Machine learning improves early prediction of small-for-gestational-age births and reveals nuchal fold thickness as unexpected predictor

Shier Nee Saw\textsuperscript{1,2} | Arijit Biswas\textsuperscript{3} | Citra Nurfarah Zaini Mattar\textsuperscript{3} | Hwee Kuan Lee\textsuperscript{1,4,5,6,7} | Choon Hwai Yap\textsuperscript{8} ∗

\textsuperscript{1}Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore
\textsuperscript{2}Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
\textsuperscript{3}Department of Obstetrics and Gynecology, Yong Loo Lin School of Medicine, National University of Singapore, National University Health Systems, Singapore
\textsuperscript{4}Rehabilitation Research Institute of Singapore, Singapore
\textsuperscript{5}School of Computing, National University of Singapore, Singapore
\textsuperscript{6}Image and Pervasive Access Lab (IPAL), CNRS UMI, Singapore
\textsuperscript{7}Singapore Eye Research Institute, Singapore
\textsuperscript{8}Department of Bioengineering, Imperial College London, London, UK

Abstract

Objective: To investigate the performance of the machine learning (ML) model in predicting small-for-gestational-age (SGA) at birth, using second-trimester data.

Methods: Retrospective data of 347 patients, consisting of maternal demographics and ultrasound parameters collected between the 20th and 25th gestational weeks, were studied. ML models were applied to different combinations of the parameters to predict SGA and severe SGA at birth (defined as 10th and third centile birth weight).

Results: Using second-trimester measurements, ML models achieved an accuracy of 70% and 73% in predicting SGA and severe SGA whereas clinical guidelines had accuracies of 64% and 48%. Uterine PI (Ut PI) was found to be an important predictor, corroborating with existing literature, but surprisingly, so was nuchal fold thickness (NF). Logistic regression showed that Ut PI and NF were significant predictors and statistical comparisons showed that these parameters were significantly different in disease. Further, including NF was found to improve ML model performance, and vice versa.

Conclusion: ML could potentially improve the prediction of SGA at birth from second-trimester measurements, and demonstrated reduced NF to be an important predictor. Early prediction of SGA allows closer clinical monitoring, which provides an opportunity to discover any underlying diseases associated with SGA.
Small-for-gestational-age (SGA) newborn is defined as birth weight below the 10th centile. SGA can result from fetal growth restriction, where the fetus fails to achieve its individual growth potential due to placental disorder, maternal medical conditions, genetic factors, and external factors such as smoking. Newborns with birth weights below the third centile experience 5–10 times higher rate of mortality and exhibit greater risk of developing chronic diseases in adult life.

Identifying SGA fetuses before delivery is clinically important because in a population-based study by Gardosi et al., the main risk factor associated with stillbirth is un-identified SGA before birth. Lindqvist and Molin also reported that if SGA fetuses were identified before birth and were given proper antenatal surveillance, their risk of adverse outcomes could be reduced by four-fold. Early prediction of SGA will provide the opportunity to select patients for follow-up, allowing closer monitoring and better clinical management. Furthermore, new pharmacological treatments for SGA are being investigated, and if successful in the future, predicting SGA at an earlier stage would be useful to identify subjects for treatment.

Clinical risk predictive model, such as logistic regression, is commonly found in medical domains because it allows interpretation of model parameters. However, machine learning (ML) can learn data patterns in higher dimensions and has significantly better predictive ability. This explains its popularity in the biomedical domain. In terms of predicting SGA, there have been studies that used ML models. Features used for prediction in these past studies include ultrasound biometrics measurements, umbilical Doppler blood flow, pregnancy risk factors, sociodemographic, maternal characteristic and medical history, pregnancy associated plasma protein A and placental growth factor.

The objective of this study is to evaluate the performance of ML in predicting SGA at birth, using second trimester fetal ultrasound scans. In our study, we show that ML significantly improves the prediction of SGA at birth compared to the prediction via clinical guidelines. Moreover, our ML models unexpectedly reveals that nuchal fold thickness to be an important parameter in predicting SGA at birth.
highest votes across the decision trees. SVM builds a predictive model by mapping example data points in a space where points are divided into separate categories with as wide a gap as possible. The MLP model consists of several layers of interconnected nodes, including an input layer, an output layer, and one or more intermediate hidden layers, and computes its prediction via weights assigned to each node. The weights are optimized during the model training stage with an optimization algorithm.

2.2.1 | Input data to machine learning model

A combination of demographics, fetal biometry and Doppler measurements were used to train the model. We took a set of 15 parameters, which includes fetal measurements that are routinely scanned at the second trimester and maternal demographics. These parameters were reported to have correlation with fetal growth and were used as potential inputs to train the model.

| TABLE 1 Patient characteristic |
|--------------------------------|
|                              | Group A BW > 10th centile | Group B BW between 3rd and 10th centile | Group C BW < 3rd centile |
| Number of samples            | 145                        | 105                                      | 97                       |
| Gestational at birth (weeks) | 38.71 ± 1.59               | 37.99 ± 0.70\(^{**}\)                    | 38.29 ± 1.00\(^{a}\)     |
| Birth weight, BW (kg)        | 3.05 ± 0.49                | 2.40 ± 0.12\(^{**}\)                    | 2.20 ± 0.21\(^{a}\)      |
| Gestational at scan (weeks)  | 21.88 ± 0.57               | 21.87 ± 0.59                            | 21.98 ± 0.78             |
| Mother’s age                 | 31.83 ± 4.17               | 31.49 ± 4.62                            | 31.04 ± 4.85             |

