Automatic Detection of Valvular Heart Disease on ECG

Tessel Huibregtsen

Amsterdam Center for Computational Cardiology
November 7, 2024

Abstract

Valvular heart disease (VHD) is a serious cardiac disease often diagnosed at advanced stages due to the late onset of symptoms. The current gold standard for diagnosis, echocardiography, is resource intense and less accessible in early-stage detection. Therefore, this project explores the potential of a variational autoencoder (VAE) trained on electrocardiographic (ECG) data instead, to detect VHD. Using a VAE pretrained on the UK Biobank dataset, we generate latent embeddings of ECGs from the Amsterdam UMC. These embeddings demonstrate the VAE's capacity to encode relevant features of ECG data. Subsequently, an XGBoost classifier trained on these embeddings achieved a promising ROC-AUC of 0.75 (95% CI: 0.7433-0.7591). Although further research is required to improve interpretability and distinguish affected valves, these initial results suggest that VAE-based models could offer a new path for early-stage VHD detection using ECGs. This project serves as a basis for future work in automated, early-stage VHD detection using ECG. All code for this project can be found at : `https://github.com/tesselhuib/Valve_AI`.

1 Introduction

Valvular heart disease (VHD) occurs when one or more of the heart valves malfunction, preventing an efficient blood flow through the heart. As life expectancy increases, the prevalence of VHD has also risen [1]. However, many cases are only diagnosed at advanced stages, when significant damage has already occurred [2]. VHD often remains asymptomatic until the disease has caused substantial changes in heart structure and function [3]. Most symptoms are due to these secondary consequences of the disease, which leads to patients only seeking medical attention after these complications arise.

Currently, the gold standard for diagnosing VHD is an echocardiogram [4]. This allows doctors to directly observe valve abnormalities. Echocardiography, however, requires specialized equipment and trained operators, making it less accessible. It is typically performed in hospital settings, often necessitating a referral and extended wait times, which can delay diagnosis and increase healthcare costs.

The electrocardiogram (ECG) however, is widely available, and can be performed easily in a general practitioners setting. ECGs require less time and specialized training, making them both more accessible and more affordable. This accessibility allows ECGs to be performed more frequently and on a larger scale, potentially enabling earlier identification of VHD. Doctors already detect a variety of heart conditions using ECG. However, reliably diagnosing VHD from ECG remains a challenge, due to the fact that known potential ECG changes are not specific to VHD [5]. These changes often arise from structural heart alterations that can occur due to other factors, not just VHD.

Recent advancements in artificial intelligence, particularly in deep learning, have shown great promise in improving diagnostic capabilities using ECGs [6]. Deep neural networks (DNNs) have been developed to detect and predict many heart conditions from ECGs. Despite this progress, there are few, if any, pretrained networks publicly available, that are specifically trained for the detection of VHD.

Therefore, the aim of this project is to develop an open access DNN capable of detecting VHD using a database of patient data from the Amsterdam UMC (AUMC). By using this large dataset in

combination with a DNN, the goal is to develop a model that could help improve the detection of VHD through ECG. This could lead to earlier diagnosis and treatment.

2 Methodology

To be able to detect cardiac diseases on ECG, large datasets and long training times are typically required. In this project, this was addressed by implementing a pretraining strategy. First, a variational autoencoder was trained on a 12-lead ECG database from UK Biobank. This pretraining led to a trained encoder capable of determining the important characteristics of an ECG and representing it in a meaningful matter. Afterwards, this pretrained encoder was used on a Amsterdam UMC database, which consists of ECG measurements with and without a VHD diagnosis. The database was passed through the model and the meaningful representations were saved. Next, a XGBoost classification model was used to differentiate between the representations of VHD diagnoses and non-VHD diagnoses. Lastly, the performance of the VAE as well as the XGBoost model were assessed.

2.1 Database

In this project, data from the UK Biobank (UKB) and from the Amsterdam University Medical Center (AUMC) were used. The UKB consists of approximately 80,000 12-lead ECGs from a sample of the general population including a limited number of people with a VHD diagnosis. The AUMC data obtained was first preprocessed, as described in the section below, which resulted in a database of 9,423 people with an echocardiography confirmed VHD, as well as 40,319 age and sex matched people without VHD (also confirmed by echocardiography). From these 49,742 AUMC participants we had information from 136,825 12-lead ECGs, taken from a 12 month window before and after the index diagnosis (either VHD or a non-VHD diagnosis, see Table 1), excluding the ECGs after potential valvular replacement surgery. Data from the UKB were used to train a variational autoencoder (VAE) model. The encoder generated embeddings were then used to derive a VHD classification model on the AUMC data.

