INTERNATIONAL JOURNAL OF ADVANCED ARTIFICIAL INTELLIGENCE RESEARCH (IJAAIR)

eISSN: 3087-4297

Volume. 02, Issue. 08, pp. 01-08, August 2025"

A Machine Learning Approach to Identifying Maternal Risk Factors for Congenital Heart Disease

Dr. Larian D. Venorth
Department of Computational Biology, Geneva Institute for Advanced Medicine, Geneva, Switzerland

Prof. Elias J. Vance
Department of Artificial Intelligence, University of Oxford, Oxford, United Kingdom

Article received: 05/06/2025, Article Revised: 06/07/2025, Article Accepted: 01/08/2025
DOI: https://doi.org/10.55640/ ijaair-v02i08-01

ABSTRACT

Objective: This article explores the application of machine learning (ML) models to identify maternal risk factors for congenital heart defects (CHDs) in offspring. CHDs are the most common birth anomalies, affecting approximately 1 in 100 live births, and early risk identification is crucial for improving neonatal outcomes. This study aims to evaluate the performance of various ML algorithms and identify key maternal factors associated with CHD prediction.
Methods: We conducted a review of existing literature, focusing on studies that used ML models for CHD risk prediction. The analysis included various algorithms, from simpler models like Logistic Regression to more complex ensemble methods (Random Forest, Gradient Boosting) and Neural Networks. We also considered critical aspects of the ML pipeline, including data preprocessing, feature selection from maternal electronic health records and environmental registries, and the use of key evaluation metrics such as AUC-ROC, precision, recall, and F1-score to assess clinical utility.
Results: Our analysis indicates that advanced ML models, particularly ensemble methods and Neural Networks, consistently outperform traditional statistical approaches and simpler ML models. These models effectively leverage a wide range of input features, including maternal age, pre-existing medical conditions, and environmental exposures, to achieve superior predictive accuracy and recall. The enhanced performance of these models highlights their potential for identifying at-risk pregnancies, which is essential given the high stakes of false negatives.
Conclusion: Machine learning is a transformative tool for prenatal risk assessment, offering a powerful way to identify maternal risk factors for CHDs. The application of these models can facilitate targeted counseling for parents, optimize prenatal monitoring, and enable planned deliveries at specialized centers. While challenges such as data privacy and model interpretability must be addressed, the integration of ML into clinical practice holds immense promise for improving health outcomes for infants with congenital heart defects.

KEYWORDS

Congenital heart defects, Machine learning, Predictive modeling, Maternal health, Prenatal risk assessment, Ensemble methods, Explainable AI.

INTRODUCTION

Congenital heart defects (CHDs) are the most common birth anomalies worldwide, affecting approximately 1 in 100 live births and representing a significant public health challenge [16, 25]. These structural abnormalities of the heart are the leading cause of birth defect-related morbidity and mortality, placing a substantial emotional, financial, and logistical burden on affected families and healthcare systems [5, 23]. While there have been remarkable advancements in the surgical and medical

management of CHDs, early and accurate prenatal identification remains the most effective strategy for mitigating adverse neonatal outcomes. A timely prenatal diagnosis allows for targeted parental counseling, planned delivery at tertiary care centers with specialized pediatric cardiology services, and immediate postnatal intervention, all of which are correlated with improved infant survival and long-term quality of life [23, 25]. The importance of this early identification cannot be overstated, as infants with undiagnosed critical CHDs can experience catastrophic decompensation shortly after birth.

The etiology of CHDs is complex and multifactorial, involving a convoluted interplay of genetic predispositions, maternal health status, and environmental exposures [7, 16]. Traditional risk assessment methods, which primarily rely on a limited number of established factors—such as advanced maternal age, a maternal history of diabetes, or a family history of CHD—often fall short of providing a comprehensive risk profile [16]. These conventional statistical approaches are typically designed to analyze linear relationships and struggle to account for the intricate, non-linear interactions that are likely at play in the development of CHDs. For instance, the combined effect of a specific environmental exposure and a maternal comorbidity may create a risk profile that is not simply the sum of its parts, a nuance that traditional models often fail to capture. The vast and high-dimensional nature of modern healthcare data, including detailed electronic health records (EHRs), prenatal screening results, and environmental registry data, further compounds the limitations of these conventional methods, as they are not well-equipped to process such a large number of variables simultaneously [10].

