IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 2, 2020, accepted November 15, 2020, date of publication November 24, 2020,
date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3040166

Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values

LUJAIN IBRAHIM*, (Member, IEEE), MUNIB MESINOVIC*, (Member, IEEE), KAI-WEN YANG*, (Student Member, IEEE), AND MOHAMAD A. EID, (Senior Member, IEEE)
Division of Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Corresponding author: Kai-Wen Yang (kwy225@nyu.edu)

*Lujain Ibrahim, Munib Mesinovic, and Kai-Wen Yang contributed equally to this work.

This work was supported by the Division of Engineering at New York University Abu Dhabi.

ABSTRACT The early and accurate detection of the onset of acute myocardial infarction (AMI) is imperative for the timely provision of medical intervention and the reduction of its mortality rate. Machine learning techniques have demonstrated great potential in aiding disease diagnosis. In this paper, we present a framework to predict the onset of AMI using 713,447 extracted ECG samples and associated auxiliary data from the longitudinal and comprehensive ECG-ViEW II database, previously unexplored in the field of machine learning in healthcare. The framework is realized with two deep learning models, a convolutional neural network (CNN) and a recurrent neural network (RNN), and a decision-tree based model, XGBoost. Synthetic minority oversampling technique (SMOTE) was utilized to address class imbalance. High prediction accuracy of 89.9%, 84.6%, 97.5% and ROC curve areas of 90.7%, 82.9%, 96.5% have been achieved for the best CNN, RNN, and XGBoost models, respectively. Shapley values were utilized to identify the features that contributed most to the classification decision with XGBoost, demonstrating the high impact of auxiliary inputs such as age and sex. This paper demonstrates the promising application of explainable machine learning in the field of cardiovascular disease prediction.

INDEX TERMS Machine learning, biomedical informatics, predictive models, acute myocardial infarction.

I. INTRODUCTION

Cardiovascular diseases (CVD) are the number one cause of death globally, accounting for 31% of all deaths, with the World Health Organization reporting figures of 17.9 million per year and growing [1], [2]. With skyrocketing prevalent rates of obesity, diabetes, and other cardiovascular-related risk factors, the CVD mortality is projected to increase to more than 23.6 million annually by 2030 [3]. Among CVDs in the United States and other countries, (acute) myocardial infarctions account for the largest percentage of deaths with every 40 seconds an American suffering a myocardial infarction [4]. Countless examples show the incredible potential machine learning can play in timely detection and prediction of CVDs in the hopes of reducing mortality [5]–[7]. This includes approaches using electrocardiogram (ECG) measurements with machine learning to detect anomalies in the signal and thereby abnormalities related to specific diseases. ECG is commonly used in clinical diagnosis due to its noninvasiveness and low cost [8]. Challenges have come from the lack of multiple publicly available extensive and longitudinal databases of real-life ECG measurements as well as highly accurate (>89%) and comparatively diverse machine learning algorithms [9]–[12]. Most have recently been focused on applying deep neural networks, specifically convolutional neural networks (CNNs), to predicting arrhythmias and atrial fibrillations and thereby limiting the possibility of better performance of other algorithms on the more lethal acute myocardial infarction [13].

In this paper, we present a framework to analyze a comprehensive, longitudinal, and in the field of machine learning unexplored ECG database assembled from a 19-year study period, ECG-ViEW II [14]. This database has provided a solution to the lack of well-annotated and large datasets for the training and validation of powerful but data-hungry machine learning models. The study was collected using

The associate editor coordinating the review of this manuscript and approving it for publication was Shahzad Mumtaz.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

standard 12-lead ECGs and therefore the results of the predictive models can be compared to results obtained by others on data collected using similar hardware [14]. We have focused on the prediction of acute myocardial infarctions to expand the range of CVD diagnoses using deep neural networks, CNNs and recurrent neural networks (RNNs) while also including the decision-tree based model, XGBoost. In order to analyze feature importance that can be directly related to the impact of physiological factors on disease occurrence, Shapley Values were used with the XGBoost model. This framework can be used for early detection of AMI through incorporating it into a remote wearable device that constantly collects ECG measurements. Furthermore, the database is explored to develop strong predictive models for cardiovascular diseases and that our models and conclusions on feature importance incite further research.

II. RELATED WORK

ECG machine learning analysis explored mostly arrhythmias, atrial fibrillation, blockage, and rarely blood pressure estimation leaving out the prevalent disease of acute myocardial infarctions (AMIs), all the while relying on a limited pool of datasets for analysis, namely datasets from PhysioBank like MIT-BIH Arrhythmia database and others such as PTB Diagnostic ECG database [15]–[20]. Both datasets include 1 or 2-lead ECGs collected from a relatively small patient sample size, 47 in the case of the former, with less than a decade of longitudinal monitoring. Most recent work on ECG machine learning analysis includes using standard 12-lead ECG from a 6-year long study to present unidimensional deep and recurrent neural networks with F1 scoring of over 80% [10]. A subsequent study achieved 83% accuracy for the prediction of AMI through ECG images with CNN, which was demonstrated to be more accurate than physicians (70%) [21]. Here, we highlight a relatively unexplored ECG database assembled from the Department of Biomedical Informatics at Ajou University School of Medicine collected over a 19-year period with nearly 1 million 12-lead ECG samples documenting over 1400 separate diagnoses as well as diverse drug treatments [14]. A common challenge with detecting for incidence of acute conditions in biomedical datasets is the class imbalance often forcing the resorting to novelty detection algorithms [22]. To account for it here, we used Synthetic Minority Oversampling Technique (SMOTE) to randomly undersample the majority class and oversample the minority [23]. Analyzing this more balanced data for acute myocardial infarction presents novelties in sourcing as well as cardiovascular disease prediction.

