual composite metric as well as the algorithm's performance on the other assessed metrics at that cutoff.

2.9 Statistical analysis

In case of comparison of normally distributed variables we chose Welch's t-test(Welch, 1947) to test for statistical significance. If one or both variables did not follow a standard distribution we used Wilcoxon's signed rank test(Wilcoxon, 1945). The test for standard distribution was performed using the Kolmogorov-Smirnov-Test.

Categorical variables were compared using Fisher's exact test.

P-values of ≤ 0.05 were considered significant while values ≤ 0.001 were considered highly significant.

We applied Bonferroni-corrected significance in case of comparison of multiple variables meaning for n assessed features we chose a significance level of $\frac{\alpha}{n}$ with α being the overall applied significance level.(Dunn, 1961)

2.10 Feature importance & interpretation

Interpretable machine-learning is an immense concern in the current research community. Methods that rely on a black-box-approach are harder for humans to accept and their results are more difficult to justify. Our base learners, decision trees, are among the most interpretable machine-learning models, but their interpretability is greatly diminished by our use of enhancement techniques, gradient boosting and random forest (bagging and feature sampling).

We attempted to give back some explicability by analysing the decision-making via shapley-values. Shapley values(Aumann and Shapley, 2015; Hart, 1989; Merrick and Taly, 2019; Roth, 1988; Shapley, 1951), introduced in 1951, frame a prediction as a coalitional game “played” by the instance’s feature.

The basic definition of shapley values is the presence of a value function v: $S \to \mathbb{R}$ with $S \subseteq \{1,..., p\}$ with $\{1,..., p\}$ being the set of features. The shapley value or contribution to the prediction for a feature $i \in \{1,..., p\}$ is given by:

$$\delta_i(v) = \sum_{S \subseteq \{1,...,p\} \setminus \{i\}} \frac{|S|!\,(p-|S|-1)!}{p!} \left(v(S \cup \{i\}) - v(S)\right)$$

This can be interpreted as the average contribution this feature makes to a possible set of features over the total number of permutations of features which can be formed without that specific feature. (Ichiishi and Shell, n.d.)

2.11 Model calibration

A model's calibration is the quality of fit for the expression of a model’s output as a predicted probability for an observation to match the algorithm's output.

It can, for example, be expressed as the mean absolute error (MAE) of the bucketed model output vs the fraction of positives in that bucket.

$$C = \sum_{i=1}^{B} |(p_i - frac_i)|$$

B being the number of buckets, p_i being the average predicted value for a given observation in that bucket, $frac_i$ being the fraction of actual instances of the target value in that bucket.

2.12 Software

All statistical analysis and data processing was performed using the Python programming language. Machine-learning models were implemented using the xgboost-package(Chen and Guestrin, 2016), data processing relied on the pandas(team, 2020), scikit-learn(Pedregosa et al., 2011) and numpy(Harris et al., 2020) software-packages. Calculation of shapley values relied on the shap-package for Python.(Lundberg et al., 2020; Lundberg and Lee, 2017)

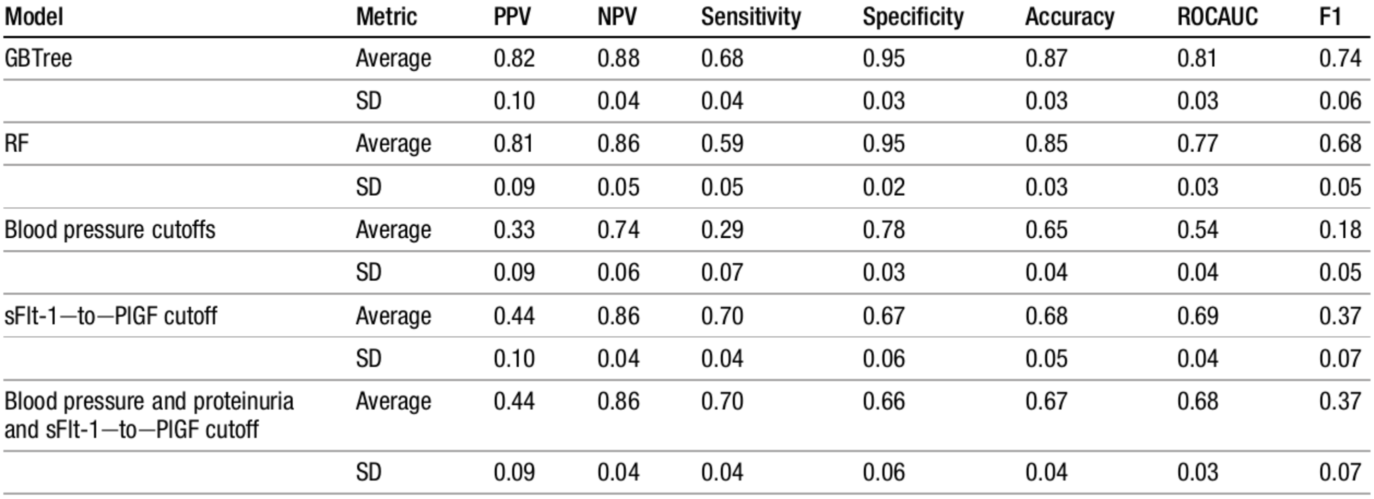
3. Results

3.1 ML-model results

We were able to show that the machine-learning classifiers greatly outperformed the conventional cutoff-based diagnostic tools on our dataset(Schmidt et al., 2022).

Table 1:

Statistical classifiers' and clinical decision making models' performance



Model	Metric	PPV	NPV	Sensitivity	Specificity	Accuracy	ROCAUC	F1
GBTree	Average	0.82	0.88	0.68	0.95	0.87	0.81	0.74
	SD	0.10	0.04	0.04	0.03	0.03	0.03	0.06
RF	Average	0.81	0.86	0.59	0.95	0.85	0.77	0.68
	SD	0.09	0.05	0.05	0.02	0.03	0.03	0.05
Blood pressure cutoffs	Average	0.33	0.74	0.29	0.78	0.65	0.54	0.18
	SD	0.09	0.06	0.07	0.03	0.04	0.04	0.05
sFlt-1—to—PlGF cutoff	Average	0.44	0.86	0.70	0.67	0.68	0.69	0.37
	SD	0.10	0.04	0.04	0.06	0.05	0.04	0.07
Blood pressure and proteinuria and sFlt-1—to—PlGF cutoff	Average	0.44	0.86	0.70	0.66	0.67	0.68	0.37
	SD	0.09	0.04	0.04	0.06	0.04	0.03	0.07

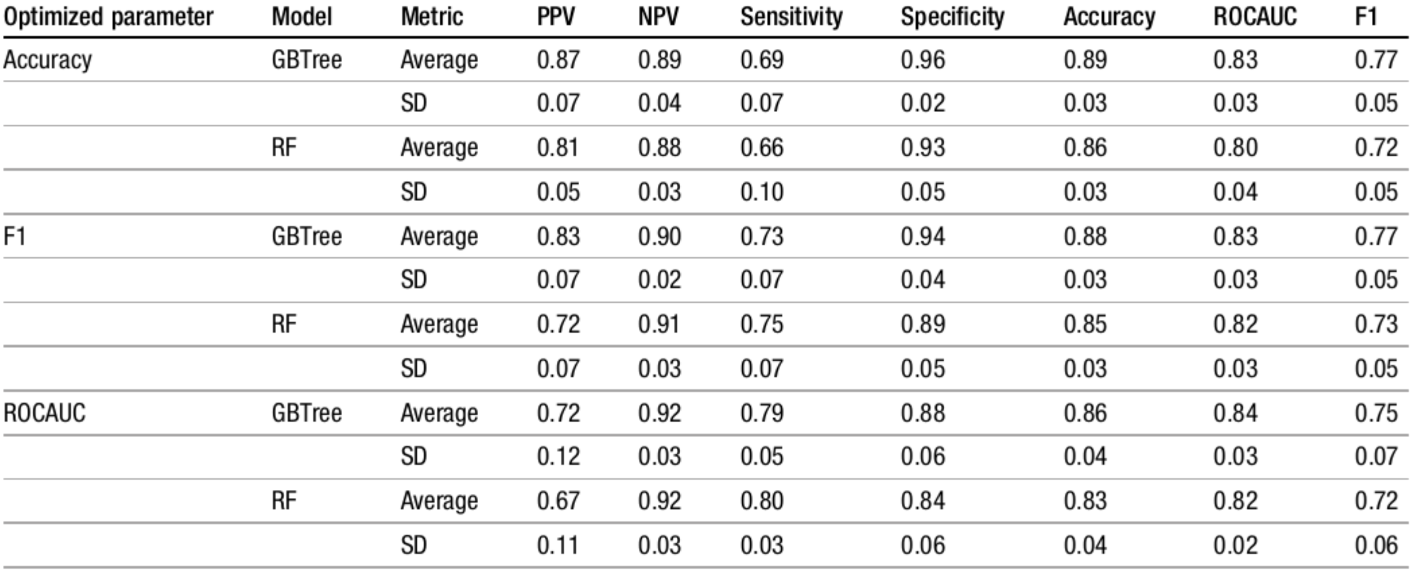
GBTree: Gradient boosted trees, RF: Random forest, PPV:positive predictive value, NPV: negative predictive value, ROCAUC: receiver operating characteristics area under the curve, SD: standard deviation, PlGF: placental growth factor, sFlt-1: soluble fms-like tyrosine kinase-1