Table 1: Top 5 Index Diagnoses of the AUMC ECG Database

Valvular Heart Disease			**Non-Valvular Heart Disease**		
ICD-10	**Description**	**N**	**ICD-10**	**Description**	**N**
I35.0	Aortic valve stenosis	4,264 (45.3%)	I10	Essential (primary) hypertension	11,344 (28.1%)
I34.0	Mitral valve insufficiency	2,093 (22.2%)	I25.1	Atherosclerotic heart disease	5,730 (14.2%)
I35.1	Aortic valve insufficiency	790 (8.3%)	E11.9	Non-insulin-dependent diabetes mellitus without complication	4,242 (10.5%)
I36.1	Nonrheumatic tricuspid valve insufficiency	483 (5.1%)	I48.9	Atrial fibrillation and atrial flutter, unspecified	3,978 (9.9%)
I07.1	Tricuspid insufficiency	254 (2.7%)	I25.2	Old myocardial infarction	3,187 (7.9%)

2.2 Data preprocessing

To prepare the ECG data for analysis, several preprocessing steps were applied to ensure data quality and compatibility across formats. ECGs suspected of significant motion artefacts (where over 10% of samples had an absolute value larger than 1.5 millivolts) and failed ECG recordings (missing median waveforms) were excluded in this project. For the AUMC database, this led to an exclusion of 16,066 ECGs. Next to that, ECGs of people diagnosed with VHD obtained after a valvular replacement surgery were also excluded as the status of their valvular disease was no longer clear. This resulted in the exclusion of another 39,079 ECGs.

The UK Biobank data is saved in XML format and the AUMC data in DICOM format. To be able to process both formats easily, while minimizing disk memory usage, the signals were first plotted using Matplotlib. An example plot can be found in Figure 1. They were then converted to compressed NumPy arrays, with the images saved as pixel arrays, and saved to disk. For pretraining the VAE, a random sample of 30,000 preprocessed ECGs was selected of the UK Biobank data to balance dataset size and computational efficiency.

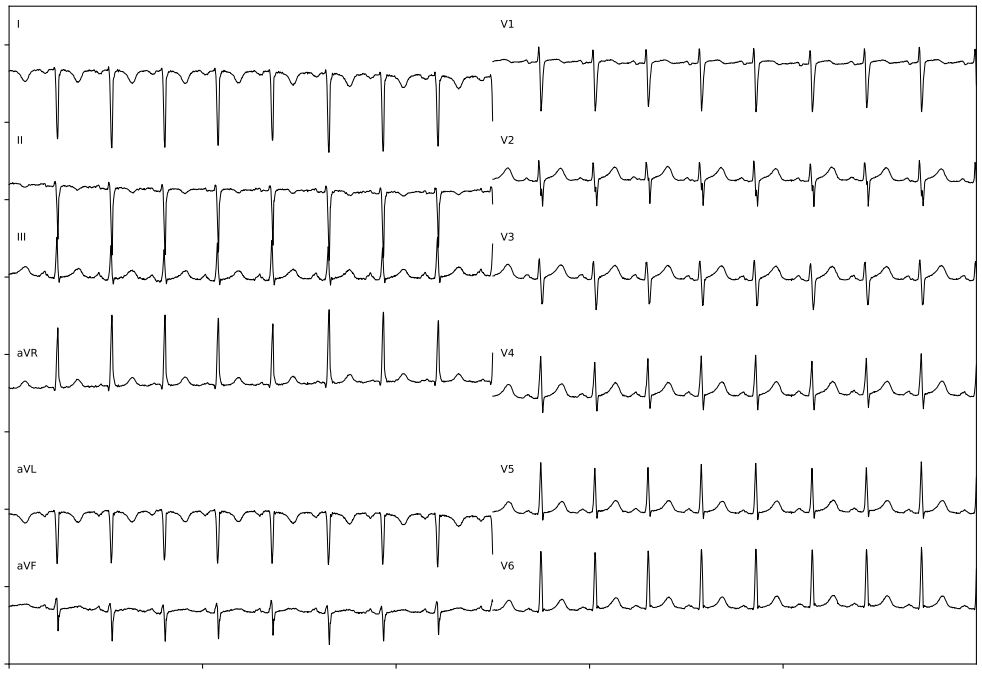

Figure 1: An example image of a 12-lead ECG signal. The leads are indicated by their name. For each subplot, time is on the x-axis and voltage on the y-axis.