This critical need for a more powerful and nuanced analytical tool has positioned machine learning (ML) as a promising solution [3, 10, 11]. As a subfield of artificial intelligence, machine learning algorithms are uniquely capable of identifying complex, non-linear patterns within large datasets without requiring a priori assumptions about the underlying data distribution. This inherent capability allows them to leverage the full richness of available data, potentially uncovering subtle risk factors and complex interactions that are invisible to traditional statistical analysis [19]. The application of ML has already demonstrated significant success in various areas of medicine, from predicting adult cardiovascular disease [4, 6] to optimizing surgical risk stratification in pediatric patients with CHD [2, 8]. These achievements underscore the potential for ML to revolutionize the field of maternal-fetal medicine. By employing ML, researchers and clinicians can move beyond simple correlations to develop highly accurate predictive models that offer a new paradigm for prenatal risk assessment, enabling a more precise, personalized, and proactive approach to care.

This article provides a comprehensive and in-depth analysis of the application of machine learning for identifying maternal risk factors for congenital heart disease. We will systematically explore the entire ML pipeline, from data acquisition and rigorous preprocessing to model selection and performance evaluation using clinically relevant metrics. Furthermore, we will delve into the profound clinical implications of these predictive models, discussing how they can be leveraged to facilitate targeted counseling and improve neonatal outcomes. We will also critically examine the significant challenges—including data privacy, model interpretability, and generalizability—that must be addressed for widespread clinical adoption. Finally, we will propose key future directions for research, highlighting the potential for integrating multimodal data and the necessity of prospective clinical validation. This synthesis aims to provide a robust framework for understanding the current state and future potential of machine learning in this vital area of maternal-fetal medicine.

METHODS

The successful implementation of a machine learning model for predicting congenital heart disease in offspring requires a meticulous and well-defined methodological framework. This section details the fundamental steps of the ML pipeline, from the sourcing and preparation of diverse data to the selection, training, and evaluation of predictive algorithms. The discussion is grounded in the principles of rigorous data science and the unique clinical challenges inherent in rare disease prediction.

1. Data Sources and Feature Engineering

The predictive power of any machine learning model is directly proportional to the quality and breadth of its input data. For prenatal CHD risk assessment, relevant data is typically derived from several key sources, each contributing a unique dimension to the overall risk profile. The primary source is electronic health records (EHRs), which provide a rich, longitudinal history of a mother's health. Data points extracted from EHRs can include:

- Demographic Information: Basic patient characteristics such as maternal age, ethnicity, and socioeconomic status. Advanced maternal age is a well-documented risk factor that consistently ranks as a highly predictive feature in ML models [10, 19].

- Medical History and Comorbidities: Data on pre-existing conditions, which are often coded using systems like ICD-10. These include maternal diabetes mellitus, hypertension, autoimmune disorders, and obesity, all of which have been associated with increased CHD risk. The granularity of EHR data allows for the inclusion of specific diagnoses and their duration, providing a more

detailed feature set.

- Medication and Substance Use: Information on a mother's medication history, particularly exposure to teratogenic drugs during pregnancy, as well as a history of alcohol, tobacco, or illicit substance use.

- Laboratory and Vitals Data: A comprehensive record of lab results, such as blood glucose levels, HbA1c, and other biomarkers, as well as vital signs throughout the pregnancy.

Complementary to EHRs are data from specialized registries and environmental databases. These sources can provide valuable context on environmental exposures—such as air and water quality data correlated with a patient's residential address [7]—and specific lifestyle factors that may not be systematically recorded in an EHR [16]. The process of feature engineering is then employed to transform this raw, heterogeneous data into a structured format suitable for ML algorithms. This involves creating meaningful variables from raw data, such as calculating the duration of a disease, or creating a binary feature for exposure to a specific environmental pollutant. The ability of ML to handle a vast number of these engineered features is a major advantage over traditional statistical methods that are often limited by the risk of multicollinearity and overfitting with too many variables [10].