Source: Schmidt et al. Machine-learning prediction in preeclampsia. Am J Obstet Gynecol 2022.

The most significant improvement could be made in terms of PPV, a metric that has traditionally been the most difficult to make significant improvements on using biomarkers or ultrasound findings. All other metrics also proved superior to conventional predictions. The gradient-boosted trees algorithm achieved the highest performance with an accuracy of 0.87 ± 0.03 %, high specificity (0.95 ± 0.03 %) and PPV (0.82 ± 0.10 %) (*see Table 1 for results*). Another noteworthy property of the ML approaches is the substantial increase in the F1-score as a metric for the correct classification of events (0.74 ± 0.06 for GBTree, 0.37 ± 0.07 for sFlt-1/PlGF-ratio, 0.37 ± 0.07 for multi-variable clinical model).

The predictive cutoff-optimisation for one of the composite metrics also yielded very promising results. Performance was generally improved in comparison to the “naive” classifiers (*see Table 2 for results*). The accuracy-optimised gradient-boosted trees performance showed great promise with a PPV of 0.87 ± 0.07 an accuracy of 0.89 ± 0.03 % and an F1-Score of 0.77 ± 0.05, equivalent to that of the F1-Score-optimised model.

Table 2:



Optimized parameter	Model	Metric	PPV	NPV	Sensitivity	Specificity	Accuracy	ROCAUC	F1
Accuracy	GBTree	Average	0.87	0.89	0.69	0.96	0.89	0.83	0.77
		SD	0.07	0.04	0.07	0.02	0.03	0.03	0.05
	RF	Average	0.81	0.88	0.66	0.93	0.86	0.80	0.72
		SD	0.05	0.03	0.10	0.05	0.03	0.04	0.05
F1	GBTree	Average	0.83	0.90	0.73	0.94	0.88	0.83	0.77
		SD	0.07	0.02	0.07	0.04	0.03	0.03	0.05
	RF	Average	0.72	0.91	0.75	0.89	0.85	0.82	0.73
		SD	0.07	0.03	0.07	0.05	0.03	0.03	0.05
ROCAUC	GBTree	Average	0.72	0.92	0.79	0.88	0.86	0.84	0.75
		SD	0.12	0.03	0.05	0.06	0.04	0.03	0.07
	RF	Average	0.67	0.92	0.80	0.84	0.83	0.82	0.72
		SD	0.11	0.03	0.03	0.06	0.04	0.02	0.06

Cutoff-optimised models' performance

GBTree: Gradient boosted trees, RF: Random forest, PPV:positive predictive value, NPV: negative predictive value, ROCAUC: receiver operating characteristics area under the curve, SD: standard deviation

Source: Schmidt et al. Machine-learning prediction in preeclampsia. Am J Obstet Gynecol 2022.

3.2 Dataset statistics

Our patient collective encompassed a total of 1647 patients with 2472 visits with the median number of visits being 1. Gestational age showed variety with a SD of 41 days and a mean of 244 days (34 weeks of gestation + 6 days). 914 patients in our dataset had a delivery before the 34th gestational week.

The total number of patients with adverse outcomes was 386 with 339 of those occurring before the woman’s 34th week of gestation. The most common adverse outcome was premature birth due to preeclampsia or related illnesses with 253 women sustaining that outcome. The second most common outcome was respiratory distress syndrome (RDS) which occurred in 190 children. Maternal outcomes were comparatively rare with the most common being HELLP-syndrome (33 women), renal failure (8 women) and lung edema (5 women).

Features that differed on a highly significant level between the group of women that sustained an adverse outcome and those that did not were diastolic and systolic blood pressure measurements, gestational age, all features related to biomarker measurements (absolute values, percentiles, deviations from mean) for sFlt-1 and PlGF and their ratio, pulsatility indexes for the umbilical and uterine arteries.

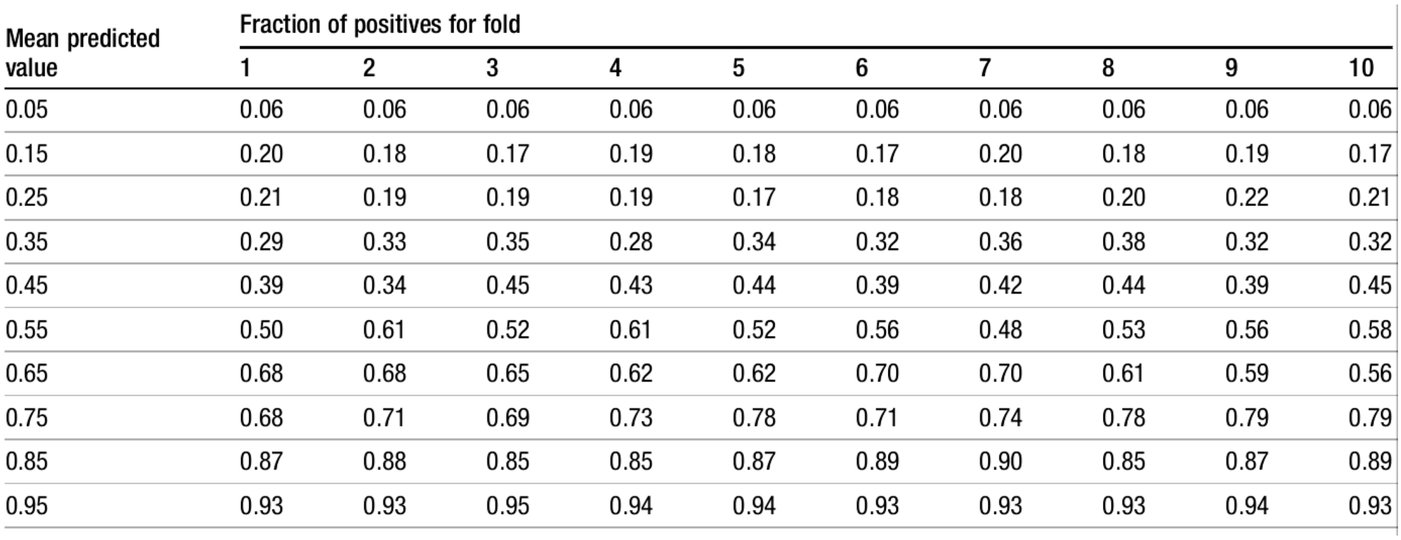
3.3 Shapley values

We are reporting the mean absolute shapley values for each feature assessed. The features that, on average, contributed the most to the algorithms output were gestational age in days (0.725 ± 0.073), sFlt-1-PlGF-ratio outside of the 95th percentile (0.237 ± 0.021), sFlt-1-PlGF-ratio multiple of the median (MoM) (0.225 ± 0.021), absolute sFlt-1 value in ng/dL (0.184 ± 0.023), PlGF value as deviation from the median (0.142 ± 0.031) and the woman’s height (0.142 ± 0.031). Please refer to Schmidt et. al(Schmidt et al., 2022) for the complete list.(Schmidt et al., 2022)

3.4 Model calibration

The model’s calibration was also satisfactory and well aligned with the ideal calibration curve (see figure 3, table 3). The mean predicted values for each fold also aligned well with an ideal calibration.