Ethnicity

| Ethnicity | Group A | Group B | Group C |
|-----------|---------|---------|---------|
| Chinese   | 66 (51.6%) | 33 (25.8%) | 29 (22.7%) |
| Malay     | 20 (25.0%) | 27 (33.8%) | 33 (41.3%) |
| Indian    | 32 (37.6%) | 26 (30.6%) | 27 (31.8%) |
| Others    | 27 (50.0%) | 19 (35.2%) | 8 (14.8%)  |

| BPD (mm)  | 52.97 ± 3.28 | 52.05 ± 3.78 | 52.69 ± 3.52 |
| HC (mm)   | 192.84 ± 9.57 | 190.63 ± 9.48 | 191.39 ± 10.95 |
| TCD (mm)  | 22.69 ± 1.19 | 22.26 ± 1.26\(^{**}\) | 22.51 ± 1.48 |
| CM (mm)   | 5.80 ± 1.28  | 5.47 ± 1.26  | 5.74 ± 1.13  |
| EFW (g)   | 447.24 ± 62.33 | 424.84 ± 59.15\(^{**}\) | 431.31 ± 74.38 |
| NF (mm)   | 4.77 ± 0.65  | 4.36 ± 0.76\(^{**}\) | 4.29 ± 0.80\(^{a}\) |
| AC (mm)   | 167.60 ± 10.40 | 163.92 ± 9.61\(^{**}\) | 163.69 ± 11.12\(^{a}\) |
| FL (mm)   | 36.81 ± 2.23  | 36.14 ± 2.25  | 36.49 ± 2.82  |
| Va (mm)   | 6.54 ± 0.87   | 6.55 ± 0.87   | 6.42 ± 0.80   |
| Vp (mm)   | 6.35 ± 1.03   | 6.20 ± 0.99   | 6.09 ± 0.81   |
| Fetal cerebral hemisphere, hem (mm) | 25.65 ± 1.97 | 25.18 ± 2.01 | 25.47 ± 2.28 |
| Ut RI     | 0.54 ± 0.09   | 0.60 ± 0.27   | 0.69 ± 0.35\(^{**}\) |
| Ut PI     | 0.89 ± 0.31   | 0.98 ± 0.43   | 1.12 ± 0.49\(^{a}\) |

Note: Values shown are mean and standard deviation.

Abbreviations: AC, abdominal circumference; BPD, Biparietal Diameter; CM, Cisterna Magna; EFW, estimated fetal weight; FL, femur length; HC, head circumference; TCD, Transverse Cerebellar Diameter; Va, anterior horn of lateral ventricle; Vp, posterior horn of lateral ventricle; Ut PI, uterine arterial pulsatility index; Ut RI, uterine arterial resistance index.

\(^{a}\)Compared to infants with BW above the 10th centile.

\(^{b}\)Compared to infants with BW between the third and 10th centile.

\(^{*}\)< 0.05, \(^{**}\) < 0.001.
gestational age, we added gestational age (GA) as an additional input (the 16th parameter) in the model, to ensure that the model could account and adjust for this variability. This brought our total number of parameters to 16.

In this study, we aimed to investigate the model’s performance in predicting SGA and severe SGA infants at birth using the second trimester data, via two separate experiments. A model to predict SGA infants was trained in the first experiment, and a separate model was trained to predict severe SGA in the second experiment.

We randomly split the data into the training and validation set (80% of data) and the testing set (20% of data). In the testing dataset, the number of samples in each class was balanced to avoid disproportional ratio of observation between classes to avoid misleading accuracy analysis. This necessitated 29 AGA and 29 SGA testing data in the first experiment and leaving 116 AGA and 173 SGA for the training and validation dataset. For the second experiment, only severe SGA cases were used, and the 105 SGA cases with birth weight between the third and 10th centile were not used. Since the number of diseased cases reduced, only 20 AGA and 20 severe SGA cases were used as testing data, leaving 125 AGA and 77 severe SGA for the training and validation dataset. A summary of the number of training and testing data used in the two experiments is shown in Figure 2.

In order to ensure that different types of data had similar orders of magnitude and that data was free of measurement units, data for every parameter were scaled to have a zero mean and unit standard deviation, according to the training data distribution. Next, training data in the class (healthy/disease) with the smaller sample size was randomly duplicated (oversampling) so that the number of samples in both classes was balanced. This was essential to improve the model’s accuracy. Note that, only the training data was oversampled for the model training process and the testing data was not.

During training, we performed five- and 10-fold cross-validation to determine the best hyper-parameters for every model. Hyper-parameters were variables that determine the ML structures, and each models had different hyper-parameters to tune. Details of hyper-parameters for each model and their tuning process are given in the Supplementary Document. In the fivefold cross-validation, the training and validation dataset (80% of all data) was evenly divided into five sets: four sets were used for training and one set was used for validation. This process was repeated five times for each hyper-parameters optimization process. We performed fivefold and 10-fold cross-validation and results showed that there were very similar in model performance, judging by the area under the receiver operating curve, which is reported in Table S2 in the Supplementary Document. The results shown here were based on fivefold cross-validation.