Pending a more recent review, ECG machine learning analysis has been mostly supervised learning and within that mostly deep neural network analysis using convolutional neural networks [24]. RNNs have also been applied for arrhythmia detection using single-lead and 12-lead ECG signals [10], [25]. A more recent study has conducted a combination of unsupervised and supervised learning for ECG signal processing [26]. ECG feature extraction and engineering is also a well-studied field with various methods being developed and used over the years like artificial neural networks, principal component analysis, and wavelet transforms and support vector machines [27]–[29].

The above-mentioned methods will be applied and supplemented by decision-tree based models with XGBoost to provide more holistic investigation of this newer dataset.

To interpret machine learning performance in terms of relative feature importance, several methods are available from extracting coefficients in linear regression, through CART algorithms, to using correlation matrices and heatmaps [30]–[32]. Shapley values, a concept from game theory, allows calculation of the contribution each feature makes towards the prediction [33], [34]. By applying the Shapley value method in AMI prediction from ECG signals, we will be able to discuss the importance of physiological factors to the development of this specific cardiovascular disease.

III. THE PROPOSED FRAMEWORK

In order to streamline the procedure conducted, we propose a framework that modularizes the classification process, as illustrated in Fig. 1 (a). A dataset that meets the needs of the given task is identified and selected. Data processing, including filtering, scaling, etc., are conducted to ensure that the dataset meets the requirement of its intended use. Features that are most relevant to constructing the model are kept while irrelevant ones are abandoned. The dataset is trained using various models and their performances are evaluated and compared. Feature analysis is conducted to evaluate the relative importance of each feature, which can then help tune feature selection to achieve optimal prediction outcome. The rest of the paper includes a realization of this framework

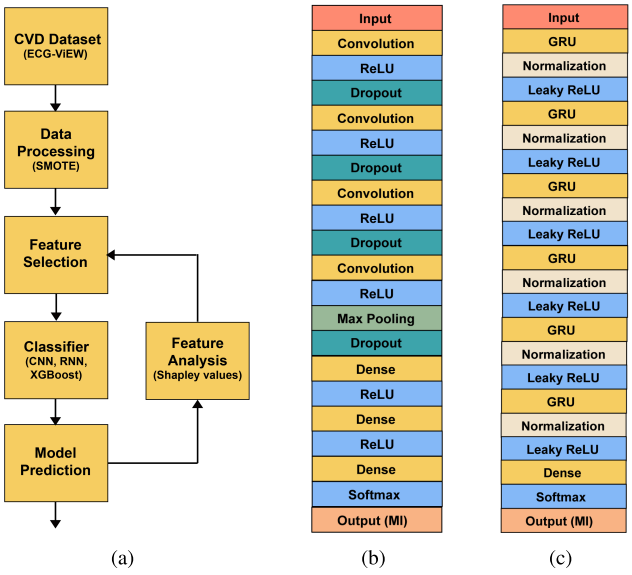


FIGURE 1. (a) The proposed framework (b) The architecture of the CNN model (c) The architecture of the RNN model.

in which a concrete implementation of these components is presented.

IV. DATA PRE-PROCESSING

The chosen dataset for this study, the Electrocardiogram Vigilance with Electronic data Warehouse (ECG-ViEW II) [14], contains 979,273 of extracted ECG measurements and other information regarding diagnoses, drug prescriptions, and selected laboratory test results collected from 371,401 patients over a period of 19 years. The ECG data, diagnosis, and personal information were matched and combined from individual files through the person ID. We decided to keep repeated measurements of the same patient to have not just the latest information but also have more data samples to learn from. Samples of incomplete extracted ECG data were excluded to ensure the quality and completeness of the dataset. Eight medical codes related to the diagnosis of acute myocardial infarction (AMI) were identified as shown in Table 1. The number in parenthesis next to the diagnosis code indicates the number of samples. Samples with these codes were labeled MI positive (N=8,395) while the rest (N=705,052) were labelled negative. The final tabular dataset consists of 12 features, including RR, PR, QRS, QT, QTc, P wave axis, QRS axis, T wave axis, Age-adjusted Charlson Comorbidity Index (ACCI), sex, age (Birthyear-group), and the label MI.

TABLE 1. The medical codes labelled MI positive in the dataset.