Table 3:



Mean predicted value	Fraction of positives for fold									
	1	2	3	4	5	6	7	8	9	10
0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
0.15	0.20	0.18	0.17	0.19	0.18	0.17	0.20	0.18	0.19	0.17
0.25	0.21	0.19	0.19	0.19	0.17	0.18	0.18	0.20	0.22	0.21
0.35	0.29	0.33	0.35	0.28	0.34	0.32	0.36	0.38	0.32	0.32
0.45	0.39	0.34	0.45	0.43	0.44	0.39	0.42	0.44	0.39	0.45
0.55	0.50	0.61	0.52	0.61	0.52	0.56	0.48	0.53	0.56	0.58
0.65	0.68	0.68	0.65	0.62	0.62	0.70	0.70	0.61	0.59	0.56
0.75	0.68	0.71	0.69	0.73	0.78	0.71	0.74	0.78	0.79	0.79
0.85	0.87	0.88	0.85	0.85	0.87	0.89	0.90	0.85	0.87	0.89
0.95	0.93	0.93	0.95	0.94	0.94	0.93	0.93	0.93	0.94	0.93

Model calibration as expressed by fraction of predicted positive and actual fraction of positives

Source: Schmidt et al. Machine-learning prediction in preeclampsia. Am J Obstet Gynecol 2022.

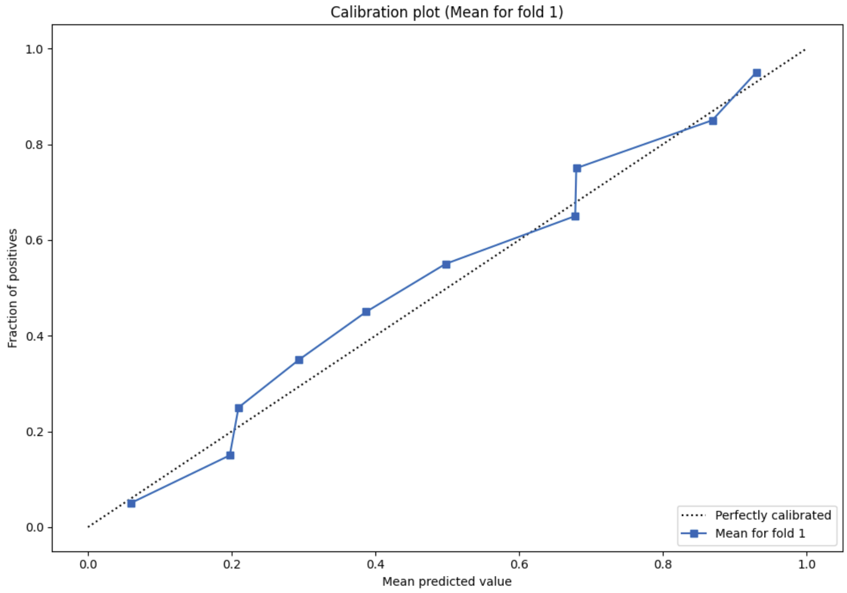


Figure 4: Exemplary mean calibration for fold 1

Exemplary calibration curve for the first fold (average of 10 models). It represents the mean predicted output of the classifier GBTree mapped onto the fraction of positives among the observations with that output.

Source: Schmidt et al. Machine-learning prediction in preeclampsia. Am J Obstet Gynecol 2022.

4. Discussion

4.1 Short summary of results

In the presented work we were able to show that models trained on the full feature set proved superior to cutoff-based decision-making currently in use in hospitals, especially in terms of PPV. The biomarkers' (sFlt-1/PlGF-ratio)(Verlohren et al., 2014) introduction already provided a gain in specificity but lacked sufficient improvement in the positive predictive value.(Dröge et al., 2021; Verlohren et al., 2014) Our models close this gap and expand the possibilities for identifying adverse outcomes throughout the pregnancy. Another major improvement is the focus on adverse outcomes rather than the prediction of preeclampsia itself. Preeclampsia constitutes a major risk factor for secondary disease, but doesn't necessarily constitute manifest harm to the mother or the child. By focussing on these direct impacts we are able to more accurately assess which women need more immediate attention by the physicians and which women can safely be discharged from the hospital.

The models also appear to be well calibrated which means they can return a rough estimate of the likelihood of sustaining an adverse outcome. This lends itself to an application as a decision-support tool because it places the output in an interpretable context for the physician. This is also strengthened by the inclusion of shapley-values to formulate the model's weighing of the different input features.

Overall we demonstrate a novel approach to the challenging topic of accurately identifying women at risk for adverse outcomes among the already high-risk-pregnancy population.

This aligns well with our hypothesis that the application of machine-learning in medicine will improve patient outcomes.

4.2 Interpretation of results

Overall we were able to show that machine-learning could present a veritable improvement for the clinical process patients undergo at the hospital. Correctly identifying women and children at risk for severe adverse outcomes still poses a significant challenge our model can help overcome. Diagnostics usually either relied on the prediction of preeclampsia itself(Jhee et al., 2019; Lundberg et al., 2018; Marić et al., 2020; Sufriyana et al., 2020a, 2020b) or predicted adverse outcomes without the use of artificial intelligence(Dröge et al., 2021; Mirkovic et al., 2020; Wright et al., 2012). Comparisons to these papers can be found in the Schmidt et al (2022)(Schmidt et al., 2022).

Research directly predicting pregnancy complications using advanced statistical methods is still in its infancy but significant steps have been made, especially in recent years. In 2020 Escobar et. al(Escobar et al., 2021) used electronic health records as the basis for development of a screening algorithm for severe pregnancy complications. They examined a variety of different algorithms, also encompassing the ones presented in our study with gradient boosted trees also returning the best performance. Their patient collective however differed greatly from the one we examined in our study, first in terms of size, constituted and second in risk-profile, which was generally lower than in our cohort. Also they predicted their outcomes in a much shorter period between 3 and 12 hours, which our models extend to at any point in the future pregnancy. (Comparison to Jhee, Lindstroem, Maric etc.)

In a recently published review Bertini et al.(Bertini et al., 2022) investigated the overall application of machine-learning in prediction of pregnancy complications. They analysed 31 studies that examined a variety of different pregnancy outcomes which all relied on machine-learning as the statistical tool of choice. They grouped the studies in three broad categories based on the underlying data: electronic health records, medical imaging and laboratory parameters, all of which we combine in our approach. Compared to the electronic-health-record (EHR) subset our dataset was rather limited in size due to the non-automated data gathering, single-centre-focus and plenitude of features of our study, but in comparison to the laboratory marker and imaging subgroups our dataset proved reasonably sizeable. In terms of results our study aligns well with the overall group, especially compared to the EHR subset, which showed a mean ROC AUC of 0.799±0.069, although this comparison encompasses a variety of different targets and should be viewed with care. All these studies, as well as our own, were simply retrospective studies and the need for prospective evaluation has to be emphasised.

In conjunction with these studies our concept further underscores the potential machine-learning has in making pregnancies safer. This applies to both identification of patients at risk for complications and identification of patients that exhibit risk factors but won't develop complications. This gives both security to women at risk and removes the need for preemptive hospitalisation and elaborate diagnostic procedures or even unnecessary induction of labour.

4.3 Embedding the results into the current state of research

Machine-learning techniques provide an immense potential to change the future of patient healthcare in pregnancy. Decision support systems and their application in the clinic will lead to safer and more individualised healthcare for both mothers and children. This capacity though is not unique to pregnancy research, OBGYN or even medicine but rather represents a general trend in technological progress due to exponential increases in computational power (see figure 2).