2. Data Preprocessing and Class Imbalance Mitigation

Raw healthcare data is rarely in a pristine state. Before any ML model can be trained, a rigorous data preprocessing phase is non-negotiable to ensure the model's robustness and reliability.

- Handling Missing Values: Missing data is a ubiquitous issue in clinical records. Simple imputation techniques, such as replacing missing values with the mean, median, or mode, are often used, but they can introduce bias. More advanced methods, such as K-Nearest Neighbors (KNN) or model-based imputation, can leverage the existing data to more accurately estimate missing values [19]. The choice of imputation strategy can significantly impact the final model's performance.

- Feature Encoding and Scaling: Machine learning algorithms operate on numerical data. Therefore, categorical features (e.g., race, type of medication) must be converted into a numerical representation. One-hot encoding is a standard practice that creates a new binary column for each category, preventing the algorithm from incorrectly inferring an ordinal relationship. Furthermore, features with vastly different scales (e.g., maternal age in years versus blood pressure in mmHg) must be scaled to a consistent range to prevent algorithms from placing undue emphasis on features with larger numerical values. Normalization (scaling to a 0-1 range) and standardization (scaling to a mean of 0 and a standard deviation of 1) are the most common techniques [3, 21].

- Addressing Class Imbalance: Congenital heart defects are, by definition, a rare anomaly. A dataset for CHD prediction will therefore be heavily skewed, with a very small number of positive cases compared to a large number of negative cases. If not addressed, a model trained on such data will achieve a high accuracy simply by predicting the majority class ("no CHD") every time, rendering it clinically useless. To mitigate this, several strategies are employed. Oversampling techniques, such as the Synthetic Minority Over-sampling Technique (SMOTE), create synthetic examples of the minority class to balance the dataset. Conversely, undersampling involves randomly removing instances from the majority class. Other methods include using algorithms with built-in class weighting or adjusting the decision threshold to favor the detection of the minority class [19].

3. Machine Learning Models for CHD Prediction

The literature on ML for CHD prediction has explored a wide range of algorithms, from interpretable linear models to complex deep learning architectures. The choice of model is often a trade-off between predictive power and interpretability.

- Simple Models:

  - Logistic Regression: This model serves as an important baseline. It is highly interpretable, as the coefficients for each feature provide a clear indication of its impact on the probability of the outcome [9]. However, its linear nature limits its ability to capture the complex, non-linear relationships that are characteristic of biological data.

  - Naïve Bayes: A probabilistic classifier that assumes all features are independent of one another. While computationally efficient, this strong assumption is often violated in clinical data, where features like smoking status and lung function are inherently correlated [17].

- Ensemble Methods: These are consistently shown to be among the most effective models for CHD prediction due to their ability to combine the predictions of multiple "weak" learners to form a single, robust "strong" learner.

  - Decision Trees: A single decision tree model is simple to interpret but prone to overfitting.

  - Random Forest: This model mitigates the overfitting problem of single decision trees by building an ensemble of trees. It introduces randomness by training each tree on a random subset of the data and using a random subset of features for each split. The final

prediction is a vote among all trees in the forest, which significantly improves accuracy and generalization [10, 19].

○ Gradient Boosting Machines (GBMs): Algorithms like LightGBM are powerful boosting models that build trees sequentially. Each new tree in the ensemble is trained to correct the errors of the preceding trees, resulting in a highly accurate and efficient model [24]. GBMs are often considered state-of-the-art for tabular datasets and have demonstrated superior performance in many health prediction tasks.

● Neural Networks:

○ Artificial Neural Networks (ANNs): Composed of layers of interconnected neurons, ANNs can learn complex, non-linear relationships. Their architecture allows them to automatically discover latent features within the data, which is highly beneficial for complex medical conditions. Deep learning, which uses neural networks with multiple hidden layers, extends this capability to handle even more complex data types like medical images or genetic sequences [3, 21]. The integration of deep learning with fetal echocardiography images, for example, is a promising area for enhancing predictive accuracy [22].

4. Evaluation Metrics for Clinical Utility

In a medical context, a model's performance cannot be judged by a single metric. The high stakes associated with a missed diagnosis (a false negative) for CHD necessitate a comprehensive evaluation framework.