Diagnosis	N = 713447
Myocardial Infarction (MI)	**N = 8395**
Acute MI	DC13972 (8128)
Acute MI NOS	DC3641 (34)
Acute subendocardial MI	DC2108 (25)
	DC8232 (12)
Acute transmural MI of anterior wall	DC910 (93)
Acute transmural MI of inferior wall	DC7547 (96)
Acute transmural MI of other sites	DC2624 (0)
Acute transmural MI of unspecified site	DC2485 (7)
Non-Myocardial Infarction	**N = 705052**

The dataset was first standardized using RobustScaler, a method using statistics that are more robust to outliers. Due to the sampling bias resulted from significant class imbalance in the dataset (only 1.18% MI positive labels out of a total of 713,447 samples), Synthetic Minority Oversampling Technique (SMOTE) was applied to under-sample the majority label while over-sampling the minority [23]. The identical SMOTE sampling ratio of 25% minority and 75% majority sampling was utilized before splitting the dataset into training and validation sets (80%/20%). The identical training and testing datasets were utilized across the various ML techniques to ensure a fair comparison.

V. MODEL ARCHITECTURE

A. CNN MODEL

The high-level architecture of our CNN model is illustrated in Fig. 1 (b). The network receives the extracted ECG signal features as well as auxiliary data (age and sex) and returns a binary output where 1 indicates the user is at risk of AMI and 0 indiciates the user is not at risk of AMI. The network was implemented using Keras' neural network library and was designed following design patterns and choices in literature on the applications of CNN modeling in prediction and classification tasks (primarily on ECG data) [35]–[38].

The proposed architecture consists of 12 layers in total: 4 one dimensional convolutional layers (with kernel sizes of 5,3,3 and 3 respectively), 4 dropout layers (with a dropout rate of 10%), a global max pooling layer, and 3 dense layers. Rectified linear unit (ReLU) activation functions are used for all the convolutional layers and the first 2 dense layers, whereas a soft max activation function is used for the last dense layer. Additionally, an L2 regularization penalty with a regularization parameter of 0.1 is applied to the dense layers' kernels to reduce overfitting.

For training, Adam was used as an optimizer with a learning rate of 0.001 which was found as optimal using grid search. The loss function used was binary cross-entropy (BCE) which is commonly used in binary classification problems and is given by [39]:

$$BCE = -\sum_{i=1}^{C'=2} t_i log(s_i) = -t_1 log(s_1) - (1-t_1) log(1-s_1) \tag{1}$$

Early stopping and reduction of learning rate upon plateau were integrated with a patience of 10 and 7 respectively to prevent overfitting alongside a 10% validation split. Training and evaluation were executed using an early 2015 MacBook Pro with an Intel Core i5-5257U CPU, an Intel Iris Graphics 6100 GPU, and a 8GB DDR3L, 1866 MHz RAM. Training took 5626 seconds and 108 epochs, 52 seconds per epoch on average.

Since accuracy can be a misleading metric for imbalanced datasets, the models was evaluated on the testing set using all of the following metrics [40]:

1) Accuracy
2) F1 Score
3) Area under the receiver operating characteristic curve (AUROC)
4) Specificity
5) Sensitivity

The results for the final model are: 89.8% for accuracy, 89.0% for F1 score, and 90.7% for AUROC. 10-fold cross-validation accuracy was 88.01 (±0.75)%. Sensitivity was 93.2% and specificity was 88.1%. The prediction or inference time was 4.21 seconds. The confusion matrix is shown in Fig. 2 .

Additionally, there is an abundance of literature that assert the role of age and sex in the development of cardiovascular diseases and AMI [41]–[43]. And so, to investigate the effect of the exclusion of age and sex on this classification problem and framework, the same network was trained and evaluated on the same testing and training data but without the age

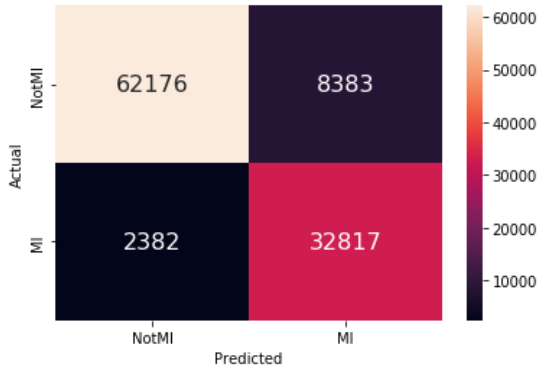


FIGURE 2. CNN model confusion matrix.

and sex features. The performance metrics went down by an average of 3.8% and reduced the sensitivity of the network by 7.9% compared to the data with age and sex. The accuracy obtained in this case was 86.9%, the F1 Score was 85.6%, and the AUROC was 86.5% with a sensitivity of 85.3% and specificity of 87.7%. The inference time taken for the model to predict the labels of the testing set was 3.60 seconds.

The effect of worsened prediction results when not including age and sex demonstrates the importance of auxiliary information for the prediction of AMI events.

B. RNN MODEL

As for the RNN, the overall architecture can be seen in Fig. 1 (c). Since RNN models are mostly applied to time-series data in sequences, here we set timesteps as 1 due it being static ECG data. We might expect this algorithm to perform better than CNN models due to its feedback nature, but since this is the case of static data, it is very likely to have similar performance. RNN models have also shown that they are subpar to CNN models when it comes to feature compatibility and power, thus this could affect their performance on this dataset. RNN models have had limited applications to ECG classification with long-sequence analysis of the MIT-BIH Arrhythmia database [44]. Due to the previously mentioned prevalence of RNN models in deep neural network analysis, we utilize their application in this case for comparative study and holistic approach to AMI prediction.

