sensors

MDPI

Article

ECG Synthesis via Diffusion-Based State Space Augmented Transformer

Md Haider Zama [1] and Friedhelm Schwenker [2,*]

[1] Department of Computer Engineering, Jamia Millia Islamia, New Delhi 110025, India; haider1272002@gmail.com
[2] Institute of Neural Information Processing, Ulm University, 89081 Ulm, Germany
* Correspondence: friedhelm.schwenker@uni-ulm.de

Abstract: Cardiovascular diseases (CVDs) are a major global health concern, causing significant morbidity and mortality. AI's integration with healthcare offers promising solutions, with data-driven techniques, including ECG analysis, emerging as powerful tools. However, privacy concerns pose a major barrier to distributing healthcare data for addressing data-driven CVD classification. To address confidentiality issues related to sensitive health data distribution, we propose leveraging artificially synthesized data generation. Our contribution introduces a novel diffusion-based model coupled with a State Space Augmented Transformer. This synthesizes conditional 12-lead electrocardiograms based on the 12 multilabeled heart rhythm classes of the PTB-XL dataset, with each lead depicting the heart's electrical activity from different viewpoints. Recent advances establish diffusion models as groundbreaking generative tools, while the State Space Augmented Transformer captures long-term dependencies in time series data. The quality of generated samples was assessed using metrics like Dynamic Time Warping (DTW) and Maximum Mean Discrepancy (MMD). To evaluate authenticity, we assessed the similarity of performance of a pre-trained classifier on both generated and real ECG samples.

Keywords: electrocardiography; generative models; diffusion models; signal processing; time series; ECG synthesis

Citation: Zama, M.H.; Schwenker, F. ECG Synthesis via Diffusion-Based State Space Augmented Transformer. *Sensors* **2023**, *23*, 8328. https://doi.org/10.3390/s23198328

Academic Editor: Andrea Facchinetti

Received: 9 September 2023
Revised: 25 September 2023
Accepted: 3 October 2023
Published: 9 October 2023

1. Introduction

Amidst the landscape of global health challenges, cardiovascular diseases (CVDs) stand as a monumental concern, contributing significantly to morbidity and mortality rates. As the leading cause of death worldwide, their impact on public health and healthcare systems cannot be overstated. In recent years, the synergy between healthcare and artificial intelligence (AI) has yielded promising avenues for addressing these challenges. Data-driven AI techniques, in particular, have emerged as powerful tools in various medical domains, including the analysis of electrocardiograms (ECGs)—a cornerstone in diagnosing and managing CVDs.

Electrocardiography, commonly known as ECG, is a non-invasive method for monitoring the electrical activity of the heart over time. It provides critical insights into cardiac health and aids in the early detection, diagnosis, and monitoring of a wide range of heart-related conditions. ECG analysis involves the systematic interpretation of the heart's electrical signals, serving as a vital tool in the diagnosis of cardiac conditions. These ECG waveforms encompass distinct components like the P waves, QRS complexes, and T waves, each of which signifies specific phases of the cardiac cycle. The analysis encompasses the evaluation of factors such as heart rate, irregular rhythms (such as arrhythmias), deviations in the heart's electrical axis, and changes in the ST segment and T wave patterns to gauge the overall health of the heart muscle. Its clinical applications are extensive and encompass the diagnosis of arrhythmias, the identification of myocardial ischemia and infarction, the

continuous monitoring of pre-existing heart conditions, the assessment of treatment effects, and the evaluation of cardiac risk before surgical procedures.

Various AI methods are applied in the analysis and classification of ECG data, with a prominent focus on convolutional architectures, as highlighted in recent reviews [1,2]. Extensive research has demonstrated the effectiveness of modern convolutional architectures like ResNet and Inception in comparison to other methods, particularly on datasets like PTB-XL [3–5]. Additionally, innovative approaches have been proposed in the field, such as using large recurrent neural networks with convolutional feature extraction [6]. Furthermore, recent advancements involve the utilization of Structured State Space Models (SSMs) for classification, showing superior performance compared to both convolutional and recurrent network-based methods [7]. However, the effective utilization of ECG data for disease detection and diagnosis is not without its challenges.