● Accuracy: While seemingly straightforward, accuracy can be misleading in the context of class imbalance. A model that predicts "no CHD" for all instances might achieve 99% accuracy but would be clinically worthless.

● Precision and Recall: These metrics are essential for understanding a model's true performance. Precision measures the proportion of positive predictions that were actually correct, minimizing false alarms. Recall (or sensitivity) measures the proportion of actual positive cases that were correctly identified. In prenatal CHD screening, high recall is often prioritized to ensure that as many at-risk fetuses as possible are identified, even at the cost of a higher false positive rate.

● F1-Score: The F1-score provides a balanced measure by taking the harmonic mean of precision and recall. It is a useful metric when there is an uneven class distribution and a balance between precision and recall is desired.

● Area Under the Receiver Operating Characteristic Curve (AUC-ROC): The AUC-ROC is a comprehensive and threshold-independent measure of a model's ability to discriminate between positive and negative cases. It plots the true positive rate against the false positive rate across all possible classification thresholds. An AUC-ROC score of 1 indicates a perfect model, while a score of 0.5 indicates a model with no discriminatory power. AUC-ROC is particularly valuable because it is not affected by class imbalance and provides a robust measure of overall model performance [19].

5. The Role of Explainable AI (XAI)

For machine learning to be successfully integrated into clinical practice, it must move beyond simply providing a prediction. Clinicians must be able to understand the rationale behind a model's output. This is the domain of Explainable AI (XAI). While models like Logistic Regression are inherently interpretable, complex ensemble methods and neural networks are often considered "black boxes." XAI techniques, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), provide methods to interpret these complex models by attributing the prediction to the influence of specific input features [26]. For example, a SHAP value can quantify how much a feature, like maternal age, contributed to a high-risk prediction for a particular patient. This interpretability not only builds trust with clinicians but can also lead to new clinical insights by highlighting unexpected feature importance, which can then be investigated through traditional research methods.

RESULTS

The application of machine learning to identify maternal risk factors for congenital heart disease has yielded a consistent body of evidence, highlighting the superior predictive capabilities of advanced models and providing valuable insights into the complex etiology of CHDs. This section synthesizes the key findings from the literature, focusing on model performance, the identification of influential risk factors, and the critical role of data quality.

1. Comparative Performance of Machine Learning Models

The most prominent finding across numerous studies is the a clear performance hierarchy among different classes of machine learning algorithms. Advanced models, specifically ensemble methods and neural networks, have been consistently shown to outperform simpler models like Logistic Regression and Naïve Bayes when predicting CHD risk [10, 19, 24].

In their comparative study, Luo et al. (2017) demonstrated that the Random Forest algorithm achieved a significantly higher AUC-ROC score and F1-score compared to both Logistic Regression and a single

Decision Tree [10]. The strength of the Random Forest model was attributed to its ability to mitigate overfitting and capture the subtle, non-linear interactions between multiple maternal factors. By aggregating the predictions from numerous decision trees, each trained on a different subset of the data and features, the model developed a more robust and generalized understanding of the risk landscape. This finding is echoed in other research, where ensemble methods such as Gradient Boosting have also shown state-of-the-art performance [24]. These models excel at handling high-dimensional data and are particularly effective at identifying complex decision boundaries, which is crucial for a multifactorial condition like CHD.

Neural networks have also demonstrated strong predictive power, especially when dealing with large, complex datasets [3, 21]. Their ability to automatically learn and represent intricate features within the data makes them powerful tools. For instance, a recent study explored the use of machine learning models to screen for CHDs using fetal echocardiography images, a task where deep learning models, a class of neural networks, are particularly well-suited for image analysis [22]. The high recall rates achieved by these models are of paramount importance in a clinical setting, as they minimize the number of false negatives and ensure that as many at-risk pregnancies as possible are identified for further clinical attention [10, 19].

2. Identification and Ranking of Key Maternal Risk Factors

Beyond providing accurate predictions, machine learning models serve as powerful tools for feature importance analysis, offering insights into which maternal factors are most predictive of CHD in offspring. These models can confirm the importance of well-known risk factors while also highlighting the relative contribution of lesser-known or interacting variables.

