Pediatric RESEARCH

CLINICAL RESEARCH ARTICLE

Enhanced early prediction of clinically relevant neonatal hyperbilirubinemia with machine learning

Imant Daunhawer[1], Severin Kasser[2], Gilbert Koch[3], Lea Sieber[2], Hatice Cakal[2], Janina Tütsch[2], Marc Pfister[3], Sven Wellmann[2,4] and Julia E. Vogt[1,5]

BACKGROUND: Machine learning models may enhance the early detection of clinically relevant hyperbilirubinemia based on patient information available in every hospital.

METHODS: We conducted a longitudinal study on preterm and term born neonates with serial measurements of total serum bilirubin in the first two weeks of life. An ensemble, that combines a logistic regression with a random forest classifier, was trained to discriminate between the two classes phototherapy treatment vs. no treatment.

RESULTS: Of 362 neonates included in this study, 98 had a phototherapy treatment, which our model was able to predict up to 48 h in advance with an area under the ROC-curve of 95.20%. From a set of 44 variables, including potential laboratory and clinical confounders, a subset of just four (bilirubin, weight, gestational age, hours since birth) suffices for a strong predictive performance. The resulting early phototherapy prediction tool (EPPT) is provided as an open web application.

CONCLUSION: Early detection of clinically relevant hyperbilirubinemia can be enhanced by the application of machine learning. Existing guidelines can be further improved to optimize timing of bilirubin measurements to avoid toxic hyperbilirubinemia in high-risk patients while minimizing unneeded measurements in neonates who are at low risk.

Pediatric Research (2019) 86:122–127; https://doi.org/10.1038/s41390-019-0384-x

INTRODUCTION

Neonatal jaundice due to hyperbilirubinemia is the most common pathology in neonates and one of the major reasons for a hospitalization in the first year of life. Almost 10% of newborn infants develop significant hyperbilirubinemia, defined as a bilirubin level above the 95th percentile at a given age in hours[1,2] and a substantial amount require phototherapy treatment.[3] Costs for national health care systems are accordingly high,[4] particularly in countries with high rates of Glucose-6-phosphate dehydrogenase (G6PD) deficiency, an established risk factor for neonatal jaundice.[5] Not treated properly, neonatal jaundice can cause major disability with life-long sequelae.[6,7] On the other hand, phototherapy treatments increase the likelihood for allergic diseases in childhood as population-based studies have shown.[8–12] Thus, in the context of neonatal jaundice, both precise patient monitoring, as well as deliberate treatment assignment are required for neonates who are at high risk of developing significant hyperbilirubinemia.

In 1999, Bhutani et al.[1] introduced nomograms for the assessment of neonatal hyperbilirubinemia; nomograms are based on percentiles of bilirubin values at a given age in hours and they are still widely used to classify neonates into risk groups. More recent risk stratification approaches include additional clinical factors for the prediction of neonatal hyperbilirubinemia shortly after birth,[13] before discharge,[14] or following inpatient phototherapy.[15] Even though approaches for risk stratification provide clinicians with a guideline for their assessment, these methods are overly general, as they do not consider whether or not an individual actually received a phototherapy treatment—a decision that often depends on further practical considerations. For instance, a clinician's assessment of visible jaundice or of the medical history of a neonate are important factors that flow into the treatment decision; therefore, we define a case of clinically relevant hyperbilirubinemia if a phototherapy treatment was delivered. We aim for a more personalized prediction by identifying neonates at risk for clinically relevant hyperbilirubinemia more accurately, thus preventing the development of severe neonatal jaundice as well as overtreatment and unnecessary hospital stays.

