Use of AI to assess Control and Diseased children at 10 Years of Age

Taher A. Biala[1,2], Ahmed Ramhi[2] Obianom Ekenedirichukwu[2], Xin Li[2], Fernando S. Schlindwein[2]
[1]Higher Institute of Medical Technology, Misrata, Libya
[2]University of Leicester, Biomedical Group, Engineering Department, UK
[3]University of Leicester, Health Science Department. UK

Abstract

This research endeavors to leverage advanced AI algorithms for the differentiation between typically developing children and those afflicted with Intrauterine Growth Restriction (IUGR) disease. We deployed three distinct AI algorithms: Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM). These algorithms were applied to the task of classifying the two target groups while discerning the pivotal parameters contributing to their classification. The obtained results yielded classification accuracy scores of 92.89%, 89.89%, and 87.67% for QDA, LDA, and SVM, respectively. Notably, the analysis revealed that parameters related to birth weight held the most substantial influence in distinguishing between the two cohorts. In light of our conclusive findings, we recommend the utilization of Quadratic Discriminant Analysis (QDA) as a valuable tool for clinicians seeking to identify children at risk of IUGR disease. This research contributes to the enhancement of diagnostic methodologies in pediatric medicine, fostering more accurate and timely interventions for affected individuals.

1. Introduction

Coronary heart diseases (CHD) represent a significant global health concern, accounting for a substantial proportion of worldwide mortality rates. Pioneering research conducted by Barker [1] has illuminated a noteworthy association between individuals who experienced intrauterine growth retardation (IUGR) and their heightened susceptibility to CHD and hypertension in adulthood. Within the realm of cardiac health, reduced Heart Rate Variability (HRV) has emerged as a key indicator, linked to an increased risk of cardiac mortality among adults [2]. HRV is defined as the fluctuation over time in the intervals between successive heartbeats, predominantly regulated by extrinsic factors governing heart rate dynamics. HRV serves as a valuable proxy for assessing overall cardiac health and the state of the Autonomic Nervous System (ANS), which plays a pivotal role in orchestrating cardiac activity [3]. The ANS exerts control over heart rate through two primary branches: the parasympathetic (vagal) and sympathetic nervous systems. The parasympathetic branch, mediated by acetylcholine release, decelerates heart rate, while the sympathetic nerves, via noradrenalin release, stimulate a faster heart rhythm [4]. Perturbations in ANS function, manifesting as reduced HRV, have been conclusively linked to an elevated risk of cardiovascular mortality.

In this study, we employ an array of models and cutting-edge Artificial Intelligence (AI) algorithms to classify and discriminate between individuals within the normal and the diseased cohorts. Our investigation aims to elucidate critical insights into the interplay between IUGR, HRV, and the risk of coronary heart diseases, ultimately contributing to the refinement of diagnostic and therapeutic strategies in the realm of cardiovascular health.

2. Method

In this study, we scrutinized the RR (NN) sequences within a distinctive dataset, encompassing 44 children with IUGR and 31 individuals in the normal cohort. We focused on figuring out when the day and night start by setting a specific RR threshold, which are indicated by the two redlines in figure 1. Upon reaching this threshold, our algorithm meticulously demarcated the commencement of daytime and night-time periods. To ensure uniformity and precision in our RR signal analysis, we applied cubic spline interpolation techniques.

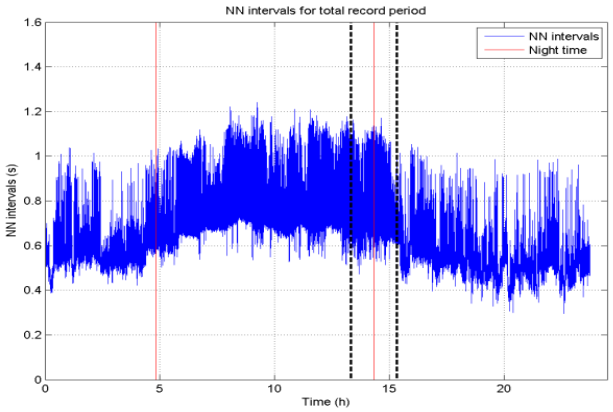

Figure 1: This figure is showing RR (NN) of day and night time, indicated by red the lines.