The model was built in Keras, with the architecture consisting of 13 layers, 6 gated recurrent unit (GRU) layers, 6 normalization layers after every GRU, and a final dense layer for binary representation. Layer normalization was used to reduce the training time and stabilize hidden neuron dynamics in the network [45]. Leaky ReLU has shown better performance than ReLU, and here used as the activation function to avoid gradient saturation for negative values [46]. GRUs have shown better performance results when applied to speech recognition at varying depths of the network [47]. Similar to the CNN model, the loss function was binary cross-entropy and the optimizer was Adam with learning rate of 0.001 found with grid search. Early stopping and reduction of learning rate upon plateau were integrated with a patience of 10 and 7 respectively to prevent overfitting alongside a 10% validation split. Training and evaluation was executed using an HP Spectre x360 with an Intel Core i7-7500U CPU, NVIDIA GeForce 940MX, and a 16GB DDR4 RAM. Training took 4402.12 seconds and 46 epochs, 95.69 seconds per epoch on average.

The results for the final model are: 84.6% for accuracy, 82.8% for F1 score, and 82.9% for AUROC. 10-fold cross-validation accuracy was 83.86 (±0.27)%. Sensitivity was 78% and specificity was 87.8%. The prediction or inference time was 6.96 seconds. The confusion matrix is shown in Fig. 3.

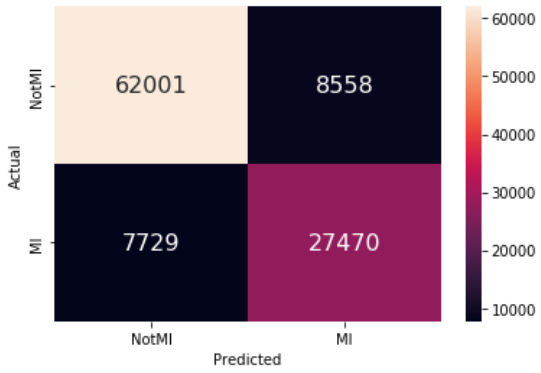


FIGURE 3. RNN model confusion matrix.

Training the RNN model on the dataset without the features of age and sex took 68 epochs and 6593.48 seconds, 96.96 seconds per epoch. The RNN model without the auxiliary inputs took longer to train probably due to taking more time to converge. The accuracy obtained in this case was 80.1%, F1 score was 77%, and the AUROC was 76.3% with a sensitivity of 64.9% and specificity of 87.7%. Once again, clearly showing the importance of auxiliary inputs, age and sex, in the prediction of AMIs.

VI. XGBoost AND SHAPLEY ANALYSIS

Another tree-based extreme gradient boosting (XGBoost) model has also been developed for the prediction of the onset of AMI. Data was processed and split akin to what has been done for the two deep learning models. For training, GBTree was used as the booster with a learning rate of 1 and number of estimator of 100, which was found as optimal using grid search. The results for the final model are: 97.5 (±0.17)% for 10-fold cross-validation accuracy, 97.1% for F1 score, and 96.5% for AUROC. Sensitivity was 93.5% and specificity was 99.4%. The prediction or inference time was 0.398 seconds. The confusion matrix is shown in Fig. 4.

Training the XGBoost on the dataset without the features of age and sex was also conducted. The model achieved high accuracy of 97.4%, F1 score of 97.5%, AUROC of 96.8% with a sensitivity of 93.9% and specificity of 99.7%, similar to the results obtained when conducting training with all

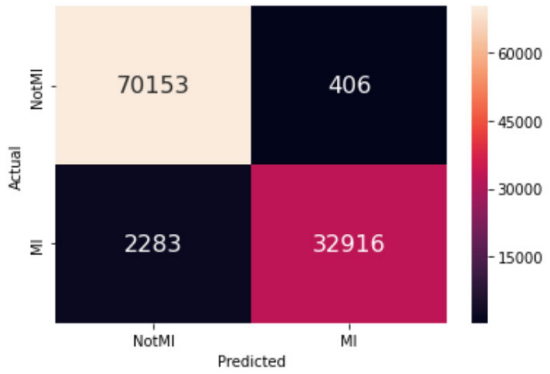


FIGURE 4. XGBoost model confusion matrix.

features. This is because XGBoost performs feature selection as a part of execution of the modeling algorithm and therefore will automatically disregard features it recognizes as less important [48]. Though feature selection is embedded in XGBoost, conducting it prior to training is still more desirable, as the dimension of the search space and therefore the complexity of the task can be reduced.