Several studies have addressed the early detection of hyperbilirubinemia. Huang et al.[16] applied a logistic regression analysis for neonates with at least 35 weeks of gestational age (GA) and exclusive breast-feeding to predict subsequent hyperbilirubinemia using GA, maximal body weight loss percentage, and peak bilirubin level during the first 72 h of life; their model achieved an AUC of 78.8%, setting a baseline for further studies. Ferreira et al.[17] demonstrated that state-of-the-art methods can be used to improve the early detection of neonatal hyperbilirubinemia: they compared different classification algorithms based on a dataset of healthy term and near-term neonates of which 15.4% received

[1]Adaptive Systems and Medical Data Science, Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland; [2]Division of Neonatology, University of Basel Children's Hospital (UKBB), Basel, Switzerland; [3]Division of Paediatric Pharmacology and Pharmacometrics, University of Basel Children's Hospital (UKBB), Basel, Switzerland; [4]Division of Neonatology, University Children's Hospital Regensburg (KUNO), University of Regensburg, Regensburg, Germany and [5]SIB Swiss Institute of Bioinformatics, Basel, Switzerland

Correspondence: Sven Wellmann (sven.wellmann@ukbb.ch)

Shared first authors: Imant Daunhawer and Severin Kasser.

Shared last authors: Sven Wellmann and Julia E. Vogt.

Received: 19 November 2018 Revised: 18 February 2019 Accepted: 26 February 2019
Published online: 31 March 2019

a phototherapy; their best model, a logistic regression that was based on 60 variables, achieved an AUC of 89%. The limitations of their model are the impractical requirement of 60 variables and the restriction that a phototherapy prediction is made at 24 h after birth for all neonates. More recently, Castillo el al.[13] developed a model for the early detection of hyperbilirubinemia in healthy term and near-term neonates younger than 24 h. Using umbilical cord bilirubin, GA, and maternal race as predictors, their model, a regularized logistic regression (LASSO), achieves a performance of 89% (±3%) AUC. The limitation of their model is that it provides only a single prediction within the first 24 h after birth, but in clinical practice one requires a prediction with every subsequent bilirubin measurement, most importantly before discharge. Thus, even though the early detection of hyperbilirubinemia has been studied, existing models are not applicable in most practical contexts, because the models can either not be applied after every new bilirubin measurement, or because they come with too much overhead (e.g., requiring too many variables).

In this study, machine learning (ML) is applied to enhance the early detection of clinically relevant hyperbilirubinemia in advance of the first phototherapy treatment. ML shows great potential in clinical applications[18,19] and in pediatrics it has been successfully applied for an improved early detection of late-onset neonatal sepsis based on medical records,[20] and of neonatal seizures based on EEG data.[21] Hence, the goal is to draw on the predictive power of state-of-the-art ML methods to provide an early identification of neonates at risk of developing clinically relevant hyperbilirubinemia, and thereby enhance the timing of bilirubin measurements and of phototherapy treatment initiation in practice.

METHODS

Study patients

We performed a retrospective study of prospectively recorded neonatal data of all neonates admitted to the University Children's Hospital Basel (UKBB) in 2015 and 2016 within the first week of life and with a gestational age of more than 31 completed weeks at birth. A second dataset was prepared for external blinded validation based on neonates admitted to the UKBB in 2017 within the first week of life, with a gestational age of more than 34 completed weeks at birth and a birthweight of at least 2500 g. The study was approved by the Institutional Review Board (EKNZ: BASEC 2018-00053). A priori exclusion criteria were: fewer than two bilirubin measurements, major malformations requiring operation within the first month of life, or the presence of any genetic syndrome

The data contained 44 variables, described in the following. Birthweight and all subsequent weight values during hospitalization; delivery mode; sex; gestational age at birth; Apgar values at 1, 5, and 10 min; arterial umbilical cord pH at birth; repeated measurements of hemoglobin and hematocrit together with total serum bilirubin, including sodium values; daily weight measurements, enteral and parenteral feeding quantities, and type of feeding (mother milk, formula milk, or both); maternal and neonatal Rh and blood group and Coombs test; ethnicity; maternal age at birth, parity, and maternal comorbidities at birth including gestational diabetes, chorioamnionitis, and preeclampsia; and presence of neonatal infection and therapy with antibiotics, respiratory morbidity including data on respiratory support, intensity and duration, and other comorbidities. From the 38 variables described above, another 6 variables were derived: ratio between bilirubin and weight, count of previous bilirubin measurements, value of and time since the previous bilirubin measurement (if available), relative weight change, and an indicator for preterm neonates (i.e., GA < 35). On phototherapies, the data contained the onset and duration of each treatment as well as the phototherapy initiation limit that was used. All bilirubin measurements were performed as total bilirubin using an ABL800 FLEX blood gas analyser (Radiometer Medical ApS, Denmark).