The dataset was divided into training and testing sets, with the testing set comprising 25% of the total data. To ensure an equitable representation of both normal and diseased data points in the test set, a stratified shuffling split method was employed. Additionally, 10-fold cross-validation was conducted to ensure that the final model's performance is not contingent on the specific data split used.

Several preprocessing steps were applied to the data. These steps included retaining only one of the parameters exhibiting a linear correlation higher than 0.99, scaling the data, and removing parameters with a variance less than 1%.

For feature selection, the ANOVA F-test was employed to rank the parameters when fitting the training data to a classification model. Subsequently, a combination of Principal Component Analysis (PCA) using various kernels (linear, polynomial, RBF, sigmoid, and cosine) was applied to reduce the dimensionality of the data. Following PCA, the Synthetic Minority Oversampling Technique (SMOTE) algorithm was employed to address data imbalance.

In the final phase of our analysis, three classifiers were trained on the balanced dataset, specifically Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM). The Area under the Receiver Operating Characteristic Curve (ROC-AUC) was used as the pivotal metric for ranking the performance of the algorithms.

3. Results and Discussion

The previously mentioned models were applied to the Heart Rate Variability (HRV) data of both normal and diseased children with the objective of classifying and distinguishing between the two cohorts.

Among the models evaluated, QDA emerged with the highest performance score, followed by LDA and, lastly, SVM. Notably, in the case of QDA, the parameters that contributed most significantly to the classification were birth weight centile and birth weight. For LDA, a subset of 15 parameters was selected, while SVM relied on 6 parameters for its classification.

Figures 2 and 3 depict the decision boundary maps generated using the QDA model, projected onto the first-second component plane and the first-third component plane, respectively. These visualizations clearly illustrate a distinct demarcation between the IUGR and normal cohorts. In the context of PCA, optimal results were achieved when employing a third-degree polynomial kernel with 7 principal components. Furthermore, the algorithm identified two paramount parameters, namely birth weight and birth centile, as the most influential in the classification process. It is noteworthy that the inclusion of additional parameters resulted in a deterioration of model performance, as demonstrated in Figure 4.

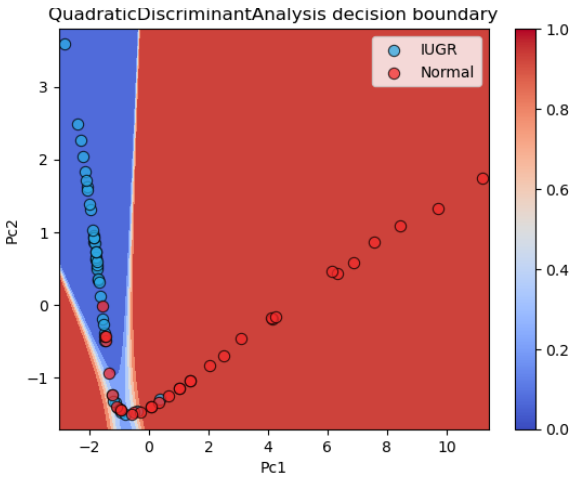

Figure 2: This figure shows the decision boundary map for principal components 1 and 2.

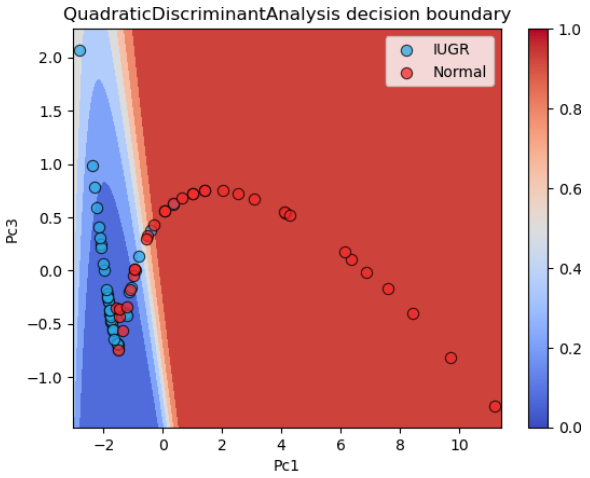

Figure 3: This figure shows the decision boundary map for principal components 1 and 3.

