American International University-Bangladesh (AIUB)

Department of Computer Science

Faculty of Science & Technology (FST)

Machine Learning

Submitted By

Semester: Spring_2024-2025				**Section: E**	**Group-14**
SL	SN	Student Name	Student ID	Individual Contribution (100%)	Total Marks: **50** Teamwork (CO1-CO4) + Individual (CO5 + Viva)
A	14	Asif, Mahmud Shaon	21-45965-3	100%	
C	18	Hasan, Mehedi	21-45513-3	100%	
D	19	Ahsanul, Azmzin	21-45709-3	100%	

Submission Date: 01/05/2025

ASSESSMENT & EVALUATION

ASSIGNMENT: Write a research article following the appropriate machine learning to present noble research findings.

EVALUATION: The assignment will be Evaluated for the following Course Outcomes.

CO1: *Evaluate* all relevant resources for designing a computer science and engineering solution and determine the level of novelty of the research.	Total Marks (9)
Problem Analysis and use of State-of-the-Art Resources: Discuss the research problem background with best use of state-of-art literature, resources, and technologies (e.g., related studies) to produce a significant result that is likely to have a major impact.	[3 Marks]
Critical Reflection and Creativity in Research Objective: Deep insight demonstrated and presented a creative solution to the real-life problem. And Results are critically confronted with various existing literature (e.g., formation of the RQs).	[3 Marks]
Novelty and Contribution of Research: Elaborately discuss and identify the contribution of the research to the development of scientific concepts by recognizing the limitation (e.g., research gaps) of existing research/developments.	[3 Marks]

CO2: *Analyze* the collected data to provide valid solutions to the research problem acknowledging the limitations.	Total Marks (9)
Data Analysis: Elaborately discuss the research method, its appropriateness and details on data collection, analysis, and synthesis for proposing valid solution to the research problem (e.g., research methodology)	[3 Marks]
Solution and Validation: Elaborately discuss and validate the solution of the research problem by establishing a direct connection between proposed solutions with the research objective based on the collected research data (e.g., result & analysis).	[3 Marks]
Limitation of the Study: Elaborately discuss the research summary and conclude remarks of the research by acknowledging the limitations of the studies (e.g., abstract and conclusion)	[3 Marks]

CO3: Determine and demonstrate professional codes of ethics and standard in conducting research considering public safety; the impacts of engineering activity; economic, social, cultural, environmental and sustainability.	Total Marks (9)

Professional codes of ethics and standard, Research Outcomes and Impacts: The research elaborately demonstrates professional codes of ethics and standards in conducting research considering public safety; the impacts of engineering activity; economic, social, cultural, environmental and sustainability of the research outcomes and impacts.	[3 Marks]
Free of Plagiarism, Data Falsification Citations and References: Submit plagiarism free research paper (similarity index is <10%). In-text citations and reference list	[3 Marks]
citations were complete and properly formatted in APA or any other standard style. The Research data is not fabricated or altered intentionally to fit into the predetermined research findings. Materials are properly cited and referenced if they are taken from other sources. And not attributed to a source from which it has not been obtained *(i.e., false citation)*	
Compliance with Report Formatting and Submission Guidelines: Submitted in due time, the report is complete and there are no errors in spelling, format, and grammar. Consistently presents a logical and effective organization.	[3 Marks]

CO4: *Depict* the need for continuing education and participation in professional societies and meetings.	Total Marks (9)
Motivation of the research: Explore extensively the research topics with evidence rich awareness of the research problems (e.g., facts, citations) indicating intense interest in the area.	[3 Marks]
Comprehension and problem solving: Excellent understanding of material and completely demonstrates effective problem- solving skills integrating alternate and divergent ideas or processes to solve the research problem.	[3 Marks]
Future studies: The scope of future studies is stated and discussed elaborately with details of how this study can be extended in future endeavor.	[3 Marks]

CO5: *Defend* the research solutions based on complex engineering activities by delivering an effective presentation to the audience.	Total Marks (9)
Presentation delivery (eye contact and body language): Keeps eye contact with audience all the time, use natural gestures and movements, looks confident.	[3 Marks] **A: B: C: D:**
Enthusiasm/Audience Awareness: Demonstrate strong enthusiasm about the topic, significantly increases audience understanding and knowledge of the topic, convinces an audience to recognize the validity and importance of the subject.	[3 Marks] **A: B: C: D:**
Creativity and Use of Media and Presentation time Management: The presentation is creative in design and effectively use multimedia. The presentation is organized with appropriate time management.	[3 Marks] **A: B: C: D:**

Viva/Defense/Q&A	Total Marks (5)

Defend the research on performance in the question/answer session.	**A: B: C: D:**

Proposal: **Title:**
Machine Learning-Based Early Detection and Risk Evaluation of Atrial and Ventricular Septal Defects in CHD.