Data preprocessing

Our goal is to make a prediction after each bilirubin measurement, since, in practice, one would like to assess a neonate's risk of developing excessive bilirubin levels right after a new bilirubin measurement is available. We argue that prior to the first measurement there is insufficient information available to provide a reasonable individual prediction. Moreover, a uniform prediction time (e.g., making a prediction at 24 h after birth for all neonates) is less practical from a clinical point of view than a prediction that can be coupled with a bilirubin measurement.

Independent variables

The dataset offers a very rich collection of information about the early state of neonates. We initially used all 44 variables—including potential laboratory and clinical confounders—for the predictive model, but noticed that a small subset was sufficient for a strong predictive performance. In particular, a backward variable selection (Figure S1 in the appendix) was employed where in each iteration the model was re-trained, evaluated, and the least influential variable was removed. As a result, the set of independent variables could be reduced to a small set of variables that should be available in every hospital: gestational age, weight, bilirubin level, and hours since birth.

Dependent variable

Given a neonate's gestational age, weight, bilirubin level, and time since birth, the model predicts the probability of the neonate receiving a phototherapy treatment within the next 48 h. The dependent variable (target) of the predictive model is a binary indicator that takes on a value of one if a neonate received a phototherapy in the 48 h following a bilirubin measurement and zero otherwise. In this definition, there are two aspects that require further explanation: first, the decision process that determines whether a neonate receives a phototherapy, and second, why we chose a prediction interval of 48 h.

At the UKBB, a phototherapy treatment is administered if the bilirubin level (μmol/L) exceeds a threshold that depends on a neonate's gestational age, birthweight, time since birth, and additional neurotoxicity risk factors. This process is based on the widely adapted AAP guidelines for phototherapy initiation[22] (which derive from the Bhutani nomogram[1]), but is slightly modified to account for preterm neonates by considering birthweight in addition to the other listed factors. The complete UKBB guidelines are summarized in Table S1 in the appendix. For term neonates Fig. 1 illustrates the similarity between the UKBB and AAP guidelines. In general, when a neonate receives a phototherapy in the UKBB, in most cases it would also receive a phototherapy according to the AAP guidelines. In the discussion it is explained in further detail how the model can be applied with other phototherapy initiation guidelines.

With respect to the prediction interval, we chose a window of 48 h, because from a clinical perspective it can be assumed that a second bilirubin measurement occurs within this timeframe, as long as the neonate is not released from hospital. In the data, almost 95% of all consecutive bilirubin measurements are separated by no more than 48 h (Table 1) for patients who received a phototherapy.

Machine learning

To improve the early detection of clinically relevant hyperbilirubinemia we applied both conventional as well as state-of-the-art ML methods—algorithms that can "learn" from data to improve their performance with respect to a particular task, which in this case is the prediction for the need of a phototherapy treatment.