Figures 5 and 6 depict the decision boundary maps generated using the LDA model, projected onto the first-second component plane and the first-third component plane, respectively. These visualizations clearly illustrate a distinct demarcation between the IUGR and normal cohorts. In the context of PCA, optimal results were achieved when employing a linear kernel with 3 components. Furthermore, figure 7 shows a heat map of the parameters yielding the highest score for the LDA model.

Similar to QDA, the inclusion of additional parameters resulted in a deterioration of model performance, as demonstrated in figure 8.

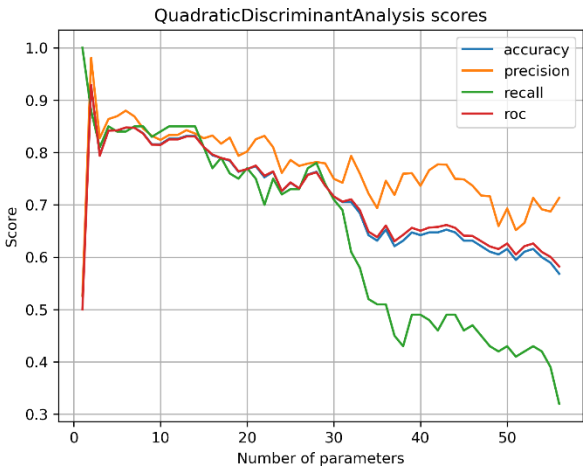

Figure 4 This figure shows the QDA model scores as a function the number of parameters used.

A heatmap illustrating the 15 parameters identified by LDA as having the most substantial influence on the separation of cohorts is presented in figure 6.

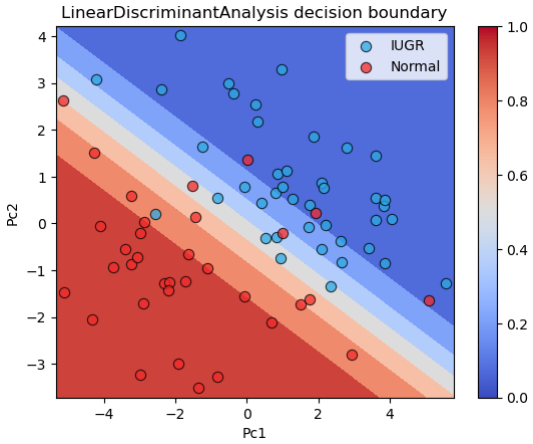

Figure 5 This figure shows the decision boundary map for principal components 1 and 2.

Finally, figures 9 and 10 shows the decision boundary maps generated using the SVM model. As in the previous two models, the figures clearly illustrate a distinct separation between the IUGR and normal cohorts. In the context of PCA, optimal results were achieved when employing a cosine kernel with 3 components. Furthermore, figure 11 shows a heat map of the parameters yielding the highest score for the SVM model. Similar to the other two models, the inclusion of additional parameters resulted in a deterioration of model performance, as demonstrated in Figure 12.

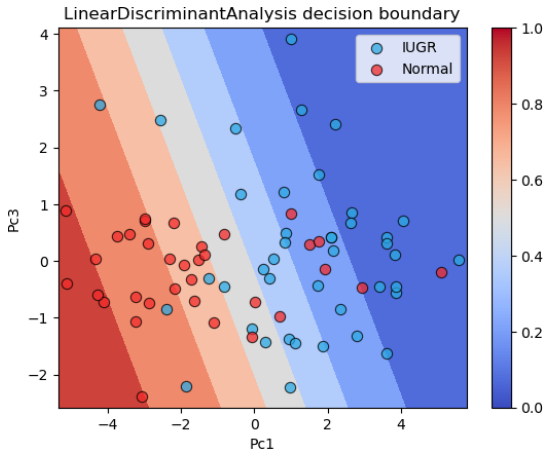

Figure 6 This figure shows the decision boundary map for principal components 1 and 3.

