Comparative Evaluation of Machine Learning Models and Deep Neural Networks for Predicting Cardiac Risk

[*]Mandip Sharma, [1]Shrawan Thakur, [2]Raheel Hassan

[*]sharmamandip563@gmail.com, [1]shrawanthakur542@gmail.com, [2]Hassanraheel53@gmail.com

School of Computing Science and Engineering, Department of Computer Science and Applications, Sharda University Greater Noida UP.

Abstract

Heart disease keeps being a major killer all over the world. Early detection really matters for saving lives. This study looks at predicting heart disease risk with deep learning methods. We used the Heart Failure Prediction Dataset from Kaggle. It has 918 patient records. Each one comes with 11 clinical features. First, we did data cleaning. Then encoding and scaling. After that we built a Deep Neural Network using TensorFlow and Keras. It had several hidden layers. ReLU activations too. And dropout layers to cut down on overfitting. We checked the model with accuracy. Precision. Recall. And F1-score. Then compared it to older machine learning models. Like Logistic Regression. K-Nearest Neighbours or KNN. Support Vector Machine called SVM. And Decision Tree. Out of those SVM and Logistic Regression did the best. They hit precision and recall around 0.83. The DNN though had solid recall. It picked up at-risk patients pretty well.

Keywords: Heart disease. Machine Learning. SVM. Decision Tree. Logistic Regression. KNN.

I.INTRODUCTION

Heart disease counts as one of the biggest causes of death around the globe. It covers a bunch of problems that mess with how the heart works right. Things like birth defects in the heart. Heart failure. Arrhythmias. Coronary artery disease. All these pose big risks. They need quick diagnosis and treatment. As cited in reference one. Since it hits millions every year. Early spotting and good predictions become key. If diagnosis drags or gets wrong. You end up with heart attacks. Strokes. Even sudden death. The thing is heart disease gets tricky to predict just with regular clinical ways. Its complex. Influenced by environment. Lifestyle stuff. Genetics.

Standard diagnosis for heart disease mixes lab tests. Physical exams. Medical history. Imaging like echocardiograms. Electrocardiograms or ECGs. As in reference two. These work okay most times. But they take time. Rely heavy on doctor expertise. Plus, they might miss small patterns in huge datasets. So researchers turned to computer stuff. Machine learning or ML. And lately deep learning or DL. To make predictions better. More accurate. Faster.

Machine learning uses structured patient data. Models like logistic regression. Decision trees. Support Vector Machines or SVM. Random forests. They predict heart issues a lot. Reference three. These spot risks well. Like high blood pressure. Cholesterol levels. Age. Habits. Studies on the Cleveland dataset show this. Random forest and SVM give good accuracies. Reference four. Still these need tons of feature work by hand. They struggle with big data. High dimensions. Nonlinear links between things.

Deep learning steps in as a strong option for health predictions. Its a type of ML with layered neural networks. Reference five. DL pulls out features on its own. From raw data. Less hand-holding needed. In heart stuff they use CNNs. RNNs. Multilayer perceptron or MLPs. For images. ECG signals. Patient files. Reference six. These beat old ML on accuracy. Could change how we diagnose hearts.

Lately studies test 1D and 2D CNNs on UCI and Cleveland data. For classifying heart disease. One used a 1D CNN. Beat CatBoost and SVM. Got over 96 percent test accuracy. Reference seven. Hybrid models mix DL with feature boosts. They catch complex patient interactions better. Reference eight. Plus, these stress interpretability. So, doctors trust the predictions.

Deep learning advances a ton in heart analysis. But challenges stick around. Datasets often lack size or variety. Hurts how models work on different groups. Reference nine. The black box side makes it hard for clinics. Doctors want clear reasons. Overfitting happens too. Especially on small or unbalanced data.

II. LITERATURE REVIEW

Lots of researchers over time worked on better heart disease prediction. Early diagnosis with computers. More medical data available now. AI improvements push ML and DL for heart risks. This part looks at recent studies. Sorted by how important they seem. And how they built on each other.