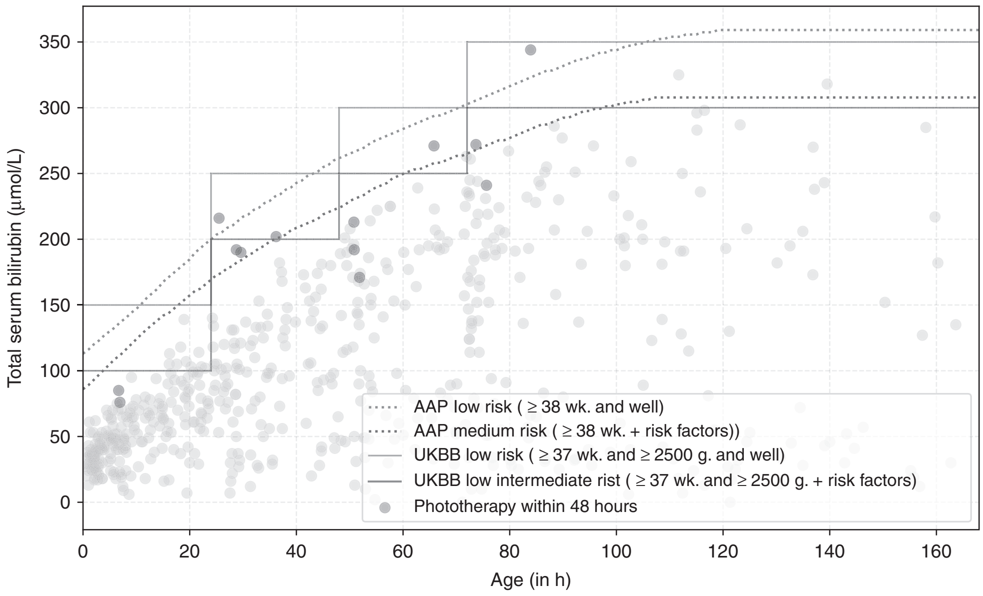

Fig. 1 Illustration of the similarity between the UKBB and AAP guidelines for term neonates. The points represent bilirubin measurements at the corresponding time since birth for all term neonates in the data. Bilirubin measurements that were followed by a phototherapy within the next 48 h are highlighted in red. In contrast to the AAP guidelines, the UKBB uses a stepwise function that approximates the AAP guidelines and in addition takes into account the birthweight of neonates. The complete UKBB guidelines are listed in the appendix (Table S1)

Table 1. Statistics about the distribution of bilirubin measurements across neonates

	All ($n = 362$)	Phototherapy only ($n = 98$)
Average number of measurements	4.26	2.34
Average time between measurements	26.82	22.13
95% quantile of time between measurements	72.43	50.46
Average time of the first bilirubin measurement	16.56	25.05

Time is measured in hours; the time of the first measurement is relative to the birthtime

In particular, we compare a regularized logistic regression—a conventional tool in the arsenal of ML methods—with a random forest classifier[23]—a state-of-the-art method that performs particularly well on tabular data and does not require huge amounts of data (compared to, for example, artificial neural networks).

The logistic regression is a gold-standard method for modeling a binary dependent variable, and an additional L1-regularization term allows for a convenient way of variable selection by shrinking regression coefficients towards zero; the resulting model is called the LASSO.[24] A random forest is an ensemble of decision trees, each of which learns to discriminate between two classes (phototherapy vs. no phototherapy) based on a different sample from the data. A tree is iteratively built by branching on the variable and threshold that separates the classes best. The separation of classes is measured by Gini impurity. For the task of prediction, a random forest decides on the class that wins the majority vote among the large number of built trees.

ML methods usually require the configuration of hyperparameters—knobs that can be turned to adjust the learner to the problem at hand. Hyperparameters are specific to the method of choice: for the LASSO we have the regularization weight, while for the random forest there are mainly the number of estimators (i.e., the number of trees in a forest), the maximum depth of each tree, and the maximum number of variables that are considered in each split. With the choice of hyperparameter values it is possible to control the tradeoff between bias and variance, i.e., the complexity of a model versus its generalizability. Usually, one seeks for a model that generalizes as well as possible (strong and stable performance on previously unseen data) and at the same time is as simple as possible (low complexity).

The best hyperparameter setting found for the LASSO has an inverse regularization weight of 0.3, and for the random forest we used 300 trees, Gini impurity as a split criterion, no depth limit for individual trees, at least two samples per split, at least one sample per leaf, and random sampling of $\sqrt{k}$ features in each tree, where k is the total number of features in the model.

All analyses were implemented in the Python3 programming language using models from the scikit-learn library.[25] The runtime for the models is very fast: for the whole time series of measurements, training time lies in the range of a few seconds, and new predictions take only a few milliseconds to compute.

Evaluation

First, the data were divided into a training and a test set using stratified sampling to keep the balance between observations with and without a phototherapy treatment. 70% of the data was used for training and the remaining 30% was left untouched for the final evaluation. It is important to note that we stratified by patient, so that observations from the same patient are not mingled across both training and test data.

On the training data, a 3-fold cross-validation was performed to select the hyperparameter values of the model; that means the training set is further divided into three disjoint chunks of approximately equal size. With a given hyperparameter setting, the model was trained on two chunks and evaluated on the remaining chunk—a procedure that was repeated three times using different combinations of chunks; this technique provides an estimate for the average performance and its variance across three partitions of the training data. Only then the performance of the model—using the hyperparameter values that had shown the best cross-validation performance—was validated on the test data.