One major impediment in data-driven CVDs is the difficulty in the distribution of healthcare data, primarily due to privacy concerns. Health data, including ECG recordings, contain sensitive information that demands strict protection to ensure patient privacy and adhere to data regulations. Consequently, the sharing and distribution of ECG datasets, crucial for training AI models, can be hampered by stringent privacy regulations. This limitation poses a significant obstacle to the development of accurate and robust ECG classification models.

The existing data protection techniques, while essential, often fall short of providing a comprehensive solution as even intricate technical strategies, such as the implementation of standalone federated learning [8], cannot unilaterally guarantee comprehensive privacy safeguards as training samples from trained models can be reconstructed, as illustrated by model inversion attacks [9]. This is where the importance of artificially synthesizing ECGs comes into play using generative machine learning models, such as Generative Adversarial Networks (GANs) [10]. GANs have demonstrated remarkable capabilities in generating synthetic data that closely resemble real-world distributions. In the context of ECGs, GANs offer the potential to synthesize samples while excluding personal information pertaining to patients, enabling researchers to develop and refine AI models without compromising patient confidentiality. Several pioneering works have explored the application of GANs for ECG synthesis [11–23].

However, GANs are not devoid of limitations. Challenges such as mode collapse, training instability, and the generation of clinically plausible ECG signals remain areas of concern. This has led to the emergence of diffusion models as an intriguing alternative. Diffusion models [24] offer a unique approach to data generation by iteratively 'diffusing' a simple distribution into the desired data distribution. The advantages of diffusion models include stability during training, controllable data generation, and the ability to capture long-range dependencies in the data.

Several recent works have demonstrated the efficacy of diffusion models in diverse applications, including image synthesis [25] and video synthesis [26]. Their potential to address some of the limitations of GANs makes them a promising technique for enhancing ECG synthesis. The major contributions of this study are as follows:

- Introduced a diffusion model for generating concise (10 s) 12-lead ECGs, incorporating a State Space Augmented Transformer [27] as a pivotal internal component.
- Enabled conditional generation of ECG samples, for different heart rhythms in a multilabel setting.
- Evaluated the quality of generated samples using metrics such as Dynamic Time Warping (DTW) and Maximum Mean Discrepancy (MMD). Additionally, to ensure authenticity, we evaluated how similar the performance of a pre-trained classifier is when applied to both generated and real ECG samples.

2. Related Work

The domain of deep generative modeling applied to time series data is an emerging and dynamic subfield in the realm of machine learning. This trajectory is substantially

guided by the persistent advancements achieved in the domain of generative models, particularly prominent in the domain of visual imagery. Amidst a spectrum of methodologies, GANs (Generative Adversarial Networks) [10] have captured considerable attention and gained prominence as the favored technique combined with LSTMs in [11–13], with the attention technique in [14], with transformers in [15], or with ordinary differential equations in [16]. Some of the recent advancements in GAN-based ECG generation proposed in [17] include a two-stage generator and dual discriminators, where the generator utilizes Gaussian noise to produce ECG representations while considering heart diseases as a conditioning factor, and in [18] the authors introduced automated solutions to integrate pre-existing shape knowledge of ECGs into the generation process. Some other GAN-based approaches are [19–23]. In addition to GANs, alternative architectures are also proposed for ECG generation like variational autoencoders [28]. Recently, diffusion models have risen as a notably potent category of deep generative models suited for the objective of ECG synthesis exhibiting benefits beyond GANs; these models offer improved training stability and excel in generating higher-quality ECG samples [29–32].

3. Materials and Methods

3.1. Dataset

The conducted experiments were carried out using the PTB-XL dataset [3–5], a publicly accessible collection of electrocardiogram (ECG) data. This dataset encompasses a total of 21,837 records obtained from 18,885 distinct patients, each of whom possesses clinical 12-lead ECG data. These records consist of concise 12-lead ECG signals, each spanning a duration of 10 s and sampled at a rate of 100 Hz.