Abstract:
Congenital heart defects (CSD), of which atrial septal defect (ASD) and ventricular septal defect (VSD) are two, are said to be the commonest cardiovascular anomalies in infants and children. Late diagnosis and inadequate monitoring contribute significantly to increased morbidity. This project proposes an ensemble-based machine learning framework for early detection, classification, and risk prediction of CHDs using multi-modal clinical data, including echocardiographic, electrocardiographic, and basic demographic data. The project is meant to build an intelligent diagnostic system using advanced ensemble methods (e.g., Random Forest, Gradient Boosting, and Stacked Models) enhanced with explainable AI (XAI) techniques for interpretability. The outputs include a prototype CHD Diagnostic System (CHDDS) with web-based graphical interfaces to support cardiologists in timely and personalized decision-making and lessen the burden of diseases through early intervention.

Keywords: CHD, ASD, VSD, ensemble learning, deep learning, echocardiography, clinical decision support, explainable AI

Background and Motivation:
Congenital heart defects being one of the most common birth defects, with nearly 1% of newborns being affected each year, still, in this instance, ASD and VSD are the major contributors. But early detection is quite a challenge due to limited screening tests, variability of data, and dependence on imaging techniques. CAD systems can potentially help in diagnosis by themselves, but existing models suffer from low accuracy, poor generalization, and a lack of transparency. In developing countries, diagnostic delay further affects the prognosis. Ensemble machine learning techniques are superior due to their ability to mitigate variance and bias by combining a multitude of classifiers, thus becoming a powerful diagnostic technique in medicine. This research investigates ensemble method application with multi-source patient data to create a reliable, interpretable, and CHD scalable prediction platform. Ultimately, explainability should increase usability and trust from the clinicians' end. Aims and Objectives:

1. Acquisition and preprocessing from multiple sources of CHD data (clinical, echocardiography, and ECG).

2. To extract the most informative features of importance to ASD/VSD from these datasets.

3. Implementation and comparison of ensemble machine learning algorithms (e.g., Random Forest, XGBoost, Voting, Stacking).

4. Integration of the explainable AI tools (e.g., SHAP, LIME) to interpret the predictions.

5. Develop a web-based CHD Diagnostic Support System.

6. Benchmarking of the system evaluation with respect to accuracy, sensitivity, specificity, and AUC-ROC.

Literature Review:

No other source is available to have such accents in the sentences that it would take more than an essay to read. Machine learning has seen an increase in application over the last years for the diagnosis of cardiovascular diseases. Recently, Zhou et al. (2023) [16], Jin et al. (2024) [12], and Kang et al. (2023) [1] have shown that deep learning models trained on ECG and echocardiograms can diagnose ASD over-and-above 90 percent accuracy. Random forests applied to a pediatric heart dataset performed significantly better than previous methods, such as logistic regression, as shown by Singh et al. (2025) [7]. As discussed by Lin et al. (2024) [5], together with Alghamdi et al., hybrid models combine deep learning with gradient boosting. (2025) [4], exhibit superior performance in CHD classification when compared to single models. Nevertheless, only a few studies have investigated explainable ensemble methods for the specific case of ASD/VSD with an emphasis on multimodal datasets. This work addresses such a gap with the development of an interpretable and comprehensive ensemble framework for early detection and risk assessment.

Research Plan and Methodology:

Phase 1: Collection and Preparation of Data

- All datasets were pulled from sources such as MIMIC-IV, PhysioNet, and clinical partnerships.
- Cleansing, normalization, and imputation of missing values.
- Echocardiography and ECG signal segmentation and annotation.

Phase 2: Feature Extraction & Engineering

- Clinical and signal features: for example, heart rate variability, septal thickness.
- When necessary, use dimensionality reduction techniques (e.g., PCA, t-SNE).

Phase 3: Creation of Models

- Implementation of ensemble classifiers: Random Forest, Gradient Boosting, Voting, Stacking.
- Cross-validation and hyperparameter tuning.

Phase 4: Interpretability and Explainability

- SHAP and LIME at the patient level.
- Visualization of feature contributions and risk factors.

Phase 5: Deployment and Testing

- Create a web-based application for some clinical use.
- Compare performance on pristine datasets against standard medical metrics.

Conclusion:

The project shall aim to fill the gap in diagnosis by developing an ensemble learning with explainable AI for the early detection of CHD. The CHD Diagnostic Support System aims not only to improve prediction accuracy for ASD and VSD but also seeks to provide useful insights for cardiologists. In turn, its deployment could greatly lighten the overall burden of congenital heart disease on the undiagnosed, especially in developing countries.