Performance was measured as AUC, i.e., the area under the receiver operator characteristic (ROC) curve, which evaluates a model based on the true positive rate versus false positive rate of its predictions at different threshold values. It is a default metric for the evaluation of binary classifiers and it has been used in related studies,[13,17,26] thus allowing us to draw a direct comparison to previous work.

RESULTS

Of the 385 neonates, we excluded 23 who received a phototherapy prior to their first bilirubin measurement. Of the remaining 362 neonates, 98 (27.07%) were subject to at least one phototherapy treatment during their initial hospitalization. All first phototherapies occur within the first week of life, on average 66 h after birth.

Table 1 summarizes the distribution of bilirubin measurements across neonates. In total, 1543 bilirubin measurements were available—more than four measurements per neonate on average. The average time between bilirubin measurements is less than 27 h, and among those who received a phototherapy, 95% of consecutive measurements were no more than 50 h apart.

Table 2 shows descriptive statistics of a selection of variables that capture key characteristics of the sample. Notably, there is a relatively large number of preterm neonates: 54.1% of the sample have a gestational age (GA) of less than 37 weeks and 41.4% have a GA of less than 35 weeks. A correlation of 0.374 was found between phototherapy and a GA of less than 37 weeks.

Variable selection

Out of all 44 variables, it was observed that only a small subset suffices to achieve a high predictive performance. Figure 2 illustrates which variables were most important for the prediction, where importance was estimated by their relevance (i.e., feature importance) in the random forest. We further performed a backward variable selection (Figure S1 in the appendix) which found that only four inputs suffice for a strong predictive performance: GA, weight, bilirubin level, and hours since birth. Notably, the ratio between bilirubin and weight was most important for the phototherapy prediction, though it was closely followed by weight and bilirubin level as individual variables, indicating that there is a more complex underlying relationship than a mere ratio. The LASSO identifies a similar set of variables to be most important for the prediction (see Figure S2 in the appendix), though with a different order of importance, suggesting that the models are sensitive to different factors and that they may complement each other when used in conjunction.

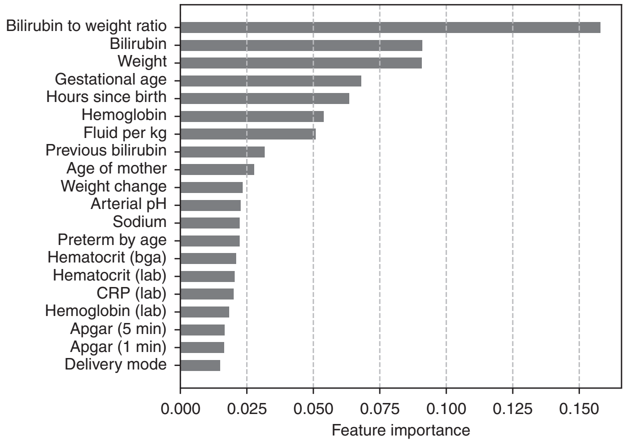

Fig. 2 A subset of 20 variables with highest feature importance. The bar chart indicates that the predictive performance of the random forest depends mostly on only a few variables with highest feature importance. The order of variables remains relatively stable during backward variable selection (Figure S1 in the appendix)

Quantitative performance

Figure 3 compares the test performance of three models: the LASSO, the random forest, and an ensemble that takes a simple average of the predictions from the former two models. The ensemble (AUC of 0.952 ± 0.013) achieves the best overall performance on the test set and has the lowest variance in the cross-validation. The second best performing model is the LASSO (AUC of 0.947 ± 0.015), followed by the random forest (AUC of 0.933 ± 0.019). The performance of each model on the holdout set is reasonably close to the respective performance in the cross-validation (see Table S2 in the appendix), suggesting that the models did not overfit to the training data, but that they generalized well to unseen data.

Additionally, the robustness of the models was checked in a series of ablation experiments. It was verified that the models perform well even if we consider only the first bilirubin measurement (AUC of 0.939 for the ensemble as well as for the LASSO, and 0.927 for the random forest). Further, the backward variable selection (Figure S1 in the appendix) showed that the model performance degrades if fewer features than the five variables described previously are used. Finally, it was tested how the models perform with less training data (Table S2 in the appendix), finding that the ensemble is more stable than the individual models when the amount of data is limited.