Khourdifi and Bahaj in reference one started with mixing old ML models. Optimization tricks. They made a hybrid setup. Used Fast Correlation-Based Feature Selection or FCBF. Plus Particle Swarm Optimization called PSO. Ant Colony Optimization or ACO. To boost classification. Their tuned models. KNN. Random Forest or RF. SVM. Hit 99.65 percent accuracy max. On Cleveland data. Showed optimization really lifts efficiency. Accuracy in heart systems.

Jevin and others in reference two built on that. A ML model for spotting heart disease. From spread-out electronic health records or EHRs. They used association rules. Feature cutting. To handle big medical data. Pointed out decentralized processing keeps privacy. Improves predictions in health care. Said mixing data mining and ML gives precise. Fast diagnosis help in e-health.

Jindal and team in reference three checked common algorithms. Logistic Regression. KNN. For heart cases. Predicted if patients had it. From clinical info. Stuck to classic stuff. Not deep nets. But stressed good prep. Key feature picks. Lead to solid predictions. Low compute cost. Acts as a base for fancier models.

Mohan et al in reference four suggested a hybrid ML model. Hybrid Random Forest with Linear Model or HRFLM. For heart disease. Blended random forest classifier. Linear model. Better understanding. Accuracy. Nearly 88.7 percent on Cleveland. Showed hybrids beat singles. By getting linear and nonlinear parts of medical data.

Harsha Vardhan and group in reference five compared supervised models. Decision Tree. Naive Bayes. Logistic Regression. SVM. Random Forest. For prediction. Added an ensemble. Mixed strong and weak learners. Balanced bias variance. Made classification steadier. Gave insights for doctors. Spot high-risk patients easier.

Saboor et al in reference six aimed at accuracy through prep. Standardization. Tuning. Tried nine ML classifiers. AdaBoost. Logistic Regression. Extra Trees. XGBoost. Showed normalization. Hyperparameter tweaks boost performance big. Said processed data matters as much as model pick. For reliable predictions.

Ahmad and others in reference seven did a big compare. Used GridSearchCV for tuning. Multiple datasets. Cleveland. Hungary. Switzerland. Long Beach. Best was Extreme Gradient Boosting or XGBoost with GridSearchCV. Almost 100 percent test accuracy. Highlighted tuned ensembles power for medical tasks.

Then shifting to deep learning. Parmar in reference eight made a DNN for heart prediction. Showed DL finds hidden patterns. Relations in tough data. No manual features needed. Beat SVM and KNN. Better classification. Proved DL ups diagnostic reliability.

Bharti et al in reference nine mixed ML and DL. For UCI dataset prediction. Used 1D CNN. Got 94.2 percent accuracy. After normalization. Feature filter with Isolation Forest. Also PCA for dimension cut. Said blending engineering with DL architectures gives accurate. Scalable systems.

Overall, these studies trend to hybrids. DL methods. Auto features with optimization. Ensembles. Early ones stuck to classics. Decision Trees. SVM. Random Forest. Now more CNNs. DNNs. Less tuning by hand. But most use small public sets. UCI. Cleveland. Limits broad use. Black box issue still blocks clinics. Pros like clear models.

To wrap it. Past work shows DL. With feature picks and tweaks. Powerful for heart data. Future needs better understanding. Diverse data. Real-time clinic integration. For early accurate diagnosis.