2.3 Variational Autoencoder

A variational autoencoder (VAE) was used for pretraining. A VAE is a neural network that learns to create compressed representations, or latent embeddings, of input data [7]. It is comparable to a principal component analysis, but instead of finding linear and mutually indepedent combinations of the original variables it uses non-linear relations to capture more intricate patterns of the input data. The found embeddings capture the important image features, allowing the VAE to reconstruct the original image. Compared to a regular autoencoder, a VAE includes a loss function that forces the latent distribution to follow the Gaussian distribution. In this way, the model creates a smooth latent distribution that generalizes well. In this project, the VAE was trained to represent ECGs in a meaningful manner.

To train the VAE, the total loss, $\mathcal{L}_{\text{total}}$, combines the Binary Cross Entropy (BCE) loss, $\mathcal{L}_{\text{BCE}}$, and the Kullback-Leibler Divergence (KLD) loss, $\mathcal{L}_{\text{KLD}}$.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{KLD}} \tag{1}$$

The BCE loss,

$$\mathcal{L}_{\text{BCE}} = -\sum_{i} \left(x_i \log(\hat{x}_i) + (1 - x_i) \log(1 - \hat{x}_i)\right) \tag{2}$$

is responsible for the reconstruction accuracy, ensuring that the output of the decoder, $\hat{x}$, closely matches the original input, x. This encourages the model to accurately reconstruct the data it receives, thereby improving the quality of the generated outputs. The logarithm in the function punishes wrong predictions stronger than for example a mean squared error loss, making it the preferred choice.

The KLD loss,

$$\mathcal{L}_{\text{KLD}} = -\frac{1}{2}\sum_{j=1}^{d}\left(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2\right) \tag{3}$$

on the other hand, regularizes the latent space by promoting similarity between the approximate posterior distribution (produced by the encoder, with mean μ and variance σ^2) of the latent space with dimension d, and a predefined prior distribution, in our case a standard unit Gaussian. This regularization helps create a smooth and continuous latent space. The two loss functions are added to create the total loss on which the VAE is trained.

2.3.1 Network Architecture

The network architecture of the VAE was designed to accurately process the information of the ECG, while not requiring too many resources. The input consists of compressed NumPy arrays, which are transformed to a size of (272, 400), normalized, and converted to grayscale. The chosen input size maximizes the information provided to the network while optimizing GPU memory usage. The image is converted to grayscale to simplify the image, as it only contains black (representing the signal) and white (representing the background) pixels, making additional color information unnecessary.

As depicted in Figure 2, the encoder and decoder consist of four (de)convolutional layers. Inbetween is the latent space, with two linear layers that output two vectors representing the mean and the log-variance of the latent space distribution. The decoder begins with a reparameterization step, where a sample is drawn from the mean and log-variance.

All convolutional and deconvolutional layers have a kernel size of 4 with a stride of 2 and padding of 1. Both the convolutional and first three deconvolutional layers use Leaky ReLU activation, while the final deconvolutional layer uses a sigmoid activation function to produce an output normalized between 0 and 1. The UKB data was split into a training, validation and test set with split ratios of 0.8, 0.1 and 0.1 respectively. This is done to ensure the model is effectively trained, tuned, and evaluated in a controlled and unbiased manner. Training was done on the training set with a batch size of 500 and a learning rate of 0.0001. An early stopping strategy with a patience of 10 epochs was implemented on the validation set and the model was allowed to run for a maximum of 300 epochs.

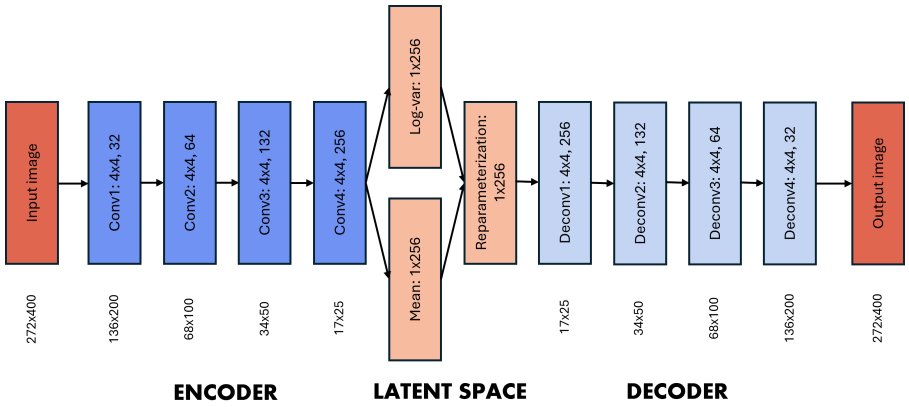

Figure 2: The architecture of the VAE. The dimensions below the blocks indicate the individual image dimensions, while the dimensions within the blocks indicate the kernel dimensions and the feature dimension respectively for the convolutional layers. For the linear layers, the dimensions within the blocks represent the vector dimensions.