Surprisingly, the LASSO, being a more conventional approach, still performed slightly better than the random forest. Further investigation (Table S2 in the appendix) suggested that with more

Table 2. Descriptive statistics of the sample

	5% Quantile	50% Quantile	95% Quantile	Percentage
Gestational age (in days)	228	253	289	—
Birthweight (in grams)	1490	2605	4120	—
Age of mother (in years)	23	32	40	—
First bilirubin level (in μmol/L)	19	46	216	—
Gender (male)	—	—	—	56.91
Preterm[a] birth	—	—	—	54.13
Multiple birth	—	—	—	23.48

Quantiles are computed for continuous or integer-valued variables; percentages for dummy variables
[a] Defined as a gestational age of less than 37 weeks

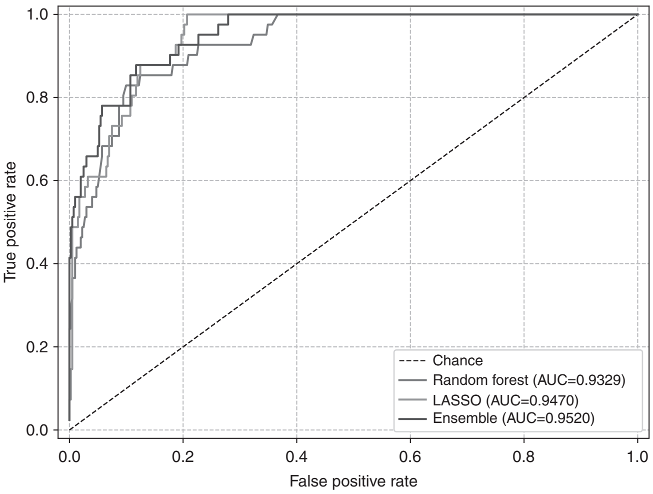

Fig. 3 Evaluation of the models on previously unseen data. The performance of an individual model is shown by the respective ROC-curve and it is summarized as a single value (AUC) in the respective label

data and a richer set of variables, the random forest would perform at least as good as the LASSO, however, in the case of limited data, conventional approaches remain a viable alternative to state-of-the-art methods. Moreover, both approaches can be combined to improve the performance over individual models.

To translate the predicted probabilities into concrete decisions, it requires a decision threshold, i.e., a value above which a predicted probability is interpreted as a positive prediction. We computed the decision threshold as the value at which the F1-score (i.e., the tradeoff between precision and recall) is maximized in the cross-validation. A decision threshold of 0.38 was chosen, based on which the model achieved a sensitivity of 0.781 and a specificity of 0.920 on the test data. Note that the decision threshold can be chosen differently based on practical considerations, especially if one could assign a cost to false positives and false negatives, respectively.

Online tool

Based on the best performing model (i.e., the ensemble) we have implemented an online tool, the early phototherapy prediction tool (EPPT), through which clinicians were able to experiment directly with the model and provided feedback on the quality of the predictions. The tool can be accessed online at https://ppt.dmi.unibas.ch.

External blinded validation

The EPPT was further validated in an experiment for which a new time series of bilirubin measurements ($n = 187$), based on $n = 57$ patients admitted in 2017 to the UKBB and fulfilling the above mentioned criteria were used. This experiment was designed to test the performance of the EPPT in a realistic setting, where the model has no potential access to information about upcoming phototherapies before it computes the predictions and in which only the four values required for the EPPT were provided.

Overall, the EPPT retained a strong predictive performance (AUC = 0.954), which is slightly better than its performance on the holdout set. Based on the decision threshold from above, the early detection was successful in 5 out of 8 phototherapy cases, for which a phototherapy was predicted on average 23 h before the bilirubin level exceeded the critical limit of the phototherapy initiation guidelines. A single case, whose bilirubin value was still 32 μmol/L below the phototherapy threshold of 300 μmol/L, turned out to be a false positive.

Overall, the blinded validation confirmed in a realistic setting that neonates at risk can be detected well in advance, which leaves time for decisions and actions, such as closer monitoring or later discharge.