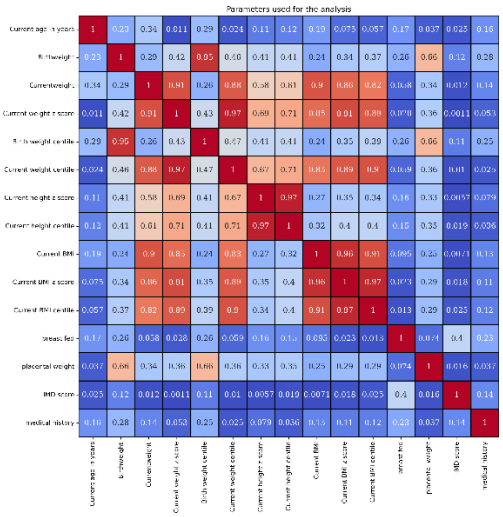

Figure 7 This figure shows a heat map of the parameters predicted by LDA as producing the highest-class separation.

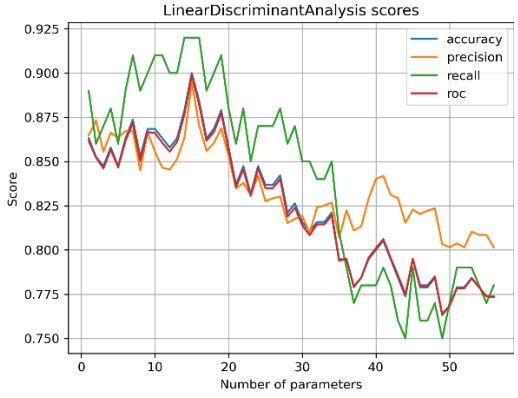

Figure 8 This figure shows the LDA model scores as a

function the number of parameters used.

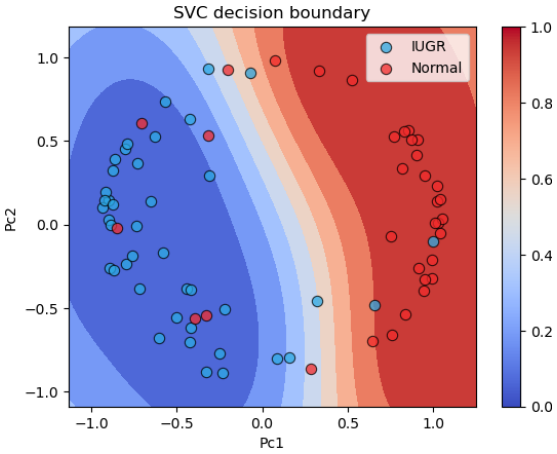

Figure 9 This figure shows the decision boundary map for principal components 1 and 2.

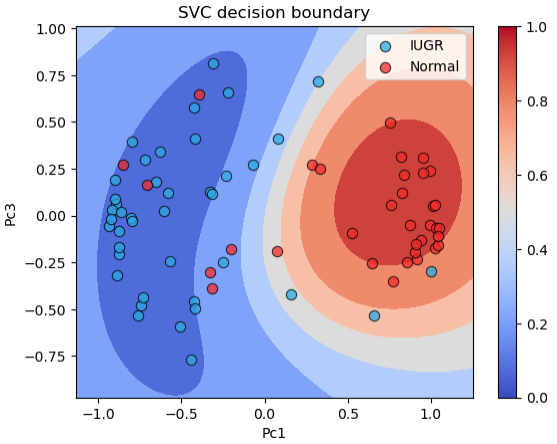

Figure 10 This figure shows the decision boundary map.