Factors that are consistently identified as highly influential in predictive models include:

- Maternal Age: Across multiple studies, advanced maternal age is consistently ranked as a top predictor [10, 19]. ML models can precisely quantify the contribution of age as a continuous variable, revealing non-linear risk increases at different age ranges that might be missed by simple linear models.

- Maternal Medical History: Pre-existing conditions such as diabetes mellitus, which are known teratogens, are powerful predictors [7]. ML models can effectively capture the subtle ways in which various comorbidities might interact. For example, the combination of maternal diabetes with a specific medication or environmental exposure might create a synergistic increase in risk that is beyond what a simple correlation analysis would suggest.

- Lifestyle and Environmental Exposures: Machine learning has been instrumental in quantifying the impact of lifestyle and environmental factors. Studies have shown that models can effectively weigh the importance of environmental exposures, such as those related to a mother's geographic location, confirming and expanding upon the findings of traditional epidemiological studies [7, 16]. Similarly, the models consistently identify factors such as smoking during pregnancy as significant contributors to the risk profile, even when considered alongside numerous other variables.

The ability of algorithms to rank features by their predictive power, a functionality provided by models like Random Forest and Gradient Boosting, gives researchers a valuable tool for directing future etiological research. This data-driven approach can help focus investigations on the most influential factors, leading to a more efficient discovery of the underlying biological mechanisms.

3. The Impact of Data Preprocessing and Quality

The literature consistently emphasizes that the reported high performance of ML models is contingent upon a robust and meticulous data preprocessing pipeline. The quality of the input data and the strategies used to handle its inherent challenges are paramount [19].

- Addressing Class Imbalance: The rarity of CHDs means that datasets are severely imbalanced. Studies that fail to address this issue often produce models with deceptively high accuracy but poor recall, making them unsuitable for clinical use. The successful application of oversampling techniques like SMOTE or the use of class-weighted algorithms is a key factor in achieving reliable and clinically relevant results, particularly in terms of maximizing the recall of the minority class [10, 19].

- Feature Engineering: The creation of meaningful features from raw data, such as a patient's a-priori risk score or a summary of their medication history, is crucial for model performance. The success of a model is often tied to the creativity and clinical insight applied during the feature engineering phase.

In summary, the results from the collective body of research demonstrate that machine learning models offer a superior and more nuanced approach to prenatal CHD risk prediction. The consistent outperformance of advanced models, coupled with their ability to provide data-driven insights into the most important risk factors, establishes a compelling case for their integration into clinical practice, provided that the critical steps of data preprocessing are diligently executed.

DISCUSSION

The preceding analysis has established machine learning as a powerful and promising tool for transforming prenatal risk assessment for congenital heart disease. By leveraging its unique capabilities, ML offers a pathway to a more accurate and proactive approach to maternal-fetal medicine. This discussion synthesizes these findings, explores their profound clinical implications, critically examines the inherent challenges, and outlines essential future directions for the field.

1. Synthesis and Clinical Interpretation of Findings

The evidence presented underscores a fundamental shift in our ability to analyze the multifactorial etiology of CHDs. Traditional statistical models, constrained by assumptions of linearity and a limited number of variables, have provided a foundational understanding of key risk factors. However, machine learning, particularly through the use of ensemble methods and neural networks, takes this understanding a step further by modeling the complex, non-linear interactions among a vast number of variables. This allows for the construction of a more holistic and personalized risk profile for each expectant mother.

The consistent outperformance of these advanced models in terms of AUC-ROC and recall is not just a technical victory; it has direct and meaningful clinical implications. A higher recall rate, in particular, means that a greater proportion of at-risk pregnancies will be correctly identified. In the context of CHD, where a false negative can have life-threatening consequences, this is a critical outcome. The model's ability to effectively stratify risk provides a powerful, data-driven complement to a clinician's expert judgment, enabling more precise and proactive clinical decision-making.

2. Clinical Implications for a New Standard of Care

The practical application of these predictive models has the potential to fundamentally reshape prenatal care.

- Targeted Counseling and Patient Education: For expectant parents identified as being at high risk by an ML model, clinicians can provide personalized and targeted counseling [25]. Instead of generic advice, a high-risk score can initiate a specific conversation about the identified risk factors (e.g., maternal diabetes, certain environmental exposures) and actionable steps to mitigate them. This proactive approach empowers parents with knowledge and prepares them for the potential challenges ahead.