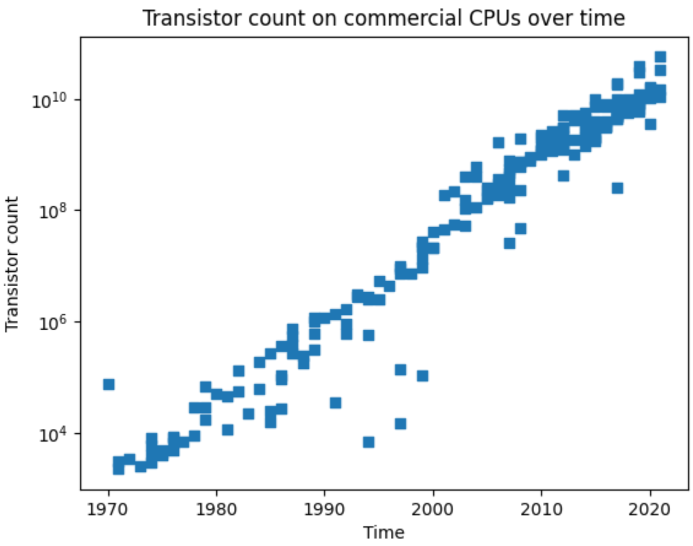


Figure 5: transistor count over time

Development of transistor counts as a proxy for computational power on commercial CPUs from 1970 to 2022

Data Source:

"https://en.wikipedia.org/wiki/Transistor_count"

Machine-learning has the potential to completely transform healthcare for pregnant women. We have already shown that approaches similar to our own can improve detection of adverse outcomes for women in preeclampsia. These results have also been shown in a variety of other pregnancy-related diseases(Bertini et al., 2022; Davidson and Boland, 2021; Iftikhar et al., 2020) and adjacent topics, for instance reproductive medicine(Wang et al., 2019).

These advances are not restricted to pregnancy research or gynecology in general but rather substantiate the potential advanced data analysis has in medicine. These data-driven applications can increase reliability, validity and accuracy of every step of the patient’s journey through the healthcare system.

4.3.1 Diagnostics

In diagnostics for example machine-learning can help in correctly identifying diseases and even give a prognostic evaluation. This is generally possible for every single modality that is currently available in the clinic, as long as training data is readily available in electronic form. Examples include skin lesion classification from simple photography in dermatology(Chan et al., 2020; Dhivyaa et al., 2020; Kassem et al., 2021; Marka et al., 2019; Pathan et al., 2018; Thomsen et al., 2020), in which digital images are being assessed for pathologies or identification of liver pathologies in CT/MRI scans, which uses similar approaches to reliably detect steatosis, fibrosis or neoplastic diseases(Ahn et al., 2021; Choi et al., 2018; Hill et al., 2021; Yasaka et al., 2018b, 2018a; Zhou et al., 2019). These show that machine-learning, relying on a data source and a target variable, can be employed successfully in a variety of different contexts. There is no conceivable limit to which data sources can be utilised, examples encompass electronic health records, physicians’ and nurses’ notes in free text, laboratory parameters, medical imaging, vital parameters etc. This first step of the patient journey after contact with a physician provides an immense potential for automated diagnostic support-tools and although prospective studies are still available only in limited quantities some suggest performance that is at least en-par with trained clinicians and sometimes even surpasses them.

4.3.2 Therapy

The next step in the patient journey after diagnosing the disease would be identification and application of the correct treatment for the particular disease. Even though for many diseases this process can take a straight-forward path, especially in the field of pharmacological therapy the increasingly complex interactions and possible treatments pose a significant challenge to clinicians. An archetypal example is antibiotic therapy which needs targeted medication specific to the causing agent, be it bacteria, viruses or fungi.

Predicting susceptibleness to available antibiotics currently depends on heuristics encompassing the clinicians intuition regarding the origin of the infection, the suspected range of possible causing agents and the locally observed resistance of these agents.(Paul-Ehrlich-Gesellschaft für Infektionstherapie e.V. (PEG), n.d.)

This process is error-prone and improvements in the statistical process can yield immense benefits(Paul et al., 2006), especially considering rising antibiotic resistances and the general lack of innovation regarding new antimicrobial agents.(World Health Organization, 2021) This complex, high-dimensional situation constitutes a prime example for the application of machine-learning algorithms. Recent works have shown that the time for correct antimicrobial resistance determination can be dramatically reduced(Lv et al., 2021) and that correct treatment decision(Komorowski et al., 2018) as well as antibiotic resistance prediction can be improved (Lewin-Epstein et al., 2021; Oonsivilai et al., 2018; Peiffer-Smadja et al., 2020; Weis et al., 2020). Though further research, especially prospective evaluations, are still needed, these examples show promising potential for the use of artificial intelligence in the therapeutic process in addition to the already shown capacities in diagnostics.

4.3.3 Doctor-patient relationship

Employing artificial intelligence in the hospital has immense potential to change the relationship between physicians and patients.

Several hypotheses have been posed in what this change will constitute for doctors. One opinion is that artificial intelligence will ultimately replace physicians in their role as diagnosticians(Pearson, n.d.) and therefore the current age marks the beginning of this shift.

Another school of thought takes a more integrative position on the future of medical AI and postulates that AI will complement doctors' work.(Karches, 2018; Pearson, n.d.; Recht and Bryan, 2017) This will in turn free the physician from much of the menial documentation tasks(Hill et al., 2013) and shift work more towards the interaction with patients and "perform more value-added tasks, such as integrating patients' clinical and imaging information, having more professional interactions, becoming more visible to patients and playing a vital role in integrated clinical teams to improve patient care."(Recht and Bryan, 2017) This view is also supported by evidence that artificial intelligence systems in combination with a clinician outperform either alone(Hekler et al., 2019; Sakai et al., 2022), which supports an application of ML as an additional tool for treating physicians.

4.4 Strengths and limitations of the study

The main limitations with this study are connected to the machine-learning statistics and to the data sources.

The main risk in any utilisation of machine-learning-techniques lies in overfitting on a particular dataset. Overfitting means that the patterns and logic derived from one dataset do not apply to other, previously unknown data, i.e. the algorithm can't reproduce the performance observed on the test or validation set when presented with input that was hitherto unknown. Another problem connected with ML-approaches is the general difficulty of comprehending the decisions made by the algorithms due to their highly complex nature. We tried to mitigate both problems by conservatively following best-practice guidelines(Higgins, 2020) and establishing secondary metrics such as shapley values to discern the algorithm's judgements.

From this we observed the algorithms placing strong emphasis on ultrasound and the biomarkers sFlt-1 and PlGF, which require a certain standard in terms of technological equipment and restrict the application to settings that exhibit a high standard of healthcare.

The limitations connected to the study population can be categorised in sampling biases, study size and the retrospective character of the study. The sampling biases that should be taken into consideration are the high-risk population, which naturally increases the frequency of adverse outcomes and introduces a possible intervention bias, the data collection at a single centre and connected to that the generally low diversity regarding ethnicity and sociocultural composition. The study's size also proves relatively small compared to other machine-learning research, which is owed to the laborious process of manual data gathering.

The main strength of our study is the definitive increase in predictive power over the clinical standard and biomarkers alone. It also introduces another level of objective evaluation apart from the current reliance on the physicians' clinical experience. Furthermore it serves as a formidable basis for a following prospective trial to validate the model's in a clinical setting and the development of a possible medical tool for application in a clinical setting. In addition, by relying on a well formulated approach and machine-learning best-practices(Higgins, 2020) we are confident to have mitigated many of the above mentioned limitations. Even though generalisation has yet to be shown by applying the algorithms to new, structurally more diverse datasets, we are confident that we indeed did obtain a veritable signal from our data.

4.5 Implications for practice and/or future research

Our study underlines the potential machine-learning techniques have in application to clinical problems. This in turn leads to two different avenues of future research, application of machine-learning with regard to the problem defined as the basis for our publication and application of machine-learning to medical problems in general. Concerning our research topic, a variety of new problems should be explored - application of the trained algorithms on new populations in relation to multiple centres and geographical diversity to test for generalisation, inclusion of other adverse outcomes and inclusion of patients with varying risk profiles. These analyses should be performed to solidify the external validity of our work and to create a larger dataset to train new algorithms from.