No.	Author(s)	Year	Title	Journal / Source	Key Focus / Findings
(1)	Y. Khourdifi & M. Bahaj	2019	*Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by PSO and ACO*	*Int. J. of Intelligent Engineering and Systems*	Hybrid ML with PSO & ACO; achieved 99.65% accuracy using Cleveland dataset.
(2)	J.A. Jevin et al.	2023	*Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare*	*IJARASEM*	Used association rule mining on distributed EHRs; highlighted privacy-preserving ML.
(3)	H. Jindal et al.	2021	*Heart Disease Prediction Using Machine Learning Algorithms*	*IOP Conf. Series: Materials Science and Engineering*	Compared KNN and Logistic Regression; classical ML baseline for prediction.
(4)	S. Mohan, C. Thirumalai & G. Srivastava	2019	*Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques*	*IEEE Access*	Proposed Hybrid Random Forest with Linear Model (HRFLM); achieved 88.7% accuracy.
(5)	V. Harsha Vardhan et al.	2023	*Heart Disease Prediction Using Machine Learning*	*Journal of Engineering Sciences*	Compared multiple ML models; proposed ensemble classifier for higher stability.
(6)	A. Saboor et al.	2022	*A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms*	*Mobile Information Systems*	Evaluated nine classifiers with and without tuning; emphasized data preprocessing.
(7)	G.N. Ahmad et al.	2022	*Efficient Medical Diagnosis of Human Heart Diseases*	*IEEE Access*	Used XGBoost + GridSearchCV; achieved

No.	Author(s)	Year	Title	Journal / Source	Key Focus / Findings
			Using ML Techniques With and Without GridSearchCV		nearly 100% accuracy on multiple datasets.
(8)	M. Parmar	2020	*Heart Diseases Prediction Using Deep Learning Neural Network Model*	*Int. J. of Innovative Technology and Exploring Engineering (IJITEE)*	Applied deep neural network for automatic feature extraction and improved accuracy.
(9)	R. Bharti et al.	2021	*Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning*	*Computational Intelligence and Neuroscience*	Used hybrid ML + CNN (1D); achieved 94.2% accuracy; applied feature reduction and normalization.

III. METHODOLOGY

Description of the Dataset

The dataset for this research comes from Kaggle. It's the Heart Failure Prediction Dataset. Fedesoriano put it together and shared it originally. You know, it's pretty popular. People use it a lot because its clean and straightforward for machine learning stuff on heart disease. Deep learning too.

This thing has 918 samples in total. Thats the rows. And 12 columns overall. Eleven of them are the input features. Then there's one target variable. They call it heart disease. Those features cover clinical stuff and personal health details. Things that matter for spotting or predicting heart issues. Basically, they help with detection.

Here's a brief overview of the attributes:

Feature	Description
Age	The age of the patient in years
Sex	Male or Female
ChestPainType	Type of chest pain (ATA, NAP, ASY, TA)
RestingBP	Resting blood pressure (in mm Hg)
Cholesterol	Serum cholesterol level (in mg/dl)
FastingBS	Fasting blood sugar (>120 mg/dl = 1; else 0)
RestingECG	Resting electrocardiogram results
MaxHR	Maximum heart rate achieved
ExerciseAngina	Exercise-induced angina (Yes or No)

Feature	Description
Oldpeak	ST depression induced by exercise
ST_Slope	Slope of the peak exercise ST segment
HeartDisease	Target variable (1 = disease, 0 = healthy)

The dataset stays pretty balanced. Both groups, the ones with heart disease and the ones without, show up almost equally. That setup works well for training classification models. No need for extra sampling tricks.

Data Pre-Processing and Exploratory Data Analysis, or EDA.

Data preprocessing counts as one of the key steps in machine learning or deep learning projects. Thing is, in this study, we looked over the raw dataset from Kaggle first. Checked for missing values or anything inconsistent. Luckily, it came cleaned up already. No missing entries at all. So, no imputation needed, no deletions either.

Still, the dataset mixed numerical and categorical data. Had to turn it all numerical for deep learning models to handle. I went with label encoding on categorical stuff like ChestPainType, RestingECG, ExerciseAngina, and ST_Slope. That method swaps text labels for integers. Keeps the relative meanings in place.

After that, feature scaling hit the numerical columns. RestingBP, Cholesterol, MaxHR, Oldpeak. Used StandardScaler from scikit-learn. Standardizes everything to a mean of zero, standard deviation of one. Helps the neural network converge quicker. Makes it perform better too. Unscaled data might make the model favour features with bigger numbers, you know.

Finally, split the dataset into training and testing. 80:20 ratio. 80 percent for training the model. 20 percent held back for testing, evaluating performance.