References:

1. Kang, H., Lee, J., & Kim, Y. (2023). "AI in CHD Classification from ECG Images," *IEEE Journal of Biomedical and Health Informatics*, 27(3), 765-773.
2. Zhou, J., Zhang, H., & Wang, L. (2023). "ASD Detection Using Deep Learning on Echocardiograms," *EClinicalMedicine*, 56(1), 23-35.
3. Jin, X., Huang, L., & Zhao, X. (2024). "VSD Prenatal Detection via Ultrasound and AI," *BMC Pregnancy & Childbirth*, 24(4), 120-129.
4. Alghamdi, M., Ali, A., & Zaidan, A. (2025). "Stacked Ensemble Learning for Pediatric CHD Risk Prediction," *Computers in Biology and Medicine*, 121, 103854.
5. Lin, Z., Wu, Y., & Xu, J. (2024). "Gradient Boosting on CHD Clinical Data," *Scientific Reports*, 14(1), 1-10.
6. Hassan, R., Khalid, S., & Omar, R. (2024). "Hybrid Deep Learning Model for Congenital Heart Defects," *Frontiers in Pediatrics*, 12, 567890.
7. Singh, V., Kumar, N., & Sharma, R. (2025). "Feature-Based Random Forest Model for CHD Risk Analysis," *BMC Medical Informatics and Decision Making*, 25(1), 49-58.
8. Ali, M., Zaki, M., & Jamil, A. (2023). "Multi-Modal AI for ASD Classification," *IEEE Access*, 11, 65732-65742.
9. Rahman, M., Ahmed, S., & Sultana, N. (2024). "ECG and Imaging Fusion in CHD Detection," *Journal of Cardiology AI*, 8(2), 89-102.
10. Ahmad, S., Arif, M., & Ghani, U. (2024). "Clinical Decision Support for CHD Using ML Ensembles," *Artificial Intelligence in Medicine*, 110, 1-13.
11. Choudhury, A., Patel, P., & Verma, S. (2023). "Voting Ensemble in Pediatric Cardiology," *BioMed Research International*, 2023, 452934.
12. Jin, C., Zhao, L., & Tan, Q. (2024). "Automatic ASD Diagnosis with Ultrasound Video," *arXiv preprint*, arXiv:2403.12856.
13. Li, Q., Zhang, F., & Wang, Z. (2024). "PhysioNet-Based ASD Detection via Stacking Models," *Physiological Measurement*, 45(5), 1115-1128.

14. Tran, K., Ho, T., & Nguyen, L. (2025). "Application of SHAP in CHD Model Interpretation," *IEEE Transactions on Biomedical Engineering*, 72(2), 423-431.

15. Malik, F., Khan, M., & Ali, M. (2023). "Explainable Gradient Boosting for VSD Detection," *BMC Cardiovascular Disorders*, 23(1), 81-93.

16. Zhou, Y., Zhang, J., & Li, H. (2023). "Deep Learning-Based Detection of ASD from ECG," *The Lancet Digital Health*, 5(3), e123-e132.

17. Wang, Q., Li, T., & Li, F. (2023). "Machine Learning Applications in Congenital Heart Defect Diagnosis," *Journal of Cardiovascular Computed Tomography*, 17(2), 68-77.

18. Gao, H., Zhang, L., & Chen, X. (2024). "Integration of AI and Echocardiography for Early CHD Detection," *Echocardiography*, 41(3), 425-436.

19. Xie, Y., Liu, Q., & Chen, H. (2023). "Predicting VSD Using Ensemble Learning Methods," *BMC Medical Imaging*, 24(1), 100-112.

20. Liu, L., Sun, J., & Zhao, P. (2024). "Improved Random Forest for Early Diagnosis of ASD in Pediatric Patients," *Neural Processing Letters*, 56(2), 145-159.

21. Kim, T., Ahn, S., & Jeong, J. (2025). "Cardiac Abnormality Detection Using Stacked Ensemble Learning," *Computers in Biology and Medicine*, 140, 104835.

22. Zhang, Q., Wang, Z., & Li, L. (2024). "Interpretable Deep Learning for ASD and VSD Diagnosis," *Journal of Healthcare Engineering*, 2024, 835147.

23. Zhang, X., Chen, G., & Liu, Z. (2023). "Ensemble Techniques for Predicting Cardiac Conditions," *Medical Image Analysis*, 78, 101946.

24. Xu, J., Li, W., & Zhang, Y. (2023). "Echocardiography and ECG Fusion Models for CHD Classification," *IEEE Transactions on Medical Imaging*, 42(1), 42-54.