In each fold, we used the Grid Search method to test different sets of hyper-parameters. The hyper-parameters combination that yielded the highest average validation accuracy (average across k-fold cross-validation) was used to train the model again using the training and validation datasets (i.e., the 80% of data, 116 AGA and

![Figure 1](null)
173 SGA in Experiment 1, and 125 AGA and 77 Severe SGA in Experiment 2. The final trained model was then used to make predictions in the testing dataset and the model’s predictive performance was evaluated, which will be further explained in the section “ML Model Performance.”

The process of data preprocessing and model development is summarized in Figure 2. Analyses of RF and SVM were performed using Scikit-learn package v0.19 available in Python language (v2.7) while analysis of MLP was performed using Keras package v2.0.6.

2.3 | Comparison between ML model and Doctor’s diagnosis

Fetuses were diagnosed as at risk of SGA at the second trimester if one of the following criteria was met in accordance to the RCOG guidelines:

- Increased uterine arterial pulsatility index (Ut PI > 95th centile), or
- Estimated fetal weight or abdominal circumference below the 10th centile

These cases were deemed to be cases of the physician-predicted disease. Here, we focused on the predictive properties of fetal ultrasound markers, and as such, other risk factors in the guidelines, such as maternal medical history and blood markers were not considered. For a fair comparison between machine learning models and clinical diagnosis, we used the same input parameters (GA, EFW, AC, and Ut PI) to train the machine learning models using the similar approach described in Figure 2. The models’ performances were evaluated using testing datasets and compared with the doctor’s diagnosis using the same testing datasets.

2.4 | ML model performance

From the experiments using all 16 parameters, we extracted the relative importance of each parameter in predicting SGA and severe SGA via the RF importance score and SVM coefficient magnitude. Both RF and SVM showed that nuchal fold thickness played a significant role in successful prediction (importance ranking in Table 3). We proceeded to test this further by comparing models with and without nuchal fold thickness.

It is known that biometric measurements, such as HC, AC, FL, EFW, and Doppler measurements (Uterine RI and PI) are the parameters correlated to SGA. As such, we selected the following combinations as models without NF:

| GA, Ut PI, Ut RI |
| GA, EFW, Ut PI, Ut RI |
| GA, EFW, AC, Ut PI, Ut RI |
| GA, EFW, HC, AC, FL, Ut PI, Ut RI |

We then repeated the similar experiments by including NF to the above combinations and evaluated the model performance.
Statistical analysis was performed to test if there were differences between the three groups (BW > 10th centile, BW between third and 10th centile and BW < third centile) in terms of the 16 parameters, birth weight and gestational age at birth, using one-way ANOVA with the Tukey post hoc test. The data were deemed significantly different if \( p < 0.05 \).

We performed a multinomial logistic regression analysis to examine the contribution of each parameter (or independent variable) in affecting the outcome variables. The Variation inflation factor (VIF) was computed for every independent variable. VIF was a measure of multicollinearity of independent variables in a multiple regression model, and was to detect variables that are correlated and explained the same variance within the datasets, so that redundant variables can be discarded. Independent variables with VIF more than 10 were excluded in the logistic regression analysis to ensure reliable estimate of regression coefficient.

### RESULTS

#### 3.1 Patient characteristics

Patient characteristics are shown in Table 1. Low birth weight babies (Group B and Group C) were born earlier with significantly lower birth weight compared to AGA (Group A). Sonographic measurements such as abdominal circumferences and nuchal fold thickness were considerably lower in Groups B and C compared to Group A. The estimated fetal weights and transverse cerebellar diameters were significantly lower in Group B compared to Group A. For Group C, these parameters were lower than Group A, but the differences were not statistically significant. Uterine RI and PI were significantly elevated in Group C but not in Group B. Other parameters showed no significant differences among the three groups.

Logistic regression model showed that there were three significant predictors, which were nuchal fold thickness (odd ratio \( = 0.53, p < 0.001 \)), abdominal circumference (odd ratio \( = 0.52, p = 0.007 \)) and uterine PI (odd ratio \( = 1.56, p = 0.004 \)). Abdominal circumference and uterine PI were known to be related to SGA but it was surprising to note that nuchal thin fold thickness also had a predictive relationship.

#### 3.2 Comparison between ML model and doctor’s diagnosis

Comparisons between the predictive capabilities of clinical diagnosis and machine learning models (4-parameters version), based on their sensitivity, specificity and accuracy, are shown in Table 2. The accuracy of models in predicting disease was consistently higher than the clinical diagnosis. In the first experiment for SGA prediction, MLP had the best predictive ability, achieving an accuracy of 71%, which was approximately 7% higher than the clinical diagnosis, which had an accuracy of 64%. In the second experiment for severe SGA prediction, SVM had the highest accuracy among all the models, 73%, which was 25% higher than the clinical diagnosis, which had an accuracy of 48%.

#### 3.3 ML model performance

With the 16-parameters models, the RF model provided an importance score that described how much the accuracy dropped when a variable was excluded from the model. The SVM coefficients described the vector coordinates which were orthogonal...
to the hyperplane and can be interpreted similarly as the importance of the parameters. As expected, Ut RI was found to be an important parameter in severe SGA prediction, this corroborated literature reports of Doppler measurements being a predictor of severe SGA. Here, nuchal fold thickness was again surprisingly found to be important in predicting SGA and severe SGA. In the RF model, nuchal fold thickness was the top important parameter, while in the SVM model, it was the fourth most important parameter. The top four highest parameters' ranking is shown in Table 3, in ascending order.