To facilitate the conditional generation of ECG samples, we employed 12 distinct heart rhythms in a multilabel setting as class conditionals. This set of labels allowed us to effectively guide the generative process.

In order to create a robust and comprehensive evaluation setup, we partitioned the dataset into three subsets: the train-set containing 17,441 samples, the test-set encompassing 2203 samples, and the validation-set comprising 2193 samples. This partitioning strategy was devised based on individual patient identifiers (person-ids), ensuring that records associated with the same person-id did not appear across different subsets. By adopting this approach, we aimed to guarantee a stringent assessment of the model's ability to generalize its learnings. Moreover, this technique helped eliminate any potential biases that could arise due to data overlap between the distinct sets. Overall, our approach to data splitting was integral to maintaining the integrity of the evaluation process.

3.2. Diffusion Models

Diffusion models [24] are a category of generative models involving two processes: a forward process and a reverse process. Throughout the forward process, Gaussian noise is incrementally infused into the input sample over a sequence of T steps, following a variance schedule β (constant or learnable). This progressive introduction of noise continues until the input distribution converges to a standard Gaussian distribution $\mathcal{N}(0, 1)$. In contrast, during the reverse diffusion process, a neural network parameterized as θ is trained with the purpose of noise reduction from the sample. The forward process can be expressed as

$$q(x_1, ..., x_T|x_0) = \prod_{t=1}^{T} q(x_t|x_t - 1) \quad (1)$$

where $q(x_t|x_t - 1) = \mathcal{N}(x_t, \sqrt{1 - \beta_t}x_t - 1, \beta_t I)$. Also. using the reparameterization trick, x_t can be expressed as $x_t = \sqrt{\overline{\alpha_t}}x_0 + (1 - \overline{\alpha_t})\epsilon$ for $\epsilon \sim \mathcal{N}(0, 1)$, $\alpha_t = 1 - \beta_t$ and $\overline{\alpha_t} = \prod_{i=1}^{T} \alpha_i$. The reverse process learns to undo this noise-inducing procedure to recover x_0 from x_t, initiating with the pure Gaussian noise $x_T \sim \mathcal{N}(0, 1)$. The reverse process can be expressed as

$$p_\theta(x_0, ..., x_T - 1|x_T) = \prod_{t=1}^{T} p_\theta(x_t - 1|x_t) \tag{2}$$

where $p_\theta(x_t - 1|x_t) = \mathcal{N}(x_t - 1, \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$. The loss function for the reverse process is obtained by computing the variational lower bound:

$$\mathbb{E}[-\log p_\theta(x_0)] \leq \mathbb{E}_q[-\log \frac{p_\theta(x_0 : T)}{q(x_1 : T|x_0)}] = \mathbb{E}_q[-\log p(x_T) - \sum_{t\geq 1} \log \frac{p_\theta(x_t - 1|x_t)}{q(x_t|x_t - 1)}] = L \tag{3}$$

It was shown in [33] that the loss function can be further simplified as

$$L_{simple} = \mathbb{E}_t, x0, \epsilon[||\epsilon - \epsilon_\theta(x_t, t)||^2] \tag{4}$$

where $\epsilon_\theta(x_t, t)$ is predicted by the neural network and ϵ is the actual noise added to the sample x_t. Class-specific diffusion models can be implemented by conditioning the reverse process on a desired set of labels, denoted as 'c'. In other words, this involves utilizing $\epsilon_\theta = \epsilon_\theta(x_t, t, c)$.