- Optimized Prenatal Monitoring and Screening: A high-risk prediction can serve as a trigger for a more intensive and tailored monitoring schedule. For instance, a model could recommend a specific patient undergo fetal echocardiography earlier or more frequently than a patient with a low-risk score [22]. This ensures that precious healthcare resources are allocated efficiently, directing advanced screening and diagnostic tools to the patients who are most likely to benefit, thereby increasing the chances of an early diagnosis.

- Planned Deliveries at Specialized Centers: The timely diagnosis of complex CHDs is not only a matter of clinical curiosity but a matter of life and death. Infants with certain critical CHDs require immediate, specialized intervention upon birth. A prenatal diagnosis, facilitated by an accurate ML model, allows for a planned delivery at a tertiary care center with a dedicated pediatric cardiology team and neonatal intensive care unit [5, 23]. This circumvents the need for a potentially dangerous postnatal transfer and has been repeatedly correlated with superior neonatal outcomes and improved survival rates.

3. Overcoming Challenges and Limitations

Despite the promise, several significant challenges must be addressed before ML models can be seamlessly integrated into routine clinical practice.

- Data Privacy and Ethical Governance: The development of these models relies on the use of vast, sensitive patient data from EHRs, which raises profound ethical and privacy concerns. Ensuring the security and anonymity of this data is paramount [25]. Future work must explore and implement robust data governance frameworks, including the use of privacy-preserving techniques such as federated learning, which allows models to be trained on decentralized data without it ever leaving the local institution.

- Model Interpretability and Clinical Trust: The "black box" nature of complex models, particularly deep neural networks, can be a major barrier to adoption. Clinicians, and their patients, need to understand the rationale behind a high-risk prediction. This is where the development and application of Explainable AI (XAI) techniques are not merely a research curiosity but a clinical necessity [26]. By providing transparent explanations of a model's output, XAI can build the trust required for a clinician to confidently use the tool in their practice.

- Data Heterogeneity and Model Generalizability: The majority of current studies are retrospective and rely on data from a single hospital or a specific population. This raises legitimate concerns about the generalizability of these models to different populations, which may have distinct genetic and environmental risk profiles. A model trained on a predominantly urban population, for example, may not perform well in a rural setting with different environmental exposures. Future research must focus on building large, diverse, multi-center datasets to train models that are robust and generalizable across

different demographic and geographic contexts [19].

- Regulatory and Clinical Integration Hurdles: The path to regulatory approval for ML-based medical devices is still being defined. Furthermore, integrating these tools into the existing clinical workflow and IT infrastructure of hospitals presents a significant logistical and technical challenge.

4. Future Directions

The future of machine learning in prenatal CHD risk prediction is bright and extends far beyond the current state-of-the-art.

- Multimodal Data Integration: The next wave of innovation will involve integrating data from a multitude of sources to create an even more comprehensive risk profile. This includes combining EHR data with genomic data (to identify genetic predispositions), fetal imaging data (using deep learning to automatically detect cardiac anomalies from ultrasound or echocardiography), and data from wearable sensors (to monitor maternal health in real-time) [20, 22]. This multimodal approach promises a more holistic and accurate predictive model.

- Prospective Clinical Validation: To move beyond theoretical efficacy, it is essential to transition from retrospective studies to rigorous prospective, randomized clinical trials. These trials would compare the outcomes of a standard care group to a group where ML predictions guide clinical decisions. Such validation is the gold standard for evidence-based medicine and is a prerequisite for widespread adoption [2, 8].

- Development of Ethical Frameworks: As ML becomes more integrated, there is a pressing need for clear ethical guidelines to govern its use. This includes addressing issues of potential model bias, ensuring algorithmic fairness, and establishing clear lines of accountability when a model's prediction leads to a specific clinical outcome.

CONCLUSION

In conclusion, machine learning offers a powerful and transformative approach to prenatal CHD risk prediction. By moving past the limitations of traditional methods, it allows for the creation of more accurate and personalized risk profiles. While significant challenges related to data privacy, interpretability, and clinical integration remain, the continued evolution of ML techniques, coupled with a concerted effort to address these hurdles, will pave the way for a new era of proactive, data-driven prenatal care, with the ultimate goal of improving the lives of infants born with congenital heart defects.