DISCUSSION

We here present empirical evidence for an enhanced early assessment of clinically relevant hyperbilirubinemia by leveraging ML methods. For this, we provided a computational model for early phototherapy prediction by applying an ensemble—consisting of a random forest and a logistic regression—which can detect with high AUC, sensitivity and specificity whether a neonate will receive a phototherapy within the next 48 h after a bilirubin measurement. The model was trained on a large dataset containing several dozen clinical and laboratory variables of which finally just four inputs suffice for a strong predictive performance. Finally, the model was provided as an open web application—the early phototherapy prediction tool—and further validated in an external blinded validation.

Although prediction is usually the main goal of ML, this study also shed light on the factors that influence the development of neonatal jaundice. It was found that bilirubin, weight, gestational age, and hours since birth are the most significant predictors in the model; in particular, it appears that there is a complex relationship between these factors, which is an interesting starting point for the analysis of bilirubin kinetics.

Compared to previous research,[13,16,17] the EPPT achieves a significant performance improvement of 6 percentage-points AUC. In addition, the EPPT has the following advantages: it can be applied after every new bilirubin measurement (not only at a predefined point in time), even if only a single measurement is provided, and requires only a small set of variables that are available in most hospitals. Consequently, the resulting model is more generally applicable in a clinical context.

In the following, we discuss the limitations of the model and point to opportunities for further research. First, the predictions were limited to a 48-hour interval, which seems adequate in a clinical context, but can be improved nevertheless. Secondly, the outcome variable was based on the phototherapy initiation guidelines of the UKBB, which are quite similar to the widely adapted AAP guidelines (as illustrated in Fig. 1), but a more generally applicable model could be trained on a dataset that follows the AAP guidelines more precisely. For institutions that use substantially different phototherapy initiation guidelines, the model would first need to be re-trained with data that contains the same set of variables (GA, weight, hours since birth, bilirubin level, phototherapy within the next 48 h).

Finally, before integrating the EPPT into clinical decision-making, a clinical study is needed to test the sensitivity and specificity prospectively, and the model should be validated before it is applied in populations that differ substantially from the study population, for example, in populations with a high prevalence of G6DH-deficiency[27] or in extreme preterm cases.

Apart from external validation on diverse populations, interesting opportunities for further research include the extension of the model to different phototherapy initiation guidelines (in this context, multi-task learning[28] might be a promising approach), or the prediction of not only phototherapy initiation, but also its duration and intensity.

CONCLUSION

Undetected and untreated hyperbilirubinemia in the neonatal period can cause major and life-long disability. Our developed EPPT can support caregivers in bilirubin surveillance to identify high-risk neonates up to 48 h in advance to the onset of clinically relevant hyperbilirubinemia. The model, on which the tool is based, has demonstrated high sensitivity and specificity for the early detection of hyperbilirubinemia. Its external validity was further confirmed in a blinded validation.

More generally, we demonstrated how a personalized medicine approach, based on state-of-the-art ML methods, allows researchers to create decision-support systems for early detection based on historical information about treatment decisions.

ACKNOWLEDGEMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS

Substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data: All authors. Drafting the article or revising it critically for important intellectual content: I.D., S.K., G.K., M.P., S.W. and J.E.V. Final approval of the version to be published: I.D., S.K., G.K., M.P., S.W. and J.E.V.