For preeclampsia our work constitutes an immense increase in statistical predictive ability, especially in terms of positive predictive value. This potentially allows for substantial changes in current clinical practice which at the moment consists of stringent hospitalisation upon formulating the suspicion of preeclampsia.(German Society of Gynecology and Obstetrics, 2019) Although only around 5% of these cases actually develop severe consequences(Magee et al., 2022; Purde et al., 2015) the current lack of rule-in tests necessitates this procedure. Our algorithms close this gap and allow for more women to stay in the comfort of their homes. It also improves overall safety due to the improved identification of patients at risk for severe outcomes. The models presented in the paper could, after careful consideration and testing, prove as viable bases for a clinical decision-support tool.

The implications for application of machine-learning in general, though our publication only represents a miniscule amount of additional knowledge, lies mainly in the further substantiation of the potential advanced statistical methods have as part of an integrated clinical system. The application of these methods should be carefully examined in both a horizontal direction, meaning a widening of the total topics ML is employed in, and a vertical direction, meaning that the entire clinical process screening, diagnostics, treatment and prevention should be considered in further analysis.

Though many limitations definitely apply to machine-learning, for example difficulty of new knowledge generation and the potential to obtain erroneous results due to statistical fluctuations, these tools are crucially underrepresented in clinical practice.(“eHealth Trend Barometer,” 2019; “Electronic Medical Record Adoption Model | HIMSS Analytics - Europe,” n.d.)

They have the potential to free resources, time and attention, and give the treating physicians more time for the doctor-patient interaction. This development, while also conferring cost-benefits by increasing the medical systems overall accuracy, might be the most important medical innovation of the 21st century.

5. Conclusions

The work presented in this text constitutes a significant improvement over the current standard of care in women with a high-risk profile for preeclampsia and associated adverse pregnancy outcomes. Though further investigation is of crucial importance, we present a stepping stone for a line of research at the end of which definitive clinical impact, in the form of decision-support or diagnostic tools, can be achieved.

Overall we have shown that the application of machine-learning to medical information can yield immense benefits for patient-care and the overall work-structure in hospitals. They can support diagnostics, increasing precision and reliability of virtually every single measurement that is taken in clinical practice. They can yield more precise treatments, as shown in the case of antibiotics, diagnoses, as demonstrated in case of preeclampsia, skin lesion- and liver radiology classification and overall improve the standard of care.

The motor behind these changes, improvements in computational power and information storage have shaped our society over the course of the last decades as few technical innovations have before. They have the potential to drastically change medicine and the way we treat health as well. This work serves as a small contribution to make this possible.

Reference list

Ahn, J.C., Connell, A., Simonetto, D.A., Hughes, C., Shah, V.H., 2021. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. Hepatology 73, 2546–2563. https://doi.org/10.1002/hep.31603

Aumann, R.J., Shapley, L.S., 2015. Values of Non-Atomic Games: Princeton University Press. https://doi.org/10.1515/9781400867080

Bertini, A., Salas, R., Chabert, S., Sobrevia, L., Pardo, F., 2022. Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review. Front. Bioeng. Biotechnol. 9, 780389. https://doi.org/10.3389/fbioe.2021.780389

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1007/BF00058655

Burkov, A., 2019. The Hundred-Page Machine Learning Book. Andriy Burkov.

Campbell, M., Hoane, A.J., Hsu, F., 2002. Deep Blue. Artif. Intell. 134, 57–83. https://doi.org/10.1016/S0004-3702(01)00129-1

Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N., Liao, W., 2020. Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations. Dermatol. Ther. 10, 365–386. https://doi.org/10.1007/s13555-020-00372-0

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD ’16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Choi, K.J., Jang, J.K., Lee, S.S., Sung, Y.S., Shim, W.H., Kim, H.S., Yun, J., Choi, J.-Y., Lee, Y., Kang, B.-K., Kim, J.H., Kim, S.Y., Yu, E.S., 2018. Development and Validation of a Deep Learning System for Staging Liver Fibrosis by Using Contrast Agent–enhanced CT Images in the Liver. Radiology 289, 688–697. https://doi.org/10.1148/radiol.2018180763

Ciobanu, A., Wright, A., Syngelaki, A., Wright, D., Akolekar, R., Nicolaides, K.H., 2019. Fetal Medicine Foundation reference ranges for umbilical artery and middle cerebral artery pulsatility index and cerebroplacental ratio. Ultrasound Obstet. Gynecol. 53, 465–472. https://doi.org/10.1002/uog.20157

Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J., 2011. Flexible, high performance convolutional neural networks for image classification, in: Twenty-Second International Joint Conference on Artificial Intelligence.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random Forests, in: Zhang, C., Ma, Y. (Eds.), Ensemble Machine Learning. Springer US, Boston, MA, pp. 157–175. https://doi.org/10.1007/978-1-4419-9326-7_5

Davidson, L., Boland, M.R., 2021. Towards deep phenotyping pregnancy: a systematic review on artificial intelligence and machine learning methods to improve pregnancy outcomes. Brief. Bioinform. 22, bbaa369. https://doi.org/10.1093/bib/bbaa369

Dhivyaa, C.R., Sangeetha, K., Balamurugan, M., Amaran, S., Vetriselvi, T., Johnpaul, P., 2020. Skin lesion classification using decision trees and random forest algorithms. J. Ambient Intell. Humaniz. Comput. https://doi.org/10.1007/s12652-020-02675-8

Dixon, M.F., Halperin, I., Bilokon, P., 2020. Machine Learning in Finance: From Theory to Practice. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-41068-1

Dröge, L.A., Perschel, F.H., Stütz, N., Gafron, A., Frank, L., Busjahn, A., Henrich, W., Verlohren, S., 2021. Prediction of Preeclampsia-Related Adverse Outcomes With the sFlt-1 (Soluble fms-Like Tyrosine Kinase 1)/PlGF (Placental Growth Factor)-Ratio in the Clinical Routine: A Real-World Study. Hypertension 77, 461–471. https://doi.org/10.1161/HYPERTENSIONAHA.120.15146

Dunn, O.J., 1961. Multiple Comparisons among Means. J. Am. Stat. Assoc. 56, 52–64. https://doi.org/10.1080/01621459.1961.10482090

eHealth Trend Barometer: AI Use in European Healthcare [WWW Document], 2019. . HIMSS Anal. - Eur. URL https://www.himssanalytics.org/europe/ehealth-barometer/ehealth-trend-barometer-ai-use-european-healthcare (accessed 7.29.22).

Electronic Medical Record Adoption Model | HIMSS Analytics - Europe [WWW Document], n.d. URL https://www.himssanalytics.org/europe/electronic-medical-record-adoption-model (accessed 7.29.22).

Escobar, G.J., Soltesz, L., Schuler, A., Niki, H., Malenica, I., Lee, C., 2021. Prediction of obstetrical and fetal complications using automated electronic health record data. Am. J. Obstet. Gynecol. 224, 137-147.e7. https://doi.org/10.1016/j.ajog.2020.10.030

Friedman, J.H., 2002. Stochastic Gradient Boosting. Comput Stat Data Anal 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29. https://doi.org/10.1214/aos/1013203451

German Society of Gynecology and Obstetrics, 2019. Hypertensive Pregnancy Disorders: Diagnosis and Therapy (S2k-Level, AWMF Registry No. 015/018).

Harris, C.R., Millman, K.J., Walt, S.J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M.H. van, Brett, M., Haldane, A., Río, J.F. del, Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hart, S., 1989. Shapley Value, in: Eatwell, J., Milgate, M., Newman, P. (Eds.), Game Theory. Palgrave Macmillan UK, London, pp. 210–216. https://doi.org/10.1007/978-1-349-20181-5_25

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference and prediction, 2nd ed. Springer.