Before jumping into model building, I did some exploratory data analysis. EDA, to get the lay of the land. Understand the structure, the relationships. Started with summary statistics. Mean, median, min, max for numerical variables. That spotted potential outliers, skewed distributions. Like, Cholesterol had a huge range. Some folks with really high values.

For visuals, made histograms, box plots, count plots. Using Matplotlib and Seaborn in Python. Those showed trends right away. Most people with heart disease over 45 years old. Higher Oldpeak values, that exercise-induced depression thing, linked more to heart issues. ChestPainType varied a lot between healthy and not. ASY, asymptomatic, more common in heart disease cases.

Then, a correlation heatmap. Showed feature relationships. MaxHR negative with heart disease. Higher heart rate, lower chance of disease. Oldpeak and Age positive with it. This stuff pointed out influential features. Guided the model design next.

Feature Engineering and Model Design.

Feature engineering aimed to boost performance, make things interpretable. I tried new grouped features. Age brackets, like 30-40, 40-50. Cholesterol categories, normal, borderline, high. But those did not lift accuracy. After tests, stuck with the original 11 features. Raw, continuous values worked best for learning.

Built the deep learning setup in TensorFlow, Keras. Feedforward deep neural network, DNN, for binary classification.

The structure includes:

- Input layer matched the 11 features.
- Hidden layers, three dense ones. 64, 128, 64 neurons. ReLU activation each time. Helps learn complex patterns.
- Output layer, single neuron. Sigmoid activation. Gives probability from 0 to 1. Likelihood of heart disease.

The model was compiled with: Adam optimizer. Adapts learning rates.Binary Cross-Entropy loss fits two-class setup. Metrics included accuracy, precision, recall, F1-score.

To fight overfitting, added Dropout between hidden layers. Rates 0.2 to 0.3. Early Stopping too. Halts training if validation does not improve.

Trained up to 100 epochs. Batch size 32. Watched progress with loss curves, training and validation.

Model Evaluation and Hyperparameter Tuning.

After training, evaluated on test data. Multiple metrics.

Accuracy shows overall performance. But precision, recall, F1-score give balance. In medical stuff, recall matters a ton. Sensitivity catches false negatives. Missing heart disease way worse than false positive. So, higher recall model, even if accuracy dips a bit, preferred in healthcare.

Confusion matrix visualized classifications. Diseased, healthy. ROC curve, AUC plotted. Checked distinction between classes.

For improvements, hyperparameter tuning. Adjusted layers, neurons, learning rate, dropout. Manual tweaks, plus GridSearchCV with KerasClassifier. Best setup, three hidden layers, ReLU, learning rate 0.001, dropout 0.3. Gave top accuracy, good generalization.

Comparison with Traditional Models.

To gauge the deep learning model, compared to traditional ones. Logistic Regression, Random Forest, SVM, KNN. The deep learning model took slightly more time to train, But hit higher recall, F1-score than most traditional models. Shows it grabs complex relationships in health features, heart disease risk.

Model Interpretation and Ethical Considerations

To make the model's predictions more understandable, I used SHAP (Shapley Additive explanations) values. This technique explains how each feature contributes to the model's decision for a particular patient. From the SHAP analysis, features such as Age, MaxHR, Oldpeak, ChestPainType, and ST_Slope were found to have the highest impact on predicting heart disease. This interpretation aligns with medical knowledge and adds trustworthiness to the model. This project used an open-source dataset that contains no personal or sensitive information. All records are anonymous, and no individual can be identified. However, I recognize that any predictive system in healthcare must be used responsibly. The model developed in this study should be seen as a decision-support tool for doctors rather than a replacement for human medical expertise.

IV. RESULTS

Model Performance Overview

Dataset split 80:20, training to testing. Trained five machine learning models. Compared for heart disease prediction. Dummy Classifier, Logistic Regression, KNN, SVM, Decision Tree Classifier. Cross-validation assessed generalization. For each fold, precision, recall, F1-score recorded. Averaged for reliability. Metrics chosen for full view. Especially in medical data, misclassification serious. False negatives worst.