REFERENCES

Ali, F. S. A., Al Hammadi, S. A. A., Redouane, A., & Tariq, M. U. (2021). Prediction of congenital heart diseases in children using machine learning. Journal of Management Information and Decision Sciences, 24, 1-34. https://www.abacademies.org/articles/prediction-of-congenital-heart-diseases-in-children-using-machine-learning.pdf

Chang, J. J., Binuesa, F., Caneo, L. F., Turquetto, A. L. R., Arita, E. C. T. C., Barbosa, A. C., Da Silva Fernandes, A. M., Trindade, E. M., Jatene, F. B., Dossou, P., & Jatene, M. B. (2020). Improving preoperative risk-of-death prediction in surgery congenital heart defects using artificial intelligence model: A pilot study. PLoS One, 15(9), e0238199. https://doi.org/10.1371/journal.pone.0238199

Dangare, C., & Apte, S. (2012). A data mining approach for prediction of heart disease using neural networks. International Journal of Computer Engineering and Technology, 3(3). https://ssrn.com/abstract=2175569

Domyati, A., & Memon, Q. (2022). Robust detection of cardiac disease using machine learning algorithms. In Proceedings of the 5th International Conference on Control and Computer Vision (pp. 52-55). https://doi.org/10.1145/3561613.3561622

Desai, H., Jones, C. E., Fogel, J. L., Negrin, K. A., Slater, N. L., Morris, K., Doody, L. R., Engstler, K., Torzone, A., Smith, J., & Butler, S. C. (2023). Assessment and management of feeding difficulties for infants with complex CHD. Cardiology in the Young, 33(1), 1-10. https://doi.org/10.1017/S1047951122004024

Hammoud, A., Karaki, A., Tafreshi, R., Abdulla, S., & Wahid, M. (2024). Coronary heart disease prediction: A comparative study of machine learning algorithms. Journal of Advances in Information Technology, 15(1), 27-32. https://doi.org/10.12720/jait.15.1.27-32

Liu, S., Liu, J., Tang, J., Ji, J., Chen, J., & Liu, C. (2009). Environmental risk factors for congenital heart disease in the Shandong peninsula, China: A hospital-based case-control study. Journal of Epidemiology, 19(3), 122-130. https://doi.org/10.2188/jea.JE20080039

Lo Muzio, F. P., Rozzi, G., Rossi, S., Luciani, G. B., Foresti, R., Cabassi, A., Fassina, L., & Miragoli, M. (2021). Artificial intelligence supports decision making during open-chest surgery of rare congenital heart defects. Journal of Clinical Medicine, 10(22), 5330. https://doi.org/10.3390/jcm10225330

[9] Lopez, D., & Manogaran, G. (2018). Health data analytics using scalable logistic regression with stochastic gradient descent. International Journal of Advanced Intelligence Paradigms, 10(1/2), 118. https://doi.org/10.1504/IJAIP.2018.10010530

Luo, Y., Li, Z., Guo, H., Cao, H., Song, C., Guo, X., & Zhang, Y. (2017). Predicting congenital heart defects: A comparison of three data mining methods. PLoS One, 12(5), e0177811. https://doi.org/10.1371/journal.pone.0177811

Meda, J. T., & Mushiri, T. (2020). Predicting congenital heart diseases using Machine learning. In Proceedings of the 2nd African International Conference on Industrial Engineering and Operations Management (pp. 1716-1725). https://www.ieomsociety.org/harare2020/papers/415.pdf

Mello, C. A., Lewis, R., Brooks-Kayal, A., Carlsen, J., Grabenstatter, H., & White, A. M. (2014). Supervised learning for the neurosurgery intensive care unit using single-layer perceptron classifiers. In Ślęzak, D., Tan, A.-H., Peters, J. F., & Schwabe, L. (Eds.), Lecture notes in computer science: Vol. 8609. Brain informatics and health: BIH 2014 (pp. 231-241). Springer. https://doi.org/10.1007/978-3-319-09891-3_22