To gain deeper insight into the model prediction and identify features that contribute most to the prediction, Shapley value, a notion from game theory, is applied to the testing dataset [33], [34]. To calculate the Shapley value of a specific feature i, sets of all possible unions are formed with all n features except feature i. The value of the i-th feature is obtained via calculating the difference between the results of the characteristic function v on N (the set of all features) and S (the subset of N without feature i). Shapley value of a particular feature i is then calculated by taking the average of the marginal contributions of all possible combinations of the feature unions. The following equation is used to calculate the Shapley value φ for feature i:

$$\varphi_i(v) = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \bigcup \{x_i\}) - v(S)) \quad (2)$$

Shapley values are useful in revealing the contribution of each feature to an individual prediction. Fig. 5 includes two sample cases, each predicting AMI negative and positive.

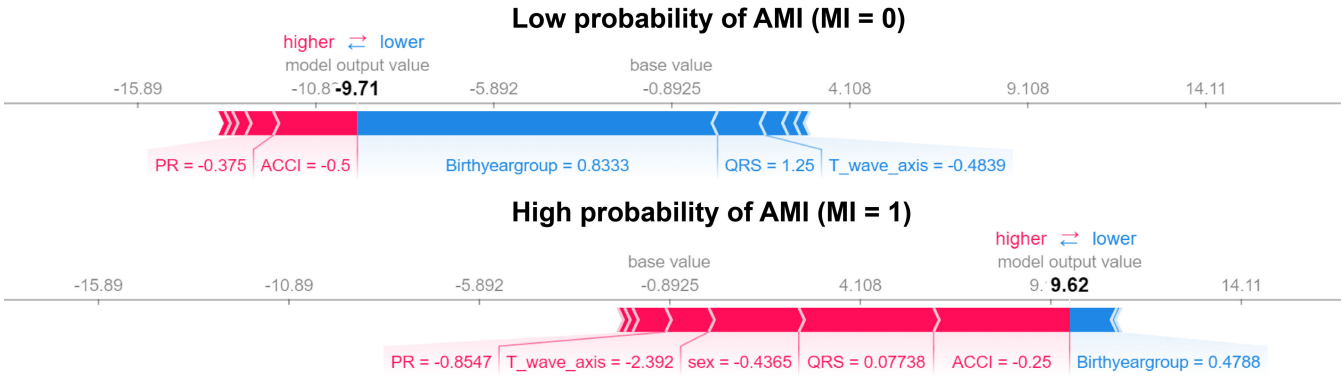
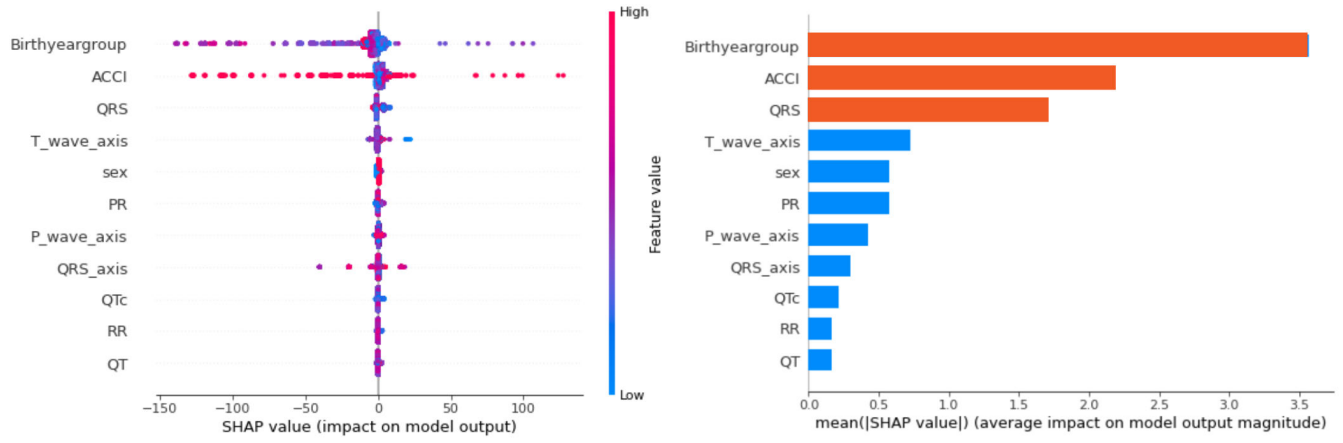


FIGURE 5. Local explanation of two sample cases.



(a) (b)

FIGURE 6. (a) Local explanation summary (b) Global feature importance.

The local explanation graphs show how each feature shift the prediction from the base value (the average model output of the dataset) to the model output. Features that contributed to higher probability of AMI onset are encoded dark pink, while blue is for the contrary.

The beeswarm plot in Fig. 6 (a) gives an overview of the impact of features on the prediction, with each dot representing the Shapley value of every feature for all samples. Fig. 6 (b) shows the average absolute of the Shapley values over the whole testing dataset. Age (Birthyeargroup), ACCI, and QRS duration were observed to be the most important features for the prediction. This confirms previous discovery that adding the age variable helps improve model accuracy, while the feature sex is found to be of less significance.