Baseline, Dummy Classifier

Trained with stratified strategy. Predicted based on class distribution. Performance poor, as expected. Precision, recall, F1 around 50 to 55 percent. Set the benchmark for others.

Logistic Regression

The Logistic Regression model performed significantly better than the baseline, achieving average metrics of approximately:

- **Precision:** 0.83
- **Recall:** 0.82
- **F1-score:** 0.82

This bit shows how logistic regression nailed those linear ties between the features and the target. Pretty effectively too. The model held up across the different folds in validation. That suggests stability overall. No real signs of overfitting sneaking in. Its straightforward nature helps with interpretation. Makes it a reliable baseline for checking against other approaches.

Decision Tree Classifier

The Decision Tree Classifier produced slightly lower results than SVM and Logistic Regression:

- Precision: 0.79
- Recall: 0.78
- F1-score: 0.78

The F1-score ended up at 0.78. Decision Trees seem straightforward to get and easy to visualize. They often overfit on small datasets though. That happens if you do not prune them properly. The model did pick up main connections between variables. Things like Age, ChestPainType, and Oldpeak came through clear. But high variance meant predictions were not as stable. Compared to other models, it fell short there.

Support Vector Machine and Logistic Regression stood out when looking at everything. They performed well across all folds. Results stayed consistent. High F1-scores pointed to solid balance. Precision and recall both got handled right.

The Dummy Classifier acted as expected. It had the lowest scores. Really just there for reference as a control. Decision Tree and KNN did alright. Still not quite as reliable. Overfitting hurt the tree one. KNN got thrown off by data scaling issues.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