Finally, table 1 summarize the results of the three models.

Table 1. The score and the number of parameters used in each model.

Model	No. parameters	Score
QDA	2	92.89
LDA	15	89.89
SVM	6	87.67

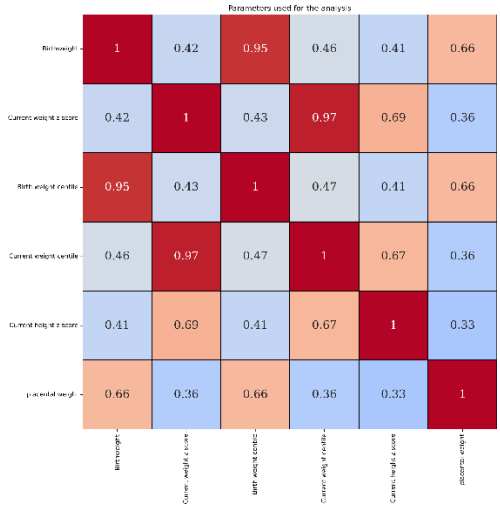

Figure 5 This figure shows a heat map of the parameters predicted by SVM as producing the highest-class separation.

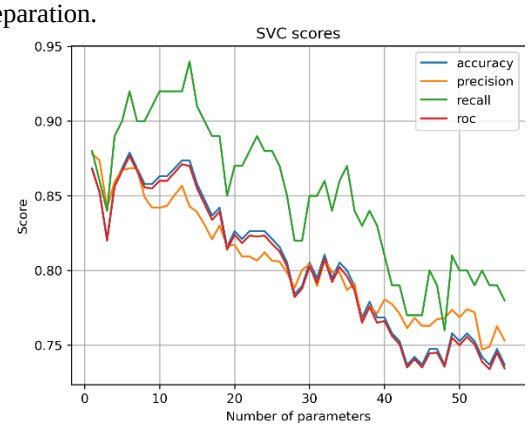

Figure 12 This figure shows the SVM model scores as a function the number of parameters used.

Conclusion

All three AI algorithms employed in this study demonstrated exceptional proficiency in accurately classifying and distinguishing between the Normal and diseased cohorts. Notably, essential parameters such as birth weight and birth centile were incorporated into the classification models, aligning with the seminal birth origin theory elucidated by Barker.

The utilization of these AI models and tools has the potential to serve as valuable aids for clinicians in the precise definition and differentiation of control and IUGR-afflicted children. It is imperative to emphasize that further data acquisition is essential to validate these results rigorously and to ascertain the viability of integrating such AI models into clinical setting.

Acknowledgments

Taher Biala is funded by the Libyan Ministry of Higher Education and Scientific Research. Michael Wailoo was supported by a grant from Leicestershire, Northampton shire and Rutland Primary Care Research Alliance.

References

[1] Barker DJ. The development origin of chronic of adult disease. Actia Paediatrica. 2004;93(11):26-33.
[2] Heitmann A, Huebner T, Schroeder R, Voss A. Ability of heart rate variability as screening tool for heart diseases in men. Computers in cardiology 2009; 36:825-826
[3] Smith RL, Wathen ER, Abaci P, Von Bergen NH, Law IH, Dick MD, Connor C, Dove EL. Analysing heart rate variability in infant using non-linear Poincaré techniques. Computer in cardiology 2009; 36:673-676
[4]. The Task Force. Heart rate variability. Standard of measurement, physiological interpretation, and clinical use. European Heart Journal. 1996;17:354-81.
[5] Rajendra Acharya U, Paul Joseph K., Kannathal N, Choo Min Lim, Jasjit S. Suri. Heart rate variability: A review. Med. Bio. Eng. Comput. (2006), 44:1031-1051.

Address for correspondence.
Taher Ali Biala
Engineering Department,
University of Leicester,
University Road, Leicester,
LE1 7RH, United Kingdom
E-mail: tab40@le.ac.uk