VII. DISCUSSION AND CONCLUSION

The majority of proposed literature frameworks for this type of prediction task use the MIT-BIH Arrhythmia database or PTB Diagnostic ECG database. However, to the authors' knowledge, this study is the first to utilize the ECG-ViEW II database to propose three machine learning models to predict AMI risk condition. As seen in Table 2, the three proposed models show promising results when evaluated across all 5 performance metrics. The RNN model underperformed when compared to the CNN model likely due to its more fitting application to time-series data and not static data. The CNN model shows competitive F1 score of $> 89\%$, sensitivity $> 88\%$, and specificity $> 93\%$ beating many state-of-the-art literature approaches. The best model was shown to be the XGBoost model with an F1 score of 97.1%, sensitivity of 93.5%, and specificity of 99.4%. Due to the tabular nature of the dataset, it is as expected that the RNN and CNN models did not perform as well as the XGBoost model.

TABLE 2. Model performance summary (No A/S refers to the training without features age and sex).

Measure	CNN		RNN		XGBoost	
	All	No A/S	All	No A/S	All	No A/S
Accuracy	89.9%	86.9%	84.6%	80.1%	97.5%	97.4%
F1 Score	89.0%	85.6%	82.2%	77.0%	97.1%	97.5%
AUROC	90.7%	86.5%	82.9%	76.3%	96.5%	96.8%
Sensitivity	88.1%	85.3%	78.0%	64.9%	93.5%	93.9%
Specificity	93.2%	87.7%	87.8%	87.7%	99.4%	99.7%

Additionally, this paper presents a deeper analysis of this dataset, the proposed models, and the AMI prediction task by examining the contribution of different features, most notably age and sex, on the prediction task through interpretable machine learning. Testing the CNN and RNN models without the age and sex features reduced performance by an average of 3.78% and 5.9% across the 5 metrics respectively. This exhibits the significant role these auxiliary inputs play in determining the AMI risk condition. Also, Shapley value analysis shows that age, ACCI, and QRS duration are the most crucial variables in the prediction of the onset of AMI, while sex is of relatively less importance. The Shapley analysis is shown to be a promising technique to uncover the intricacies and mechanisms of the prediction model, leading to higher degree of interpretation and transparency. Future collaborative work could uncover medical relatedness of these factors to the occurrence of AMIs through clinical testing and consulting.

The proposed framework of deep learning and application of XGBoost to the prediction of AMI from 12-lead ECG data presents a novel approach to this medical condition that can hopefully be integrated in systems for diagnosis and soft real-time monitoring of CVDs.

REFERENCES

[1] K. Mc Namara, H. Alzubaidi, and J. K. Jackson, "Cardiovascular disease as a leading cause of death: How are pharmacists getting involved?" *Integr. Pharmacy Res. Pract.*, vol. Volume 8, pp. 1–11, Feb. 2019.

[2] *Global action plan for the prevention and control of noncommunicable diseases 2013-2020*. World Health Organization, Geneva, Switzerland, 2013.

[3] L. J. Laslett, P. Alagona, B. A. Clark, J. P. Drozda, F. Saldivar, S. R. Wilson, C. Poe, and M. Hart, "The worldwide environment of cardiovascular disease: Prevalence, diagnosis, therapy, and policy issues," *J. Amer. College Cardiol.*, vol. 60, no. 25, pp. 1–49, Dec. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0735109712053715

[4] E. J. Benjamin *et al.*, "Heart disease and stroke statistics-2019 update a report from the american heart association," *Circulation*, vol. 139, no. 10, pp. 56–528, 2019.

[5] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?" *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0174944.

[6] S. J. Al'Aref, "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging," *Eur. Heart J.*, vol. 40, no. 24, pp. 1975–1986, Jun. 2019.

[7] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK biobank participants," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0213653.

[8] J. L. Anderson and D. A. Morrow, "Acute myocardial infarction," *New England J. Med.*, vol. 376, no. 21, pp. 2053–2064, 2017.

[9] J. Schläpfer and H. Wellens, "Computer-interpreted electrocardiograms: Benefits and limitations," *J. Amer. College Cardiol.*, vol. 70, no. 9, pp. 1183–1192, 2017.

[10] A. H. Ribeiro, M. Horta Ribeiro, G. M. M. Paixáo, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. S. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira Jr., T. B. Schön, and A. Luiz P. Ribeiro, "Automatic diagnosis of the 12-lead ECG using a deep neural network," 2019, *arXiv:1904.01949*. [Online]. Available: http://arxiv.org/abs/1904.01949

[11] C. A. Ledezma, X. Zhou, B. Rodríguez, P. J. Tan, and V. Díaz-Zuccarini, "A modeling and machine learning approach to ECG feature engineering for the detection of ischemia using pseudo-ECG," *PLoS ONE*, vol. 14, no. 8, Aug. 2019, Art. no. e0220294.

[12] R. Xiao, Y. Xu, M. M. Pelter, D. W. Mortara, and X. Hu, "A deep learning approach to examine ischemic st changes in ambulatory ecg recordings," *AMIA Summits Transl. Sci. Proc.*, vol. 2018, p. 256, Dec. 2018.

[13] D. Sadhukhan, S. Pal, and M. Mitra, "Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 10, pp. 2303–2313, Oct. 2018.

[14] Y.-G. Kim, D. Shin, M. Y. Park, S. Lee, M. S. Jeon, D. Yoon, and R. W. Park, "ECG-ViEW II, a freely accessible electrocardiogram database," *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0176222.