ADDITIONAL INFORMATION

The online version of this article (https://doi.org/10.1038/s41390-019-0384-x) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Bhutani, V. K., Johnson, L. & Sivieri, E. M. Predictive ability of a predischarge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near-term newborns. *Pediatrics* **103**, 6–14 (1999).
2. Stevenson, D. K. et al. Prediction of hyperbilirubinemia in near-term and term infants. *Pediatrics* **108**, 31–39 (2001).
3. Olusanya, B. O., Kaplan, M. & Hansen, T. W. R. Neonatal hyperbilirubinaemia: a global perspective. *Lancet Child Adolesc. Health* **2**, 610–620 (2018).
4. Burgos, A. E., Schmitt, S. K., Stevenson, D. K. & Phibbs, C. S. Readmission for neonatal jaundice in California, 1991–2000: trends and implications. *Pediatrics* **121**, e864–e869 (2008).
5. Olusanya, B. O., Emokpae, A. A., Zamora, T. G. & Slusher, T. M. Addressing the burden of neonatal hyperbilirubinaemia in countries with significant glucose-6-phosphate dehydrogenase deficiency. *Acta Paediatrica* **103**, 1102–1109 (2014).
6. Johnson, L. H., Bhutani, V. K. & Brown, A. K. System-based approach to management of neonatal jaundice and prevention of kernicterus. *J. Pediatr.* **140**, 396–403 (2002).
7. Dennery, P. A., Seidman, D. S. & Stevenson, D. K. Neonatal hyperbilirubinemia. *New Engl. J. Med.* **344**, 581–590 (2001).
8. Wei, C. C., Lin, C. L., Shen, T. C. & Kao, C. H. Neonatal jaundice and risks of childhood allergic diseases: a population-based cohort study. *Pediatr Res.* **78**, 223–230 (2015).
9. Aspberg, S., Dahlquist, G., Kahan, T. & Kallen, B. Confirmed association between neonatal phototherapy or neonatal icterus and risk of childhood asthma. *Pediatr. Allergy Immunol.* **21**, e733–e739 (2010).
10. Ku, M. S. et al. Neonatal jaundice is a risk factor for childhood asthma: a retrospective cohort study. *Pediatr Allergy Immunol.* **23**, 623–628 (2012).
11. Kuzniewicz M. W., Niki H., Walsh E. M., McCulloch C. E., Newman T. B. Hyperbilirubinemia, Phototherapy, and Childhood Asthma. *Pediatrics* 2018;142.
12. Newman T. B., Wu Y. W., Kuzniewicz M. W., Grimes B. A., McCulloch C. E. Childhood Seizures After Phototherapy. *Pediatrics* 2018;142.
13. Castillo, A. et al. Umbilical cord blood bilirubins, gestational age, and maternal race predict neonatal hyperbilirubinemia. *PLoS ONE* **13**, e0197888 (2018).
14. Han, S. et al. A model for predicting significant hyperbilirubinemia in neonates from China. *Pediatrics* **136**, e896–e905 (2015).
15. Chang P. W., Kuzniewicz M. W., McCulloch C. E., Newman T. B. A Clinical prediction rule for rebound hyperbilirubinemia following inpatient phototherapy. *Pediatrics* 2017;139.
16. Huang, H. C. et al. Model to predict hyperbilirubinemia in healthy term and near-term newborns with exclusive breast feeding. *Pediatr. Neonatol* **53**, 354–358 (2012).
17. Ferreira, D., Oliveira, A. & Freitas, A. Applying data mining techniques to improve diagnosis in neonatal jaundice. *BMC Med. Inform. Decis. Mak.* **12**, 143 (2012).
18. Obermeyer, Z. & Emanuel, E. J. Predicting the future—Big Data, machine learning, and clinical medicine. *New Engl. J. Med.* **375**, 1216–1219 (2016).
19. Chen, J. H. & Asch, S. M. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *New Engl. J. Med.* **376**, 2507–2509 (2017).
20. Mani, S. et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *J. Am. Med. Inform. Assoc.* **21**, 326–336 (2014).
21. Temko, A., Thomas, E., Marnane, W., Lightbody, G. & Boylan, G. EEG-based neonatal seizure detection with Support Vector Machines. *Clin. Neurophysiol.* **122**, 464–473 (2011).
22. American Academy of Pediatrics Subcommittee on H. Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics* **114**, 297–316 (2004).
23. Breiman, L. Random forests. *Mach Learn.* **45**, 5–32 (2001).
24. Tibshirani R. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc.* **58,** 267-288 (1996).
25. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Machi. Learn. Res.* **12**, 2825–2830 (2011).
26. Newman, T. B., Liljestrand, P. & Escobar, G. J. Combining clinical risk factors with serum bilirubin levels to predict hyperbilirubinemia in newborns. *Arch. Pediatri. Adolesc. Med.* **159**, 113–119 (2005).
27. Cappellini, M. D. & Fiorelli, G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**, 64–74 (2008).
28. Caruana, R. Multitask learning. *Mach. Learn.* **28**, 41–75 (1997).