### TABLE 4

| Model | SGA prediction | Sensitivity | Specificity | Accuracy |
|-------|----------------|-------------|-------------|----------|
| **RF** |                |             |             |          |
| All 16 parameters (GA, EFW, HC, AC, FL, Ut RI, Ut PI, NF, BPD, CM, hem, TCD, Va, Vp, maternal age, ethnics) | 0.66 | 0.83 | 0.74 |
| GA, Ut RI, Ut PI | 0.31 | 0.76 | 0.53 |
| GA, EFW, Ut RI, Ut PI | 0.55 | 0.72 | 0.64 |
| GA, EFW, AC, Ut RI, Ut PI | 0.51 | 0.79 | 0.66 |
| GA, EFW, HC, AC, FL, Ut RI, Ut PI | 0.86 | 0.41 | 0.64 |
| GA, Ut RI, Ut PI, NF | 0.45 | 0.83 | 0.64 |
| GA, EFW, Ut RI, Ut PI, NF | 0.55 | 0.90 | 0.73 |
| GA, EFW, AC, Ut RI, Ut PI, NF | 0.59 | 0.76 | 0.67 |
| GA, EFW, HC, AC, FL, Ut RI, Ut PI, NF | 0.79 | 0.72 | 0.76 |

**Note:** Input parameter combinations that yielded the highest accuracy was written in bold.

**Abbreviations:** AC, Abdominal Circumference; BPD, Biparietal Diameter; CM, Cisterna Magna; EFW, Estimated Fetal Weight; FL, Femur Length; GA, Gestational Age; HC, Head Circumference; Hem, Fetal cerebral hemisphere; NF, Nuchal Fold Thickness; SGA, small-for-gestational-age; SVM, support vector machine; TCD, Transverse cerebellar diameter; Ut RI, uterine arterial resistance index; Ut PI, uterine arterial pulsatility index; Va, Vp, Anterior and posterior horn of lateral ventricle.

Here, nuchal fold thickness was again surprisingly found to be important in predicting SGA and severe SGA. In the RF model, nuchal fold thickness was the top important parameter, while in the SVM model, it was the fourth most important parameter. The top four highest parameters' ranking is shown in Table 3, in ascending order.
The full parameters importance ranking is available in Figures S2 and S3 in the Supplementary Document.

From Table 4, it can further be seen that the addition of nuchal fold thickness to the model could consistently improve the accuracy of the model. For example, the accuracy of the SVM model without nuchal fold thickness reached a bottleneck of 0.69 and did not improve when additional parameters were included. When nuchal fold thickness was included in the model, the SVM model's accuracy improved to 0.75.

### TABLE 5  Model performance in severe SGA prediction with different combinations of input parameters

| Model  | All 16 parameters (GA, EFW, HC, AC, FL, Ut RI, Ut PI, NF, BPD, CM, hem, TCD, Va, Vp, maternal age, ethnics) | Sensitivity | Specificity | Accuracy |
|--------|---------------------------------------------------------------------------------------------------|-------------|-------------|----------|
| RF     | GA, Ut RI, Ut PI                                                                                   | 0.55        | 0.80        | 0.68     |
|        | GA, EFW, Ut RI, Ut PI                                                                             | 0.60        | 0.55        | 0.58     |
|        | GA, EFW, AC, Ut RI, Ut PI                                                                          | 0.50        | 0.75        | 0.63     |
|        | GA, EFW, HC, AC, FL, Ut RI, Ut PI                                                                     | 0.60        | 0.75        | 0.68     |
|        | GA, Ut RI, Ut PI, NF                                                                               | 0.60        | 0.85        | 0.73     |
|        | GA, EFW, Ut RI, Ut PI, NF                                                                           | 0.70        | 0.65        | 0.68     |
|        | GA, EFW, AC, Ut RI, Ut PI, NF                                                                         | 0.65        | 0.75        | 0.70     |
|        | GA, EFW, HC, AC, FL, Ut RI, Ut PI, NF                                                                  | 0.70        | 0.75        | 0.73     |
| SVM    | All 16 parameters (GA, EFW, HC, AC, FL, Ut RI, Ut PI, NF, BPD, CM, hem, TCD, Va, Vp, maternal age, ethnics) | 0.75        | 0.80        | 0.78     |
|        | GA, Ut RI, Ut PI                                                                                   | 0.60        | 0.80        | 0.75     |
|        | GA, EFW, Ut RI, Ut PI                                                                             | 0.75        | 0.65        | 0.70     |
|        | GA, EFW, AC, Ut RI, Ut PI                                                                          | 0.95        | 0.45        | 0.70     |
|        | GA, EFW, HC, AC, FL, Ut RI, Ut PI                                                                     | 0.90        | 0.50        | 0.70     |
|        | GA, Ut RI, Ut PI, NF                                                                               | 0.80        | 0.80        | 0.80     |
|        | GA, EFW, Ut RI, Ut PI, NF                                                                           | 0.75        | 0.90        | 0.83     |
|        | GA, EFW, AC, Ut RI, Ut PI, NF                                                                         | 0.70        | 0.85        | 0.78     |
|        | GA, EFW, HC, AC, FL, Ut RI, Ut PI, Ut PI, NF                                                          | 0.75        | 0.85        | 0.80     |
| MLP    | All 16 parameters (GA, EFW, HC, AC, FL, Ut RI, Ut PI, NF, BPD, CM, hem, TCD, Va, Vp, maternal age, ethnics) | 0.95        | 0.50        | 0.73     |
|        | GA, Ut RI, Ut PI                                                                                   | 0.60        | 0.80        | 0.70     |
|        | GA, EFW, Ut RI, Ut PI                                                                             | 0.75        | 0.55        | 0.65     |
|        | GA, EFW, AC, Ut RI, Ut PI                                                                          | 1.00        | 0.20        | 0.60     |
|        | GA, EFW, HC, AC, FL, Ut RI, Ut PI                                                                     | 0.80        | 0.45        | 0.63     |
|        | GA, Ut RI, Ut PI, NF                                                                               | 0.90        | 0.55        | 0.73     |
|        | GA, EFW, Ut RI, Ut PI, NF                                                                           | 0.60        | 0.85        | 0.73     |
|        | GA, EFW, AC, Ut RI, Ut PI, NF                                                                         | 0.70        | 0.70        | 0.70     |
|        | GA, EFW, HC, AC, FL, Ut RI, Ut PI, Ut PI, NF                                                          | 0.60        | 0.75        | 0.68     |