Hekler, A., Utikal, J.S., Enk, A.H., Hauschild, A., Weichenthal, M., Maron, R.C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., Schilling, B., Holland-Letz, T., Izar, B., von Kalle, C., Fröhling, S., Brinker, T.J., Schmitt, L., Peitsch, W.K., Hoffmann, F., Becker, J.C., Drusio, C., Jansen, P., Klode, J., Lodde, G., Sammet, S., Schadendorf, D., Sondermann, W., Ugurel, S., Zader, J., Enk, A., Salzmann, M., Schäfer, S., Schäkel, K., Winkler, J., Wölbing, P., Asper, H., Bohne, A.-S., Brown, V., Burba, B., Deffaa, S., Dietrich, C., Dietrich, M., Drerup, K.A., Egberts, F., Erkens, A.-S., Greven, S., Harde, V., Jost, M., Kaeding, M., Kosova, K., Lischner, S., Maagk, M., Messinger, A.L., Metzner, M., Motamedi, R., Rosenthal, A.-C., Seidl, U., Stemmermann, J., Torz, K., Velez, J.G., Haiduk, J., Alter, M., Bär, C., Bergenthal, P., Gerlach, A., Holtorf, C., Karoglan, A., Kindermann, S., Kraas, L., Felcht, M., Gaiser, M.R., Klemke, C.-D., Kurzen, H., Leibing, T., Müller, V., Reinhard, R.R., Utikal, J., Winter, F., Berking, C., Eicher, L., Hartmann, D., Heppt, M., Kilian, K., Krammer, S., Lill, D., Niesert, A.-C., Oppel, E., Sattler, E., Senner, S., Wallmichrath, J., Wolff, H., Gesierich, A., Giner, T., Glutsch, V., Kerstan, A., Presser, D., Schrüfer, P., Schummer, P., Stolze, I., Weber, J., Drexler, K., Haferkamp, S., Mickler, M., Stauner, C.T., Thiem, A., 2019. Superior skin cancer classification by the combination of human and artificial intelligence. Eur. J. Cancer 120, 114–121. https://doi.org/10.1016/j.ejca.2019.07.019

Henry, K.E., Kornfield, R., Sridharan, A., Linton, R.C., Groh, C., Wang, T., Wu, A., Mutlu, B., Saria, S., 2022. Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. Npj Digit. Med. 5, 97. https://doi.org/10.1038/s41746-022-00597-7

Higgins, D., 2020. OnRAMP for Regulating AI in Medical Products. https://doi.org/10.48550/ARXIV.2010.07038

Hill, C.E., Biasiolli, L., Robson, M.D., Grau, V., Pavlides, M., 2021. Emerging artificial intelligence applications in liver magnetic resonance imaging. World J. Gastroenterol. 27, 6825–6843. https://doi.org/10.3748/wjg.v27.i40.6825

Hill, R.G., Sears, L.M., Melanson, S.W., 2013. 4000 Clicks: a productivity analysis of electronic medical records in a community hospital ED. Am. J. Emerg. Med. 31, 1591–1594. https://doi.org/10.1016/j.ajem.2013.06.028

Ichiishi, T., Shell, K., n.d. Game Theory for Economic Analysis. New York: Academic Press.

Iftikhar, P.M., Kuijpers, M.V., Khayyat, A., Iftikhar, A., DeGouvia De Sa, M., 2020. Artificial Intelligence: A New Paradigm in Obstetrics and Gynecology Research and Clinical Practice. Cureus. https://doi.org/10.7759/cureus.7124

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. Tree-Based Methods, in: An Introduction to Statistical Learning, Springer Texts in Statistics. Springer US, New York, NY, pp. 327–365. https://doi.org/10.1007/978-1-0716-1418-1_8

Jhee, J.H., Lee, S., Park, Y., Lee, S.E., Kim, Y.A., Kang, S.-W., Kwon, J.-Y., Park, J.T., 2019. Prediction model development of late-onset preeclampsia using machine learning-based methods. PLOS ONE 14, e0221202. https://doi.org/10.1371/journal.pone.0221202

Joyce, J., 2021. Bayes' Theorem, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Karches, K.E., 2018. Against the iDoctor: why artificial intelligence should not replace physician judgment. Theor. Med. Bioeth. 39, 91–110. https://doi.org/10.1007/s11017-018-9442-3

Kassem, M.A., Hosny, K.M., Damaševičius, R., Eltoukhy, M.M., 2021. Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review. Diagnostics 11, 1390. https://doi.org/10.3390/diagnostics11081390

Komorowski, M., Celi, L.A., Badawi, O., Gordon, A.C., Faisal, A.A., 2018. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat. Med. 24, 1716–1720. https://doi.org/10.1038/s41591-018-0213-5

Le, Q.V., 2013. Building high-level features using large scale unsupervised learning, in: 2013 IEEE International Conference on Acoustics, Speech and Signal

Processing. Presented at the ICASSP 2013 - 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Vancouver, BC, Canada, pp. 8595–8598. https://doi.org/10.1109/ICASSP.2013.6639343

Lewin-Epstein, O., Baruch, S., Hadany, L., Stein, G.Y., Obolski, U., 2021. Predicting Antibiotic Resistance in Hospitalized Patients by Applying Machine Learning to Electronic Medical Records. Clin. Infect. Dis. 72, e848–e855. https://doi.org/10.1093/cid/ciaa1576

Lisonkova, S., Joseph, K.S., 2013. Incidence of preeclampsia: risk factors and outcomes associated with early- versus late-onset disease. Am. J. Obstet. Gynecol. 209, 544.e1-544.e12. https://doi.org/10.1016/j.ajog.2013.08.019

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 2522–5839.

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774.

Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., Lee, S.-I., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. 2, 749–760. https://doi.org/10.1038/s41551-018-0304-0

Lv, J., Deng, S., Zhang, L., 2021. A review of artificial intelligence applications for antimicrobial resistance. Biosaf. Health 3, 22–31. https://doi.org/10.1016/j.bsheal.2020.08.003

Magee, L.A., Nicolaides, K.H., von Dadelszen, P., 2022. Preeclampsia. N. Engl. J. Med. 386, 1817–1832. https://doi.org/10.1056/NEJMra2109523

Marić, I., Tsur, A., Aghaeepour, N., Montanari, A., Stevenson, D.K., Shaw, G.M., Winn, V.D., 2020. Early prediction of preeclampsia via machine learning. Am. J. Obstet. Gynecol. MFM 2, 100100. https://doi.org/10.1016/j.ajogmf.2020.100100

Marka, A., Carter, J.B., Toto, E., Hassanpour, S., 2019. Automated detection of nonmelanoma skin cancer using digital images: a systematic review. BMC Med. Imaging 19, 21. https://doi.org/10.1186/s12880-019-0307-7

Mason, L., Baxter, J., Bartlett, P., Frean, M., 1999. Boosting Algorithms as Gradient Descent, in: Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS'99. MIT Press, Cambridge, MA, USA, pp. 512–518.

Merrick, L., Taly, A., 2019. The Explanation Game: Explaining Machine Learning Models Using Shapley Values. https://doi.org/10.48550/ARXIV.1909.08128

Mirkovic, L., Tulic, I., Stankovic, S., Soldatovic, I., 2020. Prediction of adverse maternal outcomes of early severe preeclampsia. Pregnancy Hypertens. 22, 144–150. https://doi.org/10.1016/j.preghy.2020.09.009

Mohammed, M., Khan, M.B., Bashier, E.B.M., 2016. Machine Learning, 0 ed. CRC Press. https://doi.org/10.1201/9781315371658

Moore's Law: The number of transistors per microprocessor [WWW Document], n.d. URL https://ourworldindata.org/grapher/transistors-per-microprocessor (accessed 5.22.22).