2.4 Classifier

The trained encoder was used to generate latent embeddings of the AUMC database. This database was also split into training, validation, and test sets with ratios of 0.8, 0.1, and 0.1, respectively, and passed through the encoder to generate embeddings. To identify the embeddings with the most useful

information, two analyses were conducted. First, a Spearman correlation matrix was generated to identify any strongly correlated embeddings, which would likely provide redundant information. For each strongly correlated pair (correlation coefficient >0.8), one was excluded. Second, the variance of each embedding was assessed to exclude embeddings with minimal variation (<0.05), as these likely contained little useful information. These steps led to the exclusion of 11 embeddings. The other 245 embeddings were then loaded into an XGBoost classification model.

XGBoost is an ensemble learning method based on gradient boosting, where multiple decision trees are sequentially created, with each new tree trying to correct the errors of the previous ones [8]. This iterative process minimizes classification errors by adjusting weights for each instance based on previous predictions. The final output is obtained by applying a sigmoid to the sum of the weighed outputs of all trees.

For the XGBoost classification model, hyperparameter tuning was performed using Optuna, an open-source hyperparameter optimization framework designed to automate the search for optimal hyperparameter values [9]. The hyperparameters were optimized based on the Receiver Operating Characteristic (ROC) - Area Under the Curve (AUC) value, as this metric is less sensitive to class imbalance and provides a more comprehensive evaluation of the model's performance across different classification thresholds. The following hyperparameters were tuned: the max depth of the decision trees, the learning rate, the number of boosting rounds done and the subsampling ratio of the training samples. The hyperparameter ranges specified for Optuna are recorded in Table 2. Optuna randomly samples from within these ranges with a uniform distribution. 50 Optuna trials were run. To evaluate the performance of XGBoost with the tuned hyperparameters, a baseline LASSO logistic regression model was also implemented for comparison.

Table 2: Hyperparameter ranges for XGBoost hyperparameter tuning with Optuna

Hyperparameter	**Value Type**	**Range**
Max depth	int	[3, 11]
Learning rate	float	[0.001, 0.15]
Number of boosting rounds	int	[250, 450]
Subsample ratio	float	[0.5, 1.0]

3 Results

The preprocessed AUMC data consisted of 49,742 patients of which 9,423 had a echocardiography confirmed VHD diagnosis. All database characteristics can be found in Table 3.

Table 3: Characteristics of the AUMC ECG Database

	Valvular Heart Disease	**Non-Valvular Heart Disease**
Number of patients	9,423	40,319
Number of ECGs	64,846	71,979
Sex		
Male (N)	5,199 (55.2%)	22,189 (55.0%)
Female (N)	4,224 (44.8%)	18,130 (45.0%)
Age	69.4 (15.6)	65.9 (15.7)
ECG parameters		
PR Interval (ms)	175.238 (45.24)	168.570 (35.85)
QRS duration (ms)	112.299 (31.00)	99.28 (21.63)
QRS duration > 120ms (N)	19,243 (29.7%)	9,629 (13.4%)
QTc Interval (ms)	463.111 (48.7)	441.875 (37.43)

Note: values are reported as mean (standard deviation) unless otherwise specified.

3.1 VAE

After pretraining the VAE with the architecture and hyperparameters mentioned in section 2.3, the performance of the VAE was assessed. This was done by calculating the average mean squared error (MSE) and the average binary cross entropy (BCE) on a withheld test set of 3,000 samples of the UKB dataset. This resulted in an average MSE of 218 per sample, and an average BCE of 5848 per sample. Also the MSE and BCE on a subset of 13,685 samples of the AUMC dataset was calculated. This resulted in an average MSE of 209 per sample, and an average BCE of 5919 per sample. Figure 3 shows two examples of the reconstructed outputs compared to their inputs.