Note: Input parameter combinations that yielded the highest accuracy was written in bold.

Abbreviations: AC, Abdominal Circumference; BPD, Biparietal Diameter; CM, Cisterna Magna; EFW, Estimated Fetal Weight; FL, Femur Length; GA, Gestational Age; HC, Head Circumference; Hem, Fetal cerebral hemisphere; NF, Nuchal Fold Thickness; SGA, small-for-gestational-age; SVM, support vector machine; TCD, Transverse cerebellar diameter; Ut RI, uterine arterial resistance index; Ut PI, uterine arterial pulsatility index; Va, Vp, Anterior and posterior horn of lateral ventricle.
increased to 0.78. For severe SGA prediction (Table 5), similar observations were obtained where models with nuchal fold thickness showed consistently higher accuracy compared to models without nuchal fold thickness.

The best model with the highest prediction accuracy for both SGA and severe SGA prediction was the SVM model, achieving an accuracy of 78% and 83%, respectively (Tables 4 and 5). For both SGA and severe SGA prediction, the input parameters combination that yielded the highest accuracy were gestational age, estimated fetal weight, uterine RI, uterine PI and nuchal fold thickness.

4 | DISCUSSION

In this study, we used birth weight as a surrogate outcome to identify the infants who were at increased risk of adverse perinatal outcomes because birth weight had been shown to have a direct relation with perinatal mortality and morbidity.2,43,44

We utilized machine learning models in predicting SGA and severe SGA at birth, using second trimester data. We compared our model's performance to a baseline—clinical diagnosis performance, and we showed that machine learning could improve the prediction accuracy by approximately 10%–20% when compared to the clinical diagnosis (Table 2), using similar parameters. However, as Tables 3 and 4 would show, there were other sets of input parameters that enabled even higher machine learning accuracies, although increasing the number of parameters did not necessarily improve accuracy.

There were past attempts to use machine learning to predict growth-restricted fetuses using ultrasound scans alone.15,17,18 These studies utilized measurements from later gestation compared to ours, resulting in an average accuracy of 80%–90%. Our findings corroborated with theirs where we achieved an accuracy of 83%. Our combined findings thus suggested that machine learning is a promising approach in detecting SGA infants. By utilizing only the second trimester measurements, we showed that early prediction of SGA is possible. Past studies, however, placed more emphasis on umbilical Doppler, due to their known relationship with SGA, but did not include nuchal fold measurements. Our results indicated that including it can improve the model's prediction performance. One previous study demonstrated, interestingly, that the inclusion of measurements over a consecutive week of up to 4 weeks in the modeling could improve predictive performance. This approach is a good strategy to improve our current models in the future.

Our study had similarly found that uterine Doppler indices were important predictors of SGA, but this was unsurprising, as many previous studies showed uterine Doppler indices to be well correlated to adverse fetal outcomes and SGA, likely due to alterations in the placental circulation resistance during SGA.48,49

The interesting part of our findings, however, was that utilizing only Doppler measurements without nuchal fold thickness did not yield sufficiently high prediction accuracy. Machine learning models' accuracy reached a bottleneck across various combinations of parameters. This bottleneck could only be breached by including nuchal fold thickness. For example, the SVM model reached an accuracy bottleneck of 0.70 for severe SGA prediction, as shown in Table 5, but adding nuchal fold thickness in the model improved the prediction accuracy to 0.83. The importance of nuchal fold thickness was also shown in the logistic regression analysis and statistical analysis. Logistic regression analysis showed that nuchal fold thickness was one of the significant predictors and statistical results revealed that SGA (Group B and C) had significantly thinner nuchal fold (p-value < 1e-5) compared to appropriately grown fetuses (Group A). We noted that another study reported corroborating evidence that second trimester Doppler alone had limited utility for predicting SGA.50

Increased nuchal fold thickness is traditionally a sonographic marker for fetal chromosomal and structural abnormalities.51,52 Interestingly, in our study, it was reduced nuchal fold thickness that was found to contribute to SGA prediction. There is scarce literature on reduced fetal nuchal fold thickness, but a recent study in adults found that nuchal fold thickness was positively correlated with body mass index.53 We hypothesize that the reduced nuchal fold thickness in SGA is likely due to reduced nuchal adipose tissue, which would be a logical consequence of the reduced growth rates. We propose that investigations on the physiology of nuchal fold thickness in SGA fetuses is well warranted, and could lead to interesting findings.