Oh, K.-S., Jung, K., 2004. GPU implementation of neural networks. Pattern Recognit. 37, 1311–1314. https://doi.org/10.1016/j.patcog.2004.01.013

Oonsivilai, M., Mo, Y., Luangasanatip, N., Lubell, Y., Miliya, T., Tan, P., Loeuk, L., Turner, P., Cooper, B.S., 2018. Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia. Wellcome Open Res. 3, 131. https://doi.org/10.12688/wellcomeopenres.14847.1

Pathan, S., Prabhu, K.G., Siddalingaswamy, P.C., 2018. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review. Biomed. Signal Process. Control 39, 237–262. https://doi.org/10.1016/j.bspc.2017.07.010

Paul, M., Andreassen, S., Tacconelli, E., Nielsen, A.D., Almanasreh, N., Frank, U., Cauda, R., Leibovici, L., 2006. Improving empirical antibiotic treatment using TREAT, a computerized decision support system: cluster randomized trial. J. Antimicrob. Chemother. 58, 1238–1245. https://doi.org/10.1093/jac/dkl372

Paul-Ehrlich-Gesellschaft für Infektionstherapie e.V. (PEG), n.d. AWMF Leitlinie: Kalkulierte parenterale Initialtherapie bakterieller Erkrankungen bei Erwachsenen - Update 2018 [WWW Document]. URL https://www.awmf.org/leitlinien/detail/ll/082-006.html (accessed 7.30.22).

Pearson, D., n.d. Artificial Intelligence in Radiology: The Game-Changer on Everyone's Mind [WWW Document]. URL https://www.radiologybusiness.com/topics/medical-imaging/artificial-intelligence-radiology-game-changer-everyones-mind (accessed 6.18.22).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peiffer-Smadja, N., Dellière, S., Rodriguez, C., Birgand, G., Lescure, F.-X., Fourati, S., Ruppé, E., 2020. Machine learning in the clinical microbiology laboratory: has the time come for routine practice? Clin. Microbiol. Infect. 26, 1300–1309. https://doi.org/10.1016/j.cmi.2020.02.006

Purde, M., Baumann, M., Wiedemann, U., Nydegger, U., Risch, L., Surbek, D., Risch, M., 2015. Incidence of preeclampsia in pregnant Swiss women. Swiss Med. Wkly. https://doi.org/10.4414/smw.2015.14175

Recht, M., Bryan, R.N., 2017. Artificial Intelligence: Threat or Boon to Radiologists? J. Am. Coll. Radiol. 14, 1476–1480. https://doi.org/10.1016/j.jacr.2017.07.007

Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-Validation, in: Liu, L., Özsu, M.T. (Eds.), Encyclopedia of Database Systems. Springer US, Boston, MA, pp. 532–538. https://doi.org/10.1007/978-0-387-39940-9_565

Roth, A.E. (Ed.), 1988. The Shapley Value: Essays in Honor of Lloyd S. Shapley, 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511528446

Sakai, A., Komatsu, M., Komatsu, R., Matsuoka, R., Yasutomi, S., Dozen, A., Shozu, K., Arakaki, T., Machino, H., Asada, K., Kaneko, S., Sekizawa, A., Hamamoto, R., 2022. Medical Professional Enhancement Using Explainable Artificial Intelligence in Fetal Cardiac Ultrasound Screening. Biomedicines 10, 551. https://doi.org/10.3390/biomedicines10030551

Samuel, A.L., 1959. Some studies in machine learning using the game of Checkers. IBM J. Res. Dev. 71–105.

Schaller, R.R., 1997. Moore's law: past, present and future. IEEE Spectr. 34, 52–59. https://doi.org/10.1109/6.591665

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw. 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schmidt, L.J., Rieger, O., Neznansky, M., Hackelöer, M., Dröge, L.A., Henrich, W., Higgins, D., Verlohren, S., 2022. A machine-learning–based algorithm improves prediction of preeclampsia-associated adverse outcomes. Am. J. Obstet. Gynecol. S0002937822000503. https://doi.org/10.1016/j.ajog.2022.01.026

Shapley, L.S., 1951. Notes on the N-Person Game &mdash; II: The Value of an N-Person Game. RAND Corporation, Santa Monica, CA. https://doi.org/10.7249/RM0670

Shaw, G.L., 1986. Donald Hebb: The Organization of Behavior, in: Palm, G., Aertsen, A. (Eds.), Brain Theory. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 231–233. https://doi.org/10.1007/978-3-642-70911-1_15

Singh, H., Meyer, A.N.D., Thomas, E.J., 2014. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual. Saf. 23, 727–731. https://doi.org/10.1136/bmjqs-2013-002627

Stevens, W., Shih, T., Incerti, D., Ton, T.G.N., Lee, H.C., Peneva, D., Macones, G.A., Sibai, B.M., Jena, A.B., 2017. Short-term costs of preeclampsia to the United States health care system. Am. J. Obstet. Gynecol. 217, 237-248.e16. https://doi.org/10.1016/j.ajog.2017.04.032

Sufriyana, H., Wu, Y.-W., Su, E.C.-Y., 2020a. Prediction of Preeclampsia and Intrauterine Growth Restriction: Development of Machine Learning Models on a Prospective Cohort. JMIR Med. Inform. 8, e15411. https://doi.org/10.2196/15411

Sufriyana, H., Wu, Y.-W., Su, E.C.-Y., 2020b. Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia. EBioMedicine 54, 102710. https://doi.org/10.1016/j.ebiom.2020.102710

Taigman, Y., Yang, M., Ranzato, M., Wolf, L., 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Columbus, OH, USA, pp. 1701–1708. https://doi.org/10.1109/CVPR.2014.220

team, T. pandas development, 2020. pandas-dev/pandas: Pandas. https://doi.org/10.5281/zenodo.3509134

The Fetal Medicine Foundation Risk for preeclampsia calculator [WWW Document], n.d. URL https://fetalmedicine.org/research/assess/preeclampsia/first-trimester (accessed 6.22.22).

Thomsen, K., Iversen, L., Titlestad, T.L., Winther, O., 2020. Systematic review of machine learning for diagnosis and prognosis in dermatology. J. Dermatol. Treat. 31, 496–510. https://doi.org/10.1080/09546634.2019.1682500

Tin Kam Ho, 1995. Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition. Presented at the 3rd International Conference on Document Analysis and Recognition, IEEE Comput. Soc. Press, Montreal, Que., Canada, pp. 278–282. https://doi.org/10.1109/ICDAR.1995.598994

Tipping, M.D., Forth, V.E., O’Leary, K.J., Malkenson, D.M., Magill, D.B., Englert, K., Williams, M.V., 2010. Where did the day go?-A time-motion study of hospitalists. J. Hosp. Med. 5, 323–328. https://doi.org/10.1002/jhm.790

Verlohren, S., Herraiz, I., Lapaire, O., Schlembach, D., Zeisler, H., Calda, P., Sabria, J., Markfeld-Erol, F., Galindo, A., Schoofs, K., Denk, B., Stepan, H., 2014. New Gestational Phase–Specific Cutoff Values for the Use of the Soluble fms-Like Tyrosine Kinase-1/Placental Growth Factor Ratio as a Diagnostic Test for

Preeclampsia. Hypertension 63, 346–352. https://doi.org/10.1161/HYPERTENSIONAHA.113.01787

Wang, R., Pan, W., Jin, L., Li, Y., Geng, Y., Gao, C., Chen, G., Wang, H., Ma, D., Liao, S., 2019. Artificial intelligence in reproductive medicine. Reproduction 158, R139–R154. https://doi.org/10.1530/REP-18-0523

Weis, C.V., Jutzeler, C.R., Borgwardt, K., 2020. Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review. Clin. Microbiol. Infect. 26, 1310–1317. https://doi.org/10.1016/j.cmi.2020.03.014

Welch, B.L., 1947. The generalisation of student's problems when several different population variances are involved. Biometrika 34, 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. Biom. Bull. 1, 80. https://doi.org/10.2307/3001968

Winters, B., Custer, J., Galvagno, S.M., Colantuoni, E., Kapoor, S.G., Lee, H., Goode, V., Robinson, K., Nakhasi, A., Pronovost, P., Newman-Toker, D., 2012. Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. BMJ Qual. Saf. 21, 894–902. https://doi.org/10.1136/bmjqs-2012-000803

World Health Organization, 2021. 2020 antibacterial agents in clinical and preclinical development: an overview and analysis. World Health Organization, Geneva.

World Health Organization, n.d. Hypertension [WWW Document]. URL https://www.who.int/news-room/fact-sheets/detail/hypertension (accessed 7.3.22).