[15] C. D. Cantwell, Y. Mohamied, K. N. Tzortzis, S. Garasto, C. Houston, R. A. Chowdhury, F. S. Ng, A. A. Bharath, and N. S. Peters, "Rethinking multiscale cardiac electrophysiology with machine learning and predictive modelling," *Comput. Biol. Med.*, vol. 104, pp. 339–351, Jan. 2019.

[16] M. Alfaras, M. C. Soriano, and S. Ortín, "A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection," *Frontiers Phys.*, vol. 7, p. 103, Jul. 2019.

[17] M. Simjanoska, M. Gjoreski, M. Gams, and A. M. Bogdanova, "Non-invasive blood pressure estimation from ECG using machine learning techniques," *Sensors*, vol. 18, no. 4, p. 1160, Apr. 2018.
[18] S. Celin and K. Vasanth, "ECG signal classification using various machine learning techniques," *J. Med. Syst.*, vol. 42, no. 12, p. 241, Dec. 2018.
[19] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," 2020, *arXiv:2004.13701*. [Online]. Available: http://arxiv.org/abs/2004.13701
[20] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-signaldatenbank CARDIODAT der PTB Über das Internet," in *Proc. Biomedizinische Technik/Biomed. Eng.*, Jul. 2009, pp. 317–318.
[21] H. Makimoto, M. Höckmann, T. Lin, D. Glöckner, S. Gerguri, L. Clasen, J. Schmidt, A. Assadi-Schmidt, A. Bejinariu, P. Müller, S. Angendohr, M. Babady, C. Brinkmeyer, A. Makimoto, and M. Kelm, "Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 8445. [Online]. Available: http://search.ebscohost.com/logi%n.aspx?direct=true&db=edb&AN=143359225&site=eds-live
[22] S. M. A. Elrahman and A. Abraham, "A review of class imbalance problem," *J. Netw. Innov. Comput.*, vol. 1, no. 8, pp. 332–340, 2013.
[23] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinf.*, vol. 14, no. 1, p. 106, Dec. 2013.
[24] S. Syama, G. S. Sweta, P. I. K. Kavyasree, and K. J. M. Reddy, "Classification of ECG signal using machine learning techniques," in *Proc. 2nd Int. Conf. Power Embedded Drive Control (ICPEDC)*, Aug. 2019, pp. 122–128.
[25] T. Teijeiro, C. A. Garcia, D. Castro, and P. Fálix, "Arrhythmia classification from the abductive interpretation of short single-lead ECG records," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2017, pp. 1–4.
[26] M. M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, pp. 340–354, Jun. 2016.
[27] Q. Zhao and L. Zhang, "ECG feature extraction and classification using wavelet transform and support vector machines," in *Proc. Int. Conf. Neural Netw. Brain*, 2008, pp. 1089–1092.
[28] K. K. Patro and P. R. Kumar, "Effective feature extraction of ecg for biometric application," *Procedia Comput. Sci.*, vol. 115, pp. 296–306, Dec. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187705091731966X
[29] R. Rodríguez, A. Mexicano, J. Bila, S. Cervantes, and R. Ponce, "Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis," *J. Appl. Res. Technol.*, vol. 13, no. 2, pp. 261–269, Apr. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1665642315000103
[30] D. Nguyen, N. A. Smith, and C. P. Rosé, "Author age prediction from text using linear regression," in *Proc. 5th ACL-HLT workshop Lang. Technol. Cultural Heritage, Social Sci.*, 2011, pp. 115–123.
[31] D. Lavanya and D. K. U. Rani, "Analysis of feature selection with classification: Breast cancer datasets," *Indian J. Comput. Sci. Eng.*, vol. 2, no. 5, pp. 756–763, Oct. 2011.
[32] M. M.-C. Vidovic, N. Görnitz, K.-R. Müller, and M. Kloft, "Feature importance measure for non-linear learning algorithms," 2016, *arXiv:1611.07567*. [Online]. Available: http://arxiv.org/abs/1611.07567
[33] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4765–4774.
[34] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 2522–5839, 2020.
[35] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Inf. Sci.*, vols. 415–416, pp. 190–198, Nov. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025517308009
[36] N. Strodthoff and C. Strodthoff, "Detecting and interpreting myocardial infarction using fully convolutional neural networks," *Physiol. Meas.*, vol. 40, no. 1, p. 015001, Jan. 2019, doi: 10.1088%2F1361-6579%2Faaf34d.
[37] T. Reasat and C. Shahnaz, "Detection of inferior myocardial infarction using shallow convolutional neural networks," in *Proc. IEEE Region 10 Humanitarian Technol. Conf.*, Dec. 2017, pp. 718–721.
[38] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2018, pp. 443–444.
[39] S. Mannor, D. Peleg, and R. Rubinstein, "The cross entropy method for classification," in *Proc. 22nd Int. Conf. Mach. Learn.*, New York, NY, USA, 2005, pp. 561–568, doi: 10.1145/1102351.1102422.
[40] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, p. 27, Dec. 2019.
[41] J. G. Canto, W. J. Rogers, R. J. Goldberg, E. D. Peterson, N. K. Wenger, V. Vaccarino, C. I. Kiefe, P. D. Frederick, G. Sopko, and Z.-J. Zheng, "Association of age and sex with myocardial infarction symptom presentation and in-hospital mortality," *JAMA*, vol. 307, no. 8, pp. 813–822, Feb. 2012.
[42] V. Culić, D. Eterović, D. Mirić, and N. Silić, "Symptom presentation of acute myocardial infarction: Influence of sex, age, and risk factors," *Amer. Heart J.*, vol. 144, no. 6, pp. 1012–1017, Dec. 2002.
[43] J.-Y. Shih, Z.-C. Chen, H.-Y. Chang, Y.-W. Liu, C.-H. Ho, and W.-T. Chang, "Risks of age and sex on clinical outcomes post myocardial infarction," *IJC Heart Vasculature*, vol. 23, Jun. 2019, Art. no. 100350.
[44] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, and H. Yang, "Patient-specific ECG classification based on recurrent neural networks and clustering technique," in *Proc. Biomed. Eng.*, 2017, pp. 63–67.
[45] J. Lei Ba, J. Ryan Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*. [Online]. Available: http://arxiv.org/abs/1607.06450
[46] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*. [Online]. Available: http://arxiv.org/abs/1505.00853
[47] S. Khandelwal, B. Lecouteux, and L. Besacier. *Comparing Gruand LSTM for Automatic Speech Recognition*. HAL. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01633254
[48] R. J. Urbanowicz, M. Meeker, Wsss. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Informat.*, vol. 85, pp. 189–203, Sep. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1532046418301400