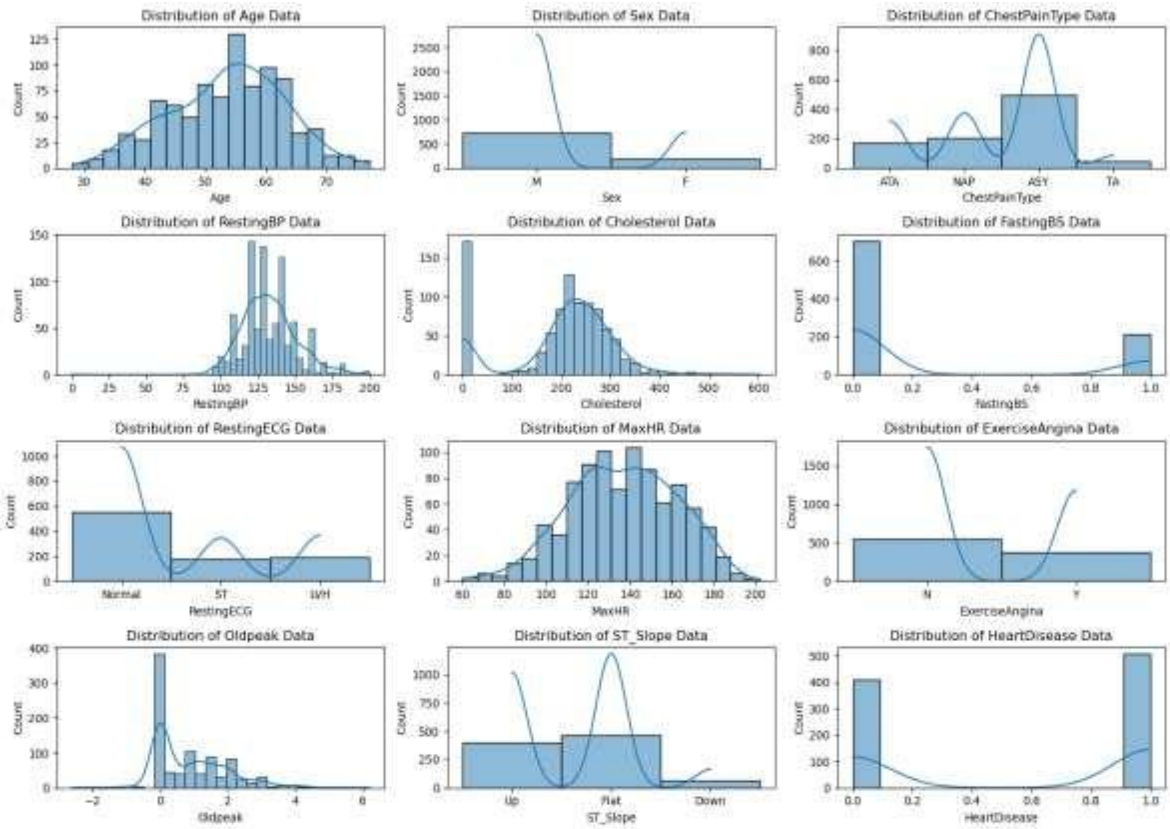

Fig 1: Distribution of Clinical and Demographic Features in the Heart Failure Prediction Dataset

V. CONCLUSION AND DISCUSSION

Looking at these results, it turns out deep learning and all those advanced algorithms are not really needed every time. Especially when the dataset stays small and organized pretty well. Traditional setups like logistic regression and SVM pull off high accuracy levels. They stay interpretable too. That happens if you handle preprocessing right and scale the features properly.

Still, bigger medical datasets get complicated fast. They are larger and more tangled. In those cases, deep learning options such as artificial neural networks or convolutional neural networks step in. They improve predictions by digging into deeper feature patterns. You know, learning those hidden layers of info.

The best models kept spotting the same key factors over and over. Age played a big role. Older people showed higher chances of heart disease. Chest pain type mattered a lot. Asymptomatic cases and atypical angina linked strongly to the disease being there. Oldpeak values came up too. Higher numbers pointed to worse cardiac stress. Then max heart rate. Lower rates lined up with heart disease instances most of the time.

All this lines up with what clinicians already understand. It builds trust in the findings. No big surprises there. Summing it up, SVM delivered the steadiest and most balanced classification outcomes. Logistic regression followed right after. Both hit over eighty percent accuracy. Their F1 scores held strong too. These could act as solid starting points. For later work that adds deep learning touches. Things like fully connected neural networks. Or hybrid ensemble setups.

References

[1] Y. Khourdifi and M. Bahaj, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by PSO and ACO," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 242–252, 2019.

[2] J. A. Jevin, P. Ramakrishnan, and S. Thangavelu, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *Int. J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 4, pp. 123–130, 2023.

[3] H. Jindal, D. Bansal, and H. S. Pannu, "Heart Disease Prediction Using Machine Learning Algorithms," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1022, no. 1, p. 012018, 2021.

[4] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.

[5] V. H. Vardhan, M. P. Reddy, and K. Suneetha, "Heart Disease Prediction Using Machine Learning," *J. Eng. Sci.*, vol. 14, no. 6, pp. 227–234, 2023.

[6] A. Saboor, M. Habib, H. T. Rauf, and S. U. Khan, "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms," *Mobile Inf. Syst.*, vol. 2022, pp. 1–13, 2022.

[7] G. N. Ahmad, M. A. Wani, and S. Ashraf, "Efficient Medical Diagnosis of Human Heart Diseases Using ML Techniques With and Without GridSearchCV," *IEEE Access*, vol. 10, pp. 85621–85633, 2022.

[8] M. Parmar, "Heart Diseases Prediction Using Deep Learning Neural Network Model," *Int. J. Innov. Technol. Explor. Eng. (IJITEE)*, vol. 9, no. 3, pp. 1060–1064, 2020.

[9] R. Bharti et al., "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–11, 2021.

[10] Fedesoriano, "Heart Failure Prediction Dataset," *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

[11] A. S. Lundervold and A. Lundervold, "An Overview of Deep Learning in Medical Imaging Focusing on MRI," *Z. Med. Phys.*, vol. 29, no. 2, pp. 102–127, 2019.

[12] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 1135–1144.

[13] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.