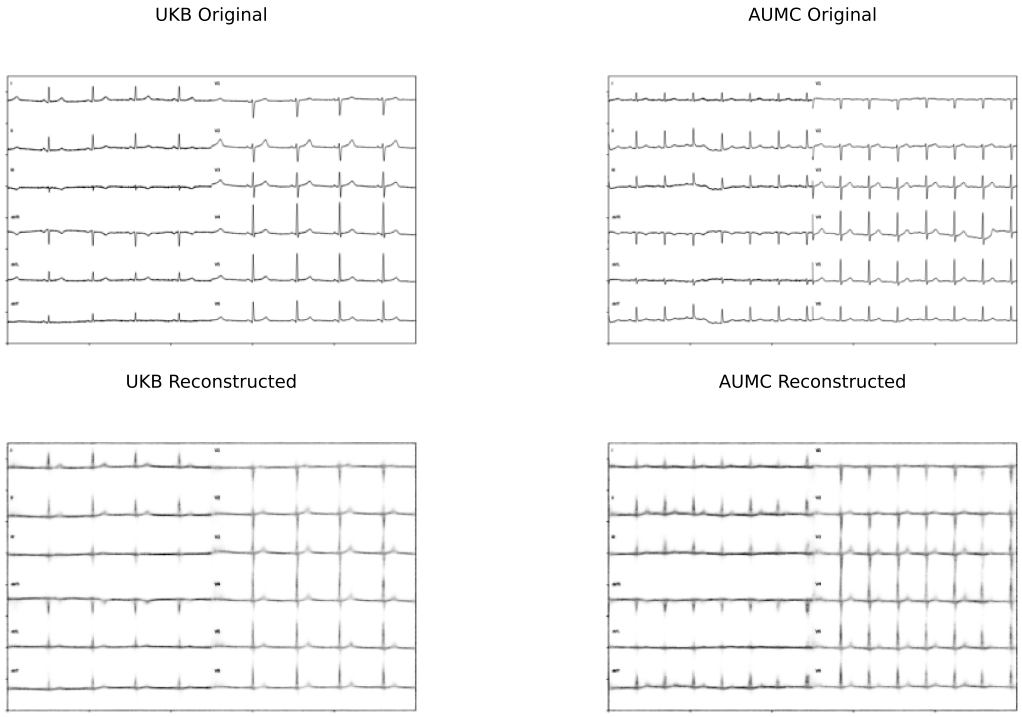

Figure 3: Two sample original images (on top) vs their reconstruction (on the bottom) by the trained VAE of the UK Biobank dataset (left) and the AUMC dataset (right). The resolution of the images is low due to their compressed size (272x400) needed to train the VAE efficiently. Each of the 12 leads in each image has time on the x-axis and voltage on the y-axis.

3.2 XGBoost Classification

To research the information density of the latent embeddings generated by the trained encoder on the AUMC database, a Spearman correlation matrix was created comparing these embeddings to ECG parameters associated to the ECG (see Figure 4). Only the correlations with an adjusted p-value (using False Discovery Rate correction for multiple comparisons) below 0.05 are shown. Five ECG parameters, QT interval, RR interval, QTc interval, QRS duration and PR interval, were chosen to compare with as these parameters were already available for most of the AUMC ECGs. These ECG parameters indicate the duration between different parts of an ECG signal and therefore give key information about the shape of an ECG.

Next, the XGBoost Classifier was trained using the train and validation set of the AUMC database, using Optuna for hyperparameter optimization. This resulted in the following hyperparameter values: a max depth of 10, a learning rate of 0.03, 424 boosting rounds and a subsampling ratio of 0.91.

To assess the performance of the XGBoost classification model, predictions were made on the AUMC database test set, which was withheld from any training. The ROC curve in Figure 5 shows its performance, with a 95% confidence interval for the ROC-AUC score between 0.7433 and 0.7591. For comparison, we also ran a LASSO logistic regression, which yielded a ROC-AUC value of 0.6698.

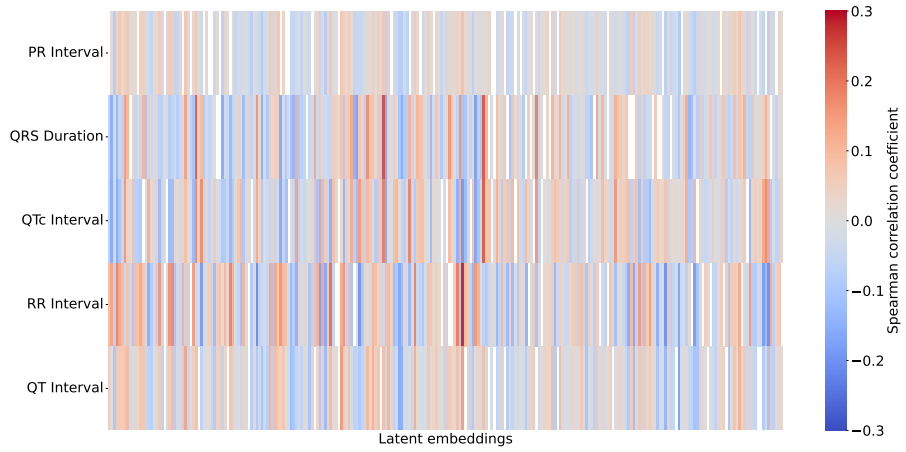

Figure 4: Significant Spearman correlation matrix. The correlation between five ECG parameters, on the y-axis, with the 256 latent embeddings, on the x-axis, is shown. Correlations with an adjusted p-value above 0.05 are removed and therefore white.