It has been consistently shown that SGA babies have higher rates of fetal death, preterm birth, hypertension or preeclampsia.54–56 It was reported that there was a significant association between second-trimester SGA and third-trimester SGA, suggesting that growth restriction started earlier than the third trimester in pregnancy.57 With our models, early prediction of SGA could be made, prospective controlled studies of prophylactic treatment (such as aspirin or antioxidants) and optimal nutritional supplementation58 could be conducted in the high-risk population. Furthermore, efforts are underway to develop pharmacological59 and other therapies60,61 for growth-restricted fetuses, and having an early detection tool will be invaluable should a therapy proves effective, to help select subjects for the therapy. Much effort has also been made to understand this idiopathic disease using various tools, such as using computational modeling,62,63 biomechanics analyses64–66 and genetic analyses.66 Early detection of SGA would allow clinicians to monitor at-risk fetuses more closely before obvious features developed, which could improve perinatal outcomes and could lead to the discovery of other underlying factors causing the disease which may otherwise have been missed.

This study, however, is only a first step in improving prediction of SGA babies at earlier gestation. There are several more ways to extend the current study, such as increasing the sample size and including other risk factors, such as pregnancy history, maternal blood pressure, smoking status, pregnancy complications such as preeclampsia, gestational diabetes mellitus, and pregnancy-associated plasma protein A (PAPP-A) level.56 With these additional data, it may be possible
to further increase the accuracy and the generalizability of the models. To enable readers to test our machine learning algorithm, we provided an Excel file in the supplementary material that can compute the probability of the babies being born healthy or SGA using the MLP model. Further, the Python codes for our models are available in https://github.com/shiernee/Small-for-Gestational-Age_Prediction. We emphasize that this should be used for reference, and not yet for diagnosis purposes.

A further limitation of this study was that we were unable to evaluate inter- and intra-observer variability of the measurements because it was a retrospective study. However, ultrasound measurements were performed by experienced sonographers in our hospital. Further, since the data had already been de-identified, there could be subjects in our study who had more than one pregnancy. Also, there was a small but significant difference in birth age of the three cohorts. While we had ensured that all subjects were not born preterm, and did not have detectable abnormalities other than SGA, we might not have ruled out all possible underlying abnormalities.

5 | CONCLUSION

We showed that machine learning can improve the accuracy of early prediction of SGA at birth and presented the combination of parameters that can optimize predictive capabilities. Furthermore, we showed that our machine learning modeling unexpectedly revealed a thin nuchal fold to be an important predictive parameter for SGA at birth, where we might neglect in the past. As such, it might be worthwhile to put attention on this parameter during the second trimester antenatal screening. Early detection of SGA is important because it can lead to improved outcomes via management strategies, enable the detection of underlying diseases and will be invaluable in the future when intrauterine therapeutic and preventive strategies become available.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Dr. Claudia Chi, REDCAP team, Maylene Tan Zipagan, Felicia Toh Xinying, Dawn Chia and Julie Ng from Fetal Care Centre, National University Hospital, Singapore for their assistance with data extraction. This research was supported by the Singapore Ministry of Education Academic Research Fund Tier 1 Grant 2016 (PI: Yap), and the Biomedical Research Council of the Agency for Science, Technology, and Research, Singapore (PI: Lee).

CONFLICT OF INTEREST

All authors have no conflict of interest to declare.

ETHICS STATEMENT

The study protocol (Ref No: 2014/01267) was approved by the National Healthcare Group (Singapore) Domain Specific Review Board in September 2018.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

ORCID

Choon Hwai Yap https://orcid.org/0000-0003-2918-3077

REFERENCES

1. American College of Obstetrics and Gynecology. ACOG Practice bulletin no. 134: fetal growth restriction. Obstetrics Gynecol. 2013;121(5):1122-1133. https://pubmed.ncbi.nlm.nih.gov/23635765/.
2. McIntire DD, Bloom SL, Casey BM, Leveno KJ. Birth weight in relation to morbidity and mortality among newborn infants. N Engl J Med. 1999;340(16):1234-1238.
3. Faraci M, Renda E, Monte S, et al. Fetal growth restriction: current perspectives. J Perinat Med. 2011;39(2):31-33.
4. Barker DJ. Fetal origins of cardiovascular disease. Ann Med. 1999;31(Suppl 1):3-6.
5. Seckl JR. "vi.; 1998:939-962. Physiologic programming of the fetus Clin perinatology4
6. Hill DJ. Fetal programming of the pancreatic β cells and the implications for postnatal diabetes. Seminars Neonatol. 1999;4(2):99-113.
7. Gardosi J, Madurasimive V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. BMJ. 2013;346:f1018.
8. Lindqvist PG, Molin J. Does antenatal identification of small-for-gestational age fetuses significantly improve their outcome? Ultrasound Obstetrics Gynecol. 2005;25(3):258-264.
9. David AL. Maternal uterine artery VEGF gene therapy for treatment of intrauterine growth restriction. Placenta. 2017;59:S44-S50.
10. Pels A, Kenny LC, Alfrevic Z, et al. STRIDER (Sildenafil TheRapy in dismal prognosis early onset fetal growth restriction): an international consortium of randomised placebo-controlled trials. BMC Pregnancy Childbirth. 2017;17(1):440.
11. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J. 2017;38(23):1805-1814.
12. Pochet N, Suykens J. Support vector machines versus logistic regression: improving prospective performance in clinical decision-making. Ultrasound Obstet Gynecol. 2006;27(6):607-608.
13. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res. 2016;18(12), e232.
14. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inf. 2002;35(5):352-359.
15. Gurgen F, Onal E, Varol FG. IUGR detection by ultrasonographic examinations using neural networks. IEEE Eng Med Biol Mag. 1997;16(3):55-58.
16. Bagi KS, Shreedhara KS. Biometric Measurement and Classification of IUGR Using Neural Networks. Paper presented at: 2014 International Conference on Contemporary Computing and Informatics; November 27-29, 2014;ICCI.
17. Gurgen N, Gurgen F, Varol FG. Decision of normal and growth retarded pregnancies by discriminant functions using umbilical arterial blood flow velocities waveforms. Paper presented at: Advance in Intelligent Systems and Computer Science; 1999; World Scientific and Engineering Society.
18. Gürgen F, Zengin Z, Varol F. Intrauterine growth restriction (IUGR) risk decision based on support vector machines. Expert Syst Appl. 2012;39(3):2872-2876.
19. Kuhle S, Maguire B, Zhang H, et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. BMC Pregnancy Childbirth. 2018;18(1):333.