Wright, D., Akolekar, R., Syngelaki, A., Poon, L.C.Y., Nicolaides, K.H., 2012. A Competing Risks Model in Early Screening for Preeclampsia. Fetal Diagn. Ther. 32, 171–178. https://doi.org/10.1159/000338470

Yasaka, K., Akai, H., Abe, O., Kiryu, S., 2018a. Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study. Radiology 286, 887–896. https://doi.org/10.1148/radiol.2017170706

Yasaka, K., Akai, H., Kunimatsu, A., Abe, O., Kiryu, S., 2018b. Deep learning for staging liver fibrosis on CT: a pilot study. Eur. Radiol. 28, 4578–4585. https://doi.org/10.1007/s00330-018-5499-7

Zhou, L.-Q., Wang, J.-Y., Yu, S.-Y., Wu, G.-G., Wei, Q., Deng, Y.-B., Wu, X.-L., Cui, X.-W., Dietrich, C.F., 2019. Artificial intelligence in medical imaging of the liver. World J. Gastroenterol. 25, 672–682. https://doi.org/10.3748/wjg.v25.i6.672

Statutory Declaration

“I, Leon Schmidt, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic “Anwendung von Machine-Learning in medizinischer Diagnostik in der Geburtshilfe / Applications of machine-learning in medical diagnostics in obstetrics”, independently and without the support of third parties, and that I used no other sources and aids than those stated.
All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

22.03.2023

Date Signature

Declaration of your own contribution to the publications

Leon Schmidt contributed the following to the below listed publications:

Publication 1: Leon J. Schmidt, Oliver Rieger, BSc, Mark Neznansky, MSc, Max Hackelöer, MD, Lisa A. Dröge, MD, Wolfgang Henrich, MD, PhD, David Higgins, PhD, Stefan Verlohren, MD, PhD,
A machine-learning–based algorithm improves prediction of preeclampsia-associated adverse outcomes, American Journal of Obstetrics & Gynecology, 2022

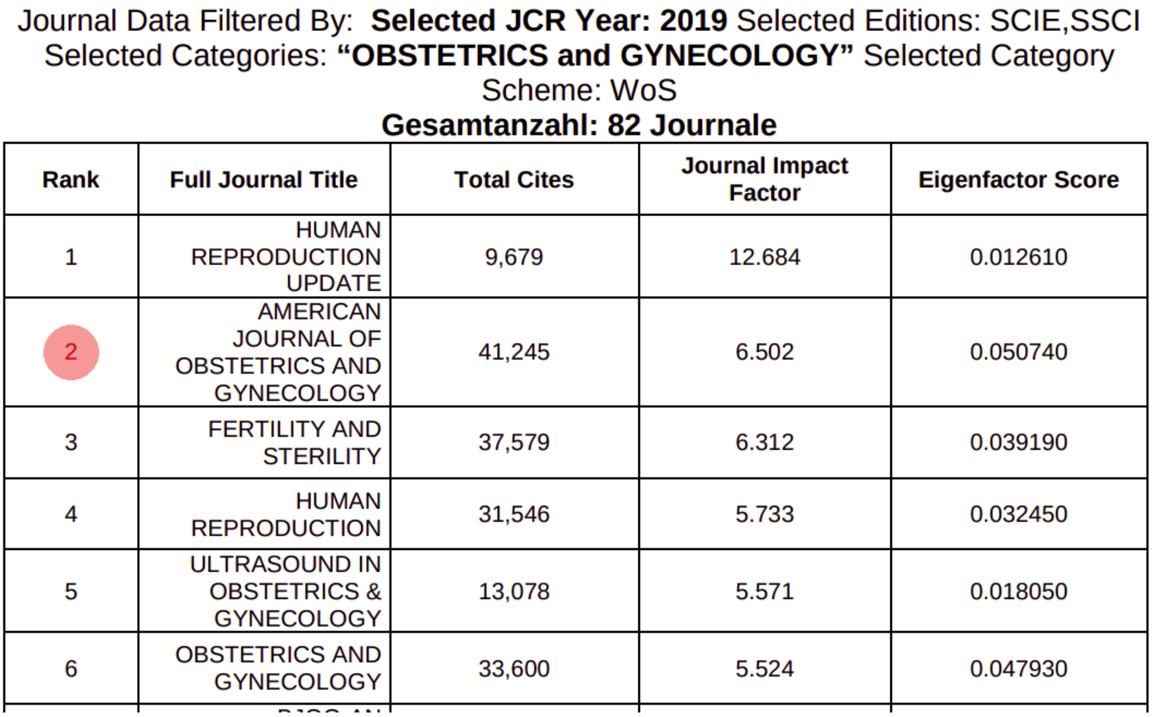
Contribution:

- Writing Draft & Finished paper
- All tables and images in the publication
- Choice of algorithms
- Data cleaning & pre-processing (entire process described)
- Programming & Implementation of hyperparameter tuning, preprocessing & training routines & statistical evaluation
- Result analysis

Signature, date and stamp of first supervising university professor / lecturer

Signature of doctoral candidate

Excerpt from Journal Summary List



Journal Data Filtered By: **Selected JCR Year: 2019** Selected Editions: SCIE,SSCI Selected Categories: **“OBSTETRICS and GYNECOLOGY”** Selected Category Scheme: WoS

Gesamtanzahl: 82 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	HUMAN REPRODUCTION UPDATE	9,679	12.684	0.012610
2	AMERICAN JOURNAL OF OBSTETRICS AND GYNECOLOGY	41,245	6.502	0.050740
3	FERTILITY AND STERILITY	37,579	6.312	0.039190
4	HUMAN REPRODUCTION	31,546	5.733	0.032450
5	ULTRASOUND IN OBSTETRICS & GYNECOLOGY	13,078	5.571	0.018050
6	OBSTETRICS AND GYNECOLOGY	33,600	5.524	0.047930

Printing copy of the publication

Schmidt, L. J., Rieger, O., Neznansky, M., Hackelöer, M., Dröge, L. A., Henrich, W., Higgins, D., & Verlohren, S. (2022). A machine-learning–based algorithm improves prediction of preeclampsia-associated adverse outcomes. In American Journal of Obstetrics and Gynecology (Vol. 227, Issue 1, p. 77.e1-77.e30). Elsevier BV. https://doi.org/10.1016/j.ajog.2022.01.026

Curriculum Vitae

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Publication list

Schmidt LJ, Rieger O, Neznansky M, Hackelöer M, Dröge LA, Henrich W, Higgins D, Verlohren S. A machine-learning-based algorithm improves prediction of preeclampsia-associated adverse outcomes. Am J Obstet Gynecol. 2022 Jul;227(1):77.e1-77.e30. doi: 10.1016/j.ajog.2022.01.026. Epub 2022 Feb 1. PMID: 35114187.

Hackelöer, M., Schmidt, L. & Verlohren, S. New advances in prediction and surveillance of preeclampsia: role of machine learning approaches and remote monitoring. *Arch Gynecol Obstet* (2022). https://doi.org/10.1007/s00404-022-06864-y

Acknowledgments

First and foremost I want to thank Prof. Stefan Verlohren for his guidance and advice throughout all stages of this work. Thank you for providing me with the opportunity to write this dissertation and to always support me in any way possible be it mentoring, writing advice, organisational support or simply guiding me in my studies. Without you none of this would have been possible, you created this.
It was a pleasure working with you and I am deeply grateful for everything you have done for me. Thank you!

Second I want to thank Mr. David Higgins - without your mentoring on machine-learning methods and educational support this work would not have come as easily as it did, not to say it would have been straight-up impossible. I learned so much from you and literally every conversation we had was a perfect combination of enjoyable and informative. I hope our paths will cross again in the future!

Third I want to thank my co-authors, especially Mr. Oliver Rieger and Mr. Mark Neznansky - you guys made it so easy doing even the most menial data cleaning tasks and I could always count on your advice regarding methodology and comradery when things did not go as planned. It was a pleasure working with you!

I also want to thank Stephanie Kühne, you supported me at every step and pushed me to do my very best, thank you so much.

Last but not least I want to thank my parents, Mr Volker Schmidt and Mrs. Marion Thill-Schmidt - thank you for your support, this is as much your achievement as it is mine.