This gives us a benchmark to see how the XGBoost model performs compared to more traditional methods. The precision-recall curve is also included in Figure 5.

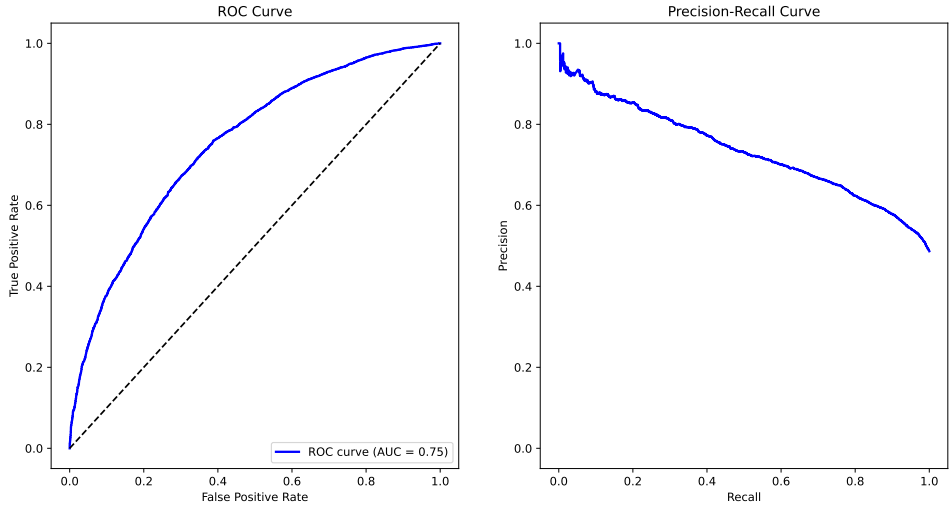

Figure 5: The ROC curve and the precision-recall curve for the XGBoost classification model predictions on the AUMC database test set. The dashed line in the ROC curve indicates the baseline performance of a random classifier.

Table 4 shows the performance metrics of the XGBoost classification model. Different thresholds were tested to classify the outcomes of the model. Predicted probabilities above the threshold were classified as VHD, predicted probabilities below the threshold were classified as non-VHD. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated for each threshold. In Figure 6, a density plot of the predicted probabilities that the XGBoost classification model outputs is shown, separated based on their true label.

4 Discussion

The results of this study show the potential of a VAE in capturing significant ECG features. After pretraining the VAE on the UKB dataset and subsequently evaluating it on both UKB and AUMC datasets, the relatively close MSE and BCE scores across the two datasets indicate that the model generalizes well. However, the average MSE scores, 218 for UKB and 209 for AUMC, and BCE scores,

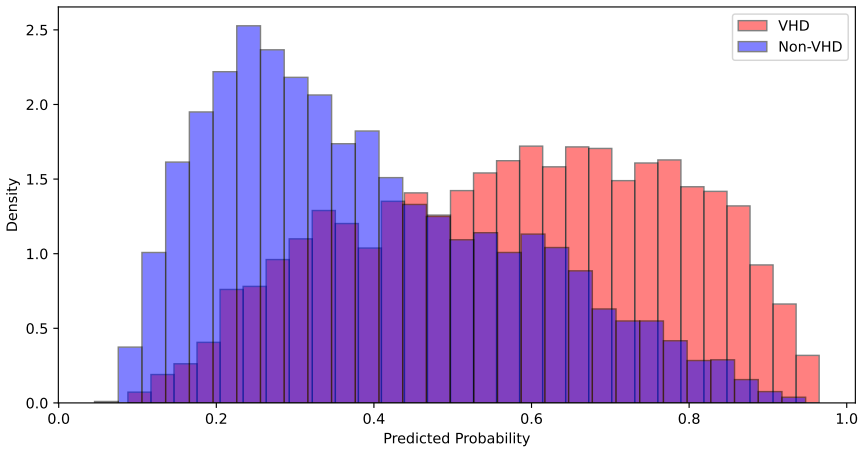

Figure 6: The overlapping density histograms of the predicted VHD probabilities for VHD patients and Non-VHD patients.