LUJAIN IBRAHIM (Member, IEEE) received the B.S. degree (*cum laude*) in computer engineering from New York University Abu Dhabi (NYUAD), United Arab Emirates, in 2020. She was a part of the NYUAD 2018 team that participated in the International Genetically Engineered Machine (iGEM) Synthetic Biology Competition, and was awarded the Gold Medal, nominated for Best Diagnostic Project and for Finalist. She is interested in explainable machine learning for social good and the societal implications of artificial intelligence. She recently co-founded Multiplicity, a publication examining the Internet's role in young people's lives, and joined the Laboratory for Computer–Human Intelligence at NYUAD as a Research Assistant to conduct research on human decision-making.

MUNIB MESINOVIC (Member, IEEE) received the B.S. degree (*summa cum laude*) in electrical engineering from New York University Abu Dhabi (NYUAD), United Arab Emirates, in 2020. He is currently pursuing the M.Sc. degree in computer science with the University of Oxford. He was awarded the NYU Founders' Day Award. He was a part of the NYUAD 2018 team in the International Genetically Engineered Machine (iGEM) Synthetic Biology Competition, and was awarded the Gold Medal, nominated for Best Diagnostic Project and for Finalist. In 2019, upon reception of the Summer Research Fellowship at the NYU Tandon School of Engineering, he conducted research on adversarialism in autonomous vehicles with Professor Anna Choromanska. He is interested in explainable machine learning for social good and digital ethics, pursuing these topics as a contributing writer for publications. Having won the Rhodes Scholarship in 2019, thereby becoming the first Bosnian to ever win this prestigious award. His research interests include clinical artificial intelligence and digital health.

KAI-WEN YANG (Student Member, IEEE) received the B.S. degree (*summa cum laude*) in electrical engineering from New York University Abu Dhabi (NYUAD), United Arab Emirates, in 2020. She was awarded the NYU Founders' Day Award and was elected to membership in the national honor society of Phi Beta Kappa, Beta Chapter of New York. She was part of the NYUAD 2018 team that participated in the International Genetically Engineered Machine (iGEM) Synthetic Biology Competition, and was awarded the Gold Medal, nominated Best Diagnostic Project and Finalist. She was an Undergraduate Research Assistant with the Integrated BioElectronics Lab, NYUAD, conducting research in the field of implantable neural interface. In 2020, she joined The Clinical AI Research Lab at NYUAD, conducting research on using clinical data for the prediction of patient outcomes. Her research interests include healthcare data science and clinical AI.

MOHAMAD A. EID (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada, in 2010. He is currently an Assistant Professor of electrical and computer engineering with New York University Abu Dhabi (NYUAD). He was previously a Teaching and Research Associate with the University of Ottawa, from June 2008 to April 2012. He has won several awards for academic and research distinction including the Natural Sciences and Engineering Research Council of Canada (NSERC) Award of Excellence, University of Ottawa Excellence scholarship, and Ontario Graduate Scholarship (OGS) scholarship. He is the coauthor of the book *Haptics Technologies: Bringing Touch to Multimedia*, Springers 2011, the technical chair of the Haptic-Audio-Visual Environment and Gaming (HAVE) symposium in several years (2013, 2014, 2015, 2017, and 2019). He was a recipient of several best paper award in several international conferences such as ICBAE 2016, DS-RT 2008, and the prestigious ACM Multimedia 2009 Grand Challenge Most Entertaining Award for "HugMe: Synchronous Haptic Teleconferencing" System. He has more than 100 conference and journal publications and six patents. His academic interests include affective haptics, neurohaptics, and novel haptic interfaces.