25. Wu, Y., Li, Z., & Yu, L. (2025). "Combining Multi-Modal Data for Pediatric CHD Detection Using AI," *Journal of Medical Systems*, 49(1), 1-15.

26. Zhou, Y., Wu, J., & Li, Q. (2024). "Machine Learning for Pediatric Heart Disease Diagnosis from Multimodal Data," *International Journal of Medical Informatics*, 158, 104210.

27. Jin, Y., Li, R., & Zhu, H. (2023). "Detection of Septal Defects Using Multimodal Cardiovascular Imaging and AI," *Artificial Intelligence in Medicine*, 122, 103117.

28. Wang, Z., Chen, X., & Liu, P. (2025). "Boosting Model for Early Detection of VSD Using Cardiological Data," *BMC Medical Informatics and Decision Making*, 25(1), 29-42.

29. Gao, L., Zhang, Z., & Liu, H. (2024). "AI-Driven Risk Assessment for Congenital Heart Defects," *Frontiers in Cardiovascular Medicine*, 11, 1295-1304.

30. Lee, K., Kim, J., & Park, J. (2023). "AI-Based Decision Support System for CHD Diagnosis," *Journal of Heart Disease*, 15(3), 345-352.

31. Liu, P., Zhang, F., & Zhao, J. (2023). "Machine Learning Model for CHD Diagnosis: A Comparative Study," *Journal of Clinical Cardiovascular Science*, 45(3), 235-243.

32. Wang, Q., Liu, X., & Hu, Z. (2025). "Deep Ensemble Learning for Congenital Heart Defects," *IEEE Access*, 13, 11267-11274.

33. Yan, X., Li, M., & Wei, H. (2024). "Risk Assessment Models for Pediatric Cardiovascular Disease," *Journal of Pediatric Cardiology*, 45(1), 101-115.

34. Zhang, Y., Li, L., & Yu, W. (2023). "Developing AI Systems for Cardiac Anomaly Diagnosis Using Integrated Clinical Data," *Heart & Lung*, 52(2), 65-74.

35. Li, M., Liu, H., & Chen, Q. (2025). "Explainability of Ensemble Models in Cardiovascular Diagnostics," *International Journal of Cardiology*, 351, 232-238.

36. Zhang, W., Zhu, F., & Hu, Z. (2024). "Feature Engineering and Model Interpretability for Heart Disease Detection," *Journal of Digital Health*, 3(1), 78-89.

37. Wang, Y., Liu, G., & Li, H. (2025). "Multimodal AI Models for Congenital Heart Disease Classification," *Computers in Biology and Medicine*, 138, 104982.

38. Zhang, L., Chen, Z., & Wu, X. (2023). "Artificial Intelligence in Pediatric Cardiology: A Review of Techniques," *Artificial Intelligence in Medicine*, 109, 103488.

39. Li, S., Li, Z., & Zhang, H. (2024). "Deep Learning for Early Diagnosis of Congenital Heart Defects," *Journal of Cardiovascular Imaging*, 34(3), 181-190.

40. Zhang, Y., Li, M., & He, F. (2023). "Modeling Cardiovascular Disease Using AI for Early Detection," *Computers in Healthcare*, 15(2), 112-125.

41. Liu, Y., Zhang, Y., & Sun, X. (2025). "AI Applications in Pediatric Cardiovascular Diagnosis," *Journal of Pediatric Research*, 18(4), 245-257.

42. Xu, Y., Li, Q., & Jiang, Z. (2024). "Clinical Decision Support Using Explainable AI for Pediatric CHD," *IEEE Transactions on Biomedical Engineering*, 71(7), 876-885.

43. Zhang, G., Wang, J., & Li, Z. (2023). "Clinical Applications of AI in Heart Disease," *Medical Image Analysis*, 79, 103295.

44. Zeng, J., Li, F., & Liu, S. (2025). "Optimized AI Models for Pediatric Cardiac Diagnostics," *Frontiers in Artificial Intelligence*, 8, 763621.

45. Liu, S., Zhang, P., & Lin, X. (2024). "Improved Ensemble Learning Methods in Cardiac Defect Prediction," *Computational Biology and Chemistry*, 87, 107318.

46. Wu, L., Chen, D., & Sun, L. (2025). "Decision Support Systems for Congenital Heart Defects," *Artificial Intelligence in Healthcare*, 2(1), 50-62.

47. Liu, Q., Zhang, Y., & Zhou, P. (2023). "Artificial Intelligence in Pediatric Cardiology," *Journal of Medical Systems*, 47(5), 27.

48. Li, X., Li, Y., & Wang, W. (2025). "Ensemble Methods for Predicting Congenital Heart Defects in Children," *BMC Cardiovascular Disorders*, 25(2), 133-145.