20. Papastefanou I, Wright D, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonate from maternal characteristics and medical history. Ultrasound Obstetrics Gynecol. 2020;56(2):196-205.

21. Papastefanou I, Wright D, Lolos M, Anampouzi K, Mamalis M, Nicolaides KH. Competing-risks model for prediction of small-for-gestational-age neonates from maternal characteristics, serum PAPP-A and PIgf at 11-13-weeks’ gestation. Ultrasound Obstetrics Gynecol. 2020. https://doi.org/10.1002/ugo.23118

22. Villar J, Cheikh Ismail L, Victora C, et al. International fetal and newborn growth consortium for the 21st century (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age and sex: the newborn cross-sectional study of the INTERGROWTH-21st project. Lancet. 2014; 384(9946):857-868.

23. Salomon LJ, Alfrevic Z, Berghella V, et al. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstetrics Gynecol. 2011;37(1):116-126.

24. Salomon LJ, Alfrevic Z, Bilardo CM, et al. ISUOG practice guidelines: performance of first-trimester fetal ultrasound scan. Ultrasound Obstetrics Gynecol. 2013;41(1):102-113.

25. Singh A, Thakur N, Sharma A. A Review of Supervised Machine Learning Algorithms 16-18 March 2016. Paper presented at: 2016 3rd International Conference on Computing for Sustainable Global Development; 2016; INDIACom.

26. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: a review of classification techniques. Emerging artificial intelligence applications in computer engineering. 2007;160:3-24.

27. Buck Louis GM, Grewal J, Albert PS, et al. Racial/ethnic standards for fetal growth; the NICHD Fetal Growth Studies. Am J Obstetrics Gynecol. 2015;213(4):449.e441.

28. Alexander GR, Wingate MS, Mor J, Boulet S. Birth outcomes of Asian-Indian-Americans. Int J Gynecol Obstetrics. 2007;97(3):215-220.

29. Albemery M, Soothill P. Management of fetal growth restriction. Archives Dis Childhood-Fetal Neonatal Ed. 2007;92(1):F62-F67.

30. Goldstein I, Reece EA, Pilu G, Bovicelli L, Hobbins JC. Cerebellar measurements with ultrasonography in the evaluation of fetal growth and development. Am J Obstetrics Gynecol. 1987;156(5):1065-1069.

31. Sug H. Improving the prediction accuracy of liver disorder disease with oversampling. Applied Mathematics in Electrical and Computer Engineering. 2012;7:331-335.

32. Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Med Inf Decis Mak. 2010;10(1):16.

33. Richert W. Building Machine Learning Systems with Python. Birmingham, UK: Packt Publishing Ltd: 2013.

34. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.

35. Chollet FK. Keras package; Version 2.0.6. GitHub; 2017/2015. https://github.com/fchollet/keras

36. Royal College of Obstetricians and Gynaecologists. RCOG. Small-for-Gestational-Age Fetuses. Investigation and Management [Green-top Guideline No. 31]. London, United Kingdom: Royal College of Obstetricians and Gynaecologists; 2013.

37. Bagley SC, White H, Colomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. J Clin Epidemiol. 2001;54(10):979-985.

38. Mansfield ER, Helms BP. Detecting multicollinearity. Am Statistician. 1982;36(3a):158-160.

39. Ferrazzi E, Bulfamante G, Mezzopane R, Barbera A, Ghidini A, Pardi G. Uterine Doppler velocimetry and placental hypoxic-ischemic lesion in pregnancies with fetal intrauterine growth restriction. Placenta. 1999;20(5):389-394.

40. Qi Y. Random forest for bioinformatics. In: Zhang C, Ma Y, eds. Ensemble Machine Learning: Methods and Applications. Boston, MA: Springer US; 2012:307-323.

41. Kelleher JD, Mac Namee B, D’arcy A. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge, Massachusetts: MIT press; 2015.

42. Crossen JS, Morris RK, ter Riet G, et al. Use of uterine artery Doppler ultrasonography to predict pre-eclampsia and intrauterine growth restriction: a systematic review and bivariate meta-analysis. Can Med Assoc J. 2008;178(6):701-711.