Table 4: Performance Metrics of XGBoost by Threshold

Threshold	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**
Sensitivity	0.89	0.77	0.64	0.49	0.32
Specificity	0.39	0.59	0.72	0.83	0.92
PPV	0.58	0.64	0.69	0.73	0.80
NPV	0.79	0.73	0.68	0.63	0.59

5848 for UKB and 5919 for AUMC, are still relatively high, suggesting that more optimization of the model could still be possible. As the current architecture is still relatively shallow, deepening this could be an interesting step. Even so, Figure 4 already shows significant correlation between the latent embeddings and ECG parameters, demonstrating the models capability to encode useful ECG information in its latent space. We do observe that there is little correlation for the PR interval and QT interval compared to the other ECG parameters. For the PR interval, this can be explained by the fact that it relies on the detection of the P wave, which is usually small, and in Figure 3 we can see that the compression of the images leads to loss of detail, including some P waves. This means the VAE is not able to encode these either. The association of latent embeddings with the QTc interval, rather than the QT interval, suggests that the model is capturing features more aligned with heart rate-adjusted cardiac characteristics. Since QTc normalizes for heart rate variability, it provides a more stable measure across samples, which may explain why the latent embeddings show stronger associations with QTc than with the uncorrected QT interval.

In evaluating the XGBoost classifier's performance, the ROC-AUC of 0.75 (95% CI: 0.7433 to 0.7591) shows that the VAE embeddings are useful for downstream classification tasks, specifically for the detection of valvular heart disease (VHD) in our case. Compared to a LASSO logistic regression (ROC-AUC 0.6698), the XGBoost classification model showed better performance, indicating its added value over a basic regression model.

The performance metrics in Table 4 provide further insight into the models performance under different classification thresholds. At higher threshold values, the specificity increases, whereas at lower threshold values, the sensitivity increases, catching more VHD cases but at the cost of increased false positives. This trade-off suggests that threshold adjustments should be made based on the application of the model. For example, in screening contexts, a higher false positive rate may be acceptable if it enables the detection of more potential VHD cases.

While the results are promising, the project also has its limitations. First of all, the control group would ideally have consisted of non-cardiology patients as ECGs of cardiology patients most likely

contain other abnormalities that could influence the model. However, the time constraint of the project did not allow for waiting for data of non-cardiology patients. Also, further research is needed to understand the underlying basis for classification. It could be possible that the model is currently associating with something else than VHD, such as secondary consequences of VHD that are not specific to VHD. Therefore, more research into the features that determine classification, and ensuring they are specific to VHD, is required to ensure accurate and generalizable VHD classification.

5 Clinical implications

This project provides a foundation for further research related to the detection of VHD. The ability to detect VHD on a 12-lead ECG allows for easier and earlier detection of VHD. Since a 12-lead ECG measurement is already standard procedure for patients presenting with suspected cardiac disease, it could easily be used to detect VHD as well. Next steps for eventual clinical implementation would be to include which valve is affected. This information is available in the databases already and could therefore be added into the model relatively easily, to detect which valve is affected based on the ECG. This would give clinicians indispensable information for potential treatment that the current model does not yet give. Table 1 shows there is currently a big imbalance in the representation of type of valve affected. This should be handled carefully when adding this information to the model, to ensure the model doesn't solely focus on one type of valve.

After the type of valve is included, it would be interesting to see if the model is able to classify the severity of the VHD as well. VHD presents with different types of severity. Mild cases might not require clinical intervention, whereas severe cases often require clinical intervention. Being able to distinguish between these types allows for better indication of the need for clinical intervention for a patient. Including VHD severity in the model opens up another research direction: investigating whether the model can identify patients with mild VHD who are likely to progress to severe cases, thereby allowing for earlier intervention. If patients with a high risk of severe VHD are recognized sooner, they can be treated earlier and extra damage to the heart can be prevented. In line with this, a future goal would be to use the model for screening, either in the hospital or even earlier, at the general practitioner (GP). The model could be used to detect VHD earlier and determine who is at risk of severe VHD. Evaluating the current model's performance on ECGs from a GP setting could be valuable as a good performance allows for care to be moved to the GP and out of the hospital. Currently patients suspected of VHD require a echocardiography in the hospital, but if ECGs from the GP could be used effectively instead, patients would no longer need to visit the hospital for diagnosis.

In terms of earlier detection, another path that could be explored is that of detecting VHD on single lead ECGs. Single lead ECGs are currently widespread in smartwatches and could allow for at-home monitoring. Therefore, if the network can be fine-tuned for single lead ECGs from smartwatches it could open up the possibility to detect VHD before patients experience symptoms and enter the clinic. This could also be extended to photoplethysmography (PPG) measurements, as previous research has shown that the performance of automatic cardiac disease detection on PPG is comparable to that of single lead ECG [10]. PPG measurements can be done by smartphones and can therefore be employed in parts of the world where access to healthcare is limited. As smartphones are more widely used than smartwatches, PPG is likely a more accessible measurement than single lead ECG. If VHD can be detected and/or predicted on PPG measurements, it could help identify patients in resource-limited areas and guide them towards appropriate medical care.

Another path that could be explored is that of the impact of valvular interventions. For this project, ECGs obtained after a valvular intervention were excluded as the label of these ECGs became ambiguous. However, it could be interesting to research what label the network gives to these ECGs and whether that is indicative of the result of the intervention. This would allow for further exploration of the reasons why and/or when an intervention could be beneficial. Does the label change to non-VHD for succeeded interventions? If so, why does the model make this decision and if not, why not. Also, can we predict what will happen before the intervention has happened?

Lastly, because the pretrained network has been made publicly available on GitHub, this project provides researchers with a baseline for further research on detecting VHD using ECG. Currently, there are limited to no publicly available pretrained networks using 2D input specifically designed for this purpose, leading to unnecessary duplication of efforts among researchers. Our open-access approach enables researchers to build on existing work, leading to more effective solutions for early detection and intervention of VHD.

6 Conclusion

To conclude, this project showcased the promises of a variational autoencoder to encode an ECG in a meaningful matter to be able to detect valvular heart disease. A relatively shallow VAE of four layers deep, in combination with a relatively simple XGBoost classification model already showed promising results with a ROC-AUC score of 0.75. This project can be seen as a starting point for futher exploration of detecting VHD on ECG. For clinical implementation, the next steps could be to include the type of valve affected, the severity of the VHD and to predict the future course of the disease. With continued research and validation, the use of deep neural networks on ECGs could lead to earlier and more accessible detection of VHD, ultimately enhancing patient outcomes through timely intervention.

References

[1] Coffey, S., Roberts-Thomson, R., Brown, A., Carapetis, J., Chen, M., Enriquez-Sarano, M., Zühlke, L., & Prendergast, B. D. (2021). Global epidemiology of valvular heart disease. Nature Reviews Cardiology, 18(12), 853–864. https://doi.org/10.1038/s41569-021-00570-z

[2] d'Arcy, J. L., Coffey, S., Loudon, M. A., Kennedy, A., Pearson-Stuttard, J., Birks, J., Frangou, E., Farmer, A. J., Mant, D., Wilson, J., Myerson, S. G., & Prendergast, B. D. (2016). Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: The OxVALVE Population Cohort Study. European Heart Journal, 37(47), 3515. https://doi.org/10.1093/eurheartj/ehw229

[3] Institute of Medicine (US) Committee on Social Security Cardiovascular Disability Criteria. Cardiovascular Disability: Updating the Social Security Listings. Washington (DC): National Academies Press (US); 2010. 12, Valvular Heart Disease. Available from: https://www.ncbi.nlm.nih.gov/books/NBK209979/

[4] Vahanian, A., Beyersdorf, F., Praz, F., Milojevic, M., Baldus, S., Bauersachs, J., Capodanno, D., Conradi, L., De Bonis, M., De Paulis, R., Delgado, V., Freemantle, N., Gilard, M., Haugaa, K. H., Jeppsson, A., Jüni, P., Pierard, L., Prendergast, B. D., Sádaba, J. R., ... ESC National Cardiac Societies. (2022). 2021 ESC/EACTS Guidelines for the management of valvular heart disease: Developed by the Task Force for the management of valvular heart disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). European Heart Journal, 43(7), 561–632. https://doi.org/10.1093/eurheartj/ehab395

[5] Maganti, K., Rigolin, V. H., Sarano, M. E., & Bonow, R. O. (2010). Valvular Heart Disease: Diagnosis and Management. Mayo Clinic Proceedings, 85(5), 483. https://doi.org/10.4065/mcp.2009.0706

[6] Ansari, Y., Mourad, O., Qaraqe, K., & Serpedin, E. (2023). Deep learning for ECG Arrhythmia detection and classification: An overview of progress for period 2017–2023. Frontiers in Physiology, 14. https://doi.org/10.3389/fphys.2023.1246746

[7] Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes (arXiv:1312.6114). arXiv. https://doi.org/10.48550/arXiv.1312.6114

[8] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System (arXiv:1603.02754). arXiv. http://arxiv.org/abs/1603.02754

[9] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework (arXiv:1907.10902). arXiv. http://arxiv.org/abs/1907.10902

[10] Gruwez, H., Evens, S., Proesmans, T., Smeets, C., Haemers, P., Pison, L., & Vandervoort, P. (2021). Head-to-head comparison of proprietary PPG and single-lead ECG algorithms for atrial fibrillation detection. EP Europace, 23(Supplement_3), euab116.524. https://doi.org/10.1093/europace/euab116.524