43. Clausson B, Gardosi J, Francis A, Cnattingius S. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. BJOG. 2001;108(8):830-834.

44. Alexander GR, Kogan M, Bader D, Carlo W, Allen M, Mor J. US birth weight/gestational age-specific neonatal mortality: 1995–1997 rates for whites, hispanics, and blacks. Pediatrics. 2003;111(1):e61-e66.

45. Dugoff L, Lynch AM, Cioffi-Ragan D, et al. First trimester uterine artery Doppler abnormalities predict subsequent intrauterine growth restriction. Am J Obstetrics Gynecol. 2005;193(Suppl 3):1208-1212.

46. Lührba C, Carreras E, Gratacos E, et al. Maternal history and uterine artery Doppler in the assessment of risk for development of early and late-onset preeclampsia and intrauterine growth restriction. Obstetrics Gynecol Int. 2009;2009.

47. Murakoshi T, Sekizuka N, Takakuwa K, Yoshizawa H, Tana K. Uterine and spiral artery flow velocity waveforms in pregnancy-induced hypertension and/or intrauterine growth retardation. Ultrasound Obstetrics Gynecol. 1996;7(2):122-128.

48. Adamson SL, Morrow RJ, Bascom PA, Mo LY, Ritchie JK. Effect of placental resistance, arterial diameter, and blood pressure on the uterine arterial velocity waveform: a computer modeling approach. Ultrasound Med Biol. 1989;15(5):437-442.

49. Mo LY, Bascom PA, Ritchie K, McCoyon LM. A transmission line modelling approach to the interpretation of uterine Doppler waveforms. Ultrasound Med Biol. 1988;14(5):365-376.

50. Payr S, Sciscione A, Haas DM, et al. Role of early second-trimester uterine artery Doppler screening to predict small-for-gestational-age babies in nulliparous women. Am J Obstet Gynecol. 2017;217(5):594.e591-594.e510.

51. Crane JP, Gray DL. Sonographically measured nuchal skinfold thickness as a screening tool for Down syndrome: results of a prospective clinical trial. Obstetrics Gynecol. 1991;77(4):533-536.

52. Benacerraf BR, Barss VA, Laboda LA. A sonographic sign for the detection in the second trimester of the fetus with Down’s syndrome. Am J Obstetrics Gynecol. 1985;151(8):1078-1079.

53. Sterkenburg AS, Hoffmann A, Reichel J, et al. Nuchal skinfold thickness: a novel parameter for assessment of body composition in childhood chondralhypertrophy. J Clin Endocrinol Metabolism. 2016; 101(12):4922-4930.

54. Misra DP. The effect of the pregnancy-induced hypertension on fetal growth: a review of the literature. Paediatr Perinat Epidemiol. 1996;10(3):244-263.

55. Li X, Zhang W, Lin J, et al. Preterm birth, low birthweight, and small for gestational age among women with preeclampsia: does maternal age matter? Pregnancy Hypertens. 2018;13:260-266.

56. Olazagutia RA, Vatten LJ, Nilsen ST, Salvesen KÅ, Austgulen R. Preeclampsia and fetal growth. Obstetrics Gynecol. 2000;96(6):950-955.

57. Fox NS, Huang M, Chasson S. Second-trimester fetal growth and the risk of poor obstetric and neonatal outcomes. Ultrasound Obstetrics Gynecol. 2008;32(1):61-65.
58. Kramer MS, Kakuma R. Energy and protein intake in pregnancy. Cochrane database Syst Rev. 2003(4):Cd000032.

59. Sharp A, Comforth C, Jackson R, et al. OC01. 05: STRIDER UK: a randomised controlled trial of sildenafil therapy in dismal prognosis early-onset intrauterine growth restriction. Ultrasound Obstetrics Gynecol. 2017;50:3-3.

60. Kingdom JC, Audette MC, Hobson SR, Windrim RC, Morgen E. A placenta clinic approach to the diagnosis and management of fetal growth restriction. Am J Obstetrics Gynecol. 2018;218(Suppl 2):S803-S817.

61. Spiroski AM, Oliver MH, Harding JE, Bloomfield FH. Intrauterine intervention for the treatment of fetal growth restriction. Curr Pediatr Rev. 2016;12(3):168-178.

62. Lin M, Mauroy B, James JL, Tawhai MH, Clark AR. A multiscale model of placental oxygen exchange: the effect of villous tree structure on exchange efficiency. J Theor Biol. 2016;408:1-12.

63. Chernyavsky IL, Jensen OE, Leach L. A mathematical model of intervillous blood flow in the human placentone. Placenta. 2010; 31(1):44-52.

64. Saw SN, Tay JJH, Poh YW, et al. Altered placental chorionic arterial biomechanical properties during intrauterine growth restriction. Sci Rep. 2018;8(1):16526.

65. Ferguson VL, Dodson RB. Bioengineering aspects of the umbilical cord. Eur J Obstetrics Gynecol Reproductive Biol. 2009;144: S108-S113.

66. Monk D, Moore GE. Intrauterine growth restriction—genetic causes and consequences. Seminars Fetal Neonatal Med. 2004;9(5): 371-378.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Saw SN, Biswas A, Mattar CNZ, Lee HK, Yap CH. Machine learning improves early prediction of small-for-gestational-age births and reveals nuchal fold thickness as unexpected predictor. Prenatal Diagnosis. 2021;41:505–516. https://doi.org/10.1002/pd.5903