3.3. Structured State Space Models

At its core, Structured State Space Models (SSSMs) rely on a linear state space transition equation that establishes a link between a one-dimensional input sequence $u(t)$ and a one-dimensional output sequence $y(t)$. This connection is mediated by an N-dimensional hidden state $x(t)$,

$$x^{'}(t) = Ax(t) + Bu(t), \;\; y(t) = Cx(t). \tag{5}$$

Here, A, B, C are the transformation matrices. It was shown in [34] that the typical utilization of randomly initialized parameters A, B, C, and D proves inadequate in effectively representing long-range dependencies. Hence, a category of matrices known as HiPPOs (high-order polynomial projection operators) was introduced to serve as the initialization for A. These HiPPO matrices are crafted with the specific objective of enabling parameter A to retain a memory of the input history $u(t)$ until time t, thereby enhancing the ability of the state $x(t)$ at time t to capture this historical context. In real-world applications, we commonly handle discrete sequences with some sequence length L. Hence, the discretized form of Equation (5) can be expressed as

$$x_k = \overline{A}x_{k-1} + \overline{B}u_k, y = \overline{C}x_k, \tag{6}$$

where $\overline{A} = (I - \delta/2 \cdot A)^{-1}(I + \delta/2 \cdot A)$, $\overline{B} = (I - \delta/2 \cdot A)^{-1}\delta B$, $\overline{C} = C$. And $\delta > 0$ is the small step size. After expanding the above recurrent representation, we obtain

$$y_k = \overline{CA^K B}u_0 + ... + \overline{CAB}u_{k-1} + \overline{CB}u_k \tag{7}$$

The convolutional representation is

$$\begin{aligned} &y = \overline{K} * u, \\ &\overline{K} \in \mathbb{R}^L = (\overline{CB}, \overline{CAB}, ..., \overline{CA^{L-1}B}). \end{aligned} \tag{8}$$

In [35], a Structured State Space Sequence model (S4) was proposed to efficiently compute Equation (8), where A and B are initialized as

$$\begin{aligned} &A = A^{(d_s)} - PP^T, \;\; B_i = (2i + 1)^{\frac{1}{2}}, \\ &\text{where } P_i = (i + 1/2)^{\frac{1}{2}}, \\ &A_{ij}^{d_s} = -\begin{cases} (i + \frac{1}{2})^{\frac{1}{2}}(j + \frac{1}{2})^{\frac{1}{2}}, & \text{if } i > j \\ \frac{1}{2}, & \text{if } i = j \\ -(i + \frac{1}{2})^{\frac{1}{2}}(j + \frac{1}{2})^{\frac{1}{2}}, & \text{if } i < j \end{cases} \end{aligned} \tag{9}$$

Following that, the convolution kernel K described in Equation (8) can be efficiently computed with a time and space complexity of $O(L)$. Therefore, when given an input u, the S4 output $y = K * u$ can also be computed efficiently.

3.4. State Space Augmented Transformer

In [27], the authors introduced SPADE (State sPace AugmenteD transformEr), a multi-layer transformer model designed to capture intricate global and local information. The architecture of SPADE is illustrated in Figure 1 (left). This model utilizes a hierarchical design. At its foundational global layer, SPADE employs an SSM (State Space Model) to grasp global dependencies. As the SSM primarily offers broad global insights, the subsequent local layers empower the model to manage more intricate and nuanced local dependencies. In essence, the SSM introduces a robust structural bias that enriches global information within the inputs. For the local layer instantiation, the authors replace the full attention found in conventional transformer layers with effective pre-existing local attention techniques. Within the global layer (as shown in Figure 1, right), when presented with the input X , the resulting output Y is obtained as follows:

$$\begin{aligned} X_{local} &= \text{Local}(\text{LN}(X)), \\ X_{global} &= \text{SSM}(\text{LN}(X)), \\ X_a &= \text{W}[\text{LN}(X_{local}), \text{LN}(X_{global})] + X, \\ Y &= \text{FFN}(\text{LN}(X_a)) + X_a. \end{aligned} \tag{10}$$

In the above Equation (10), LN(.) represents layer normalization [36], while FFN$(\cdot)$ signifies a two-layer feed-forward neural network. The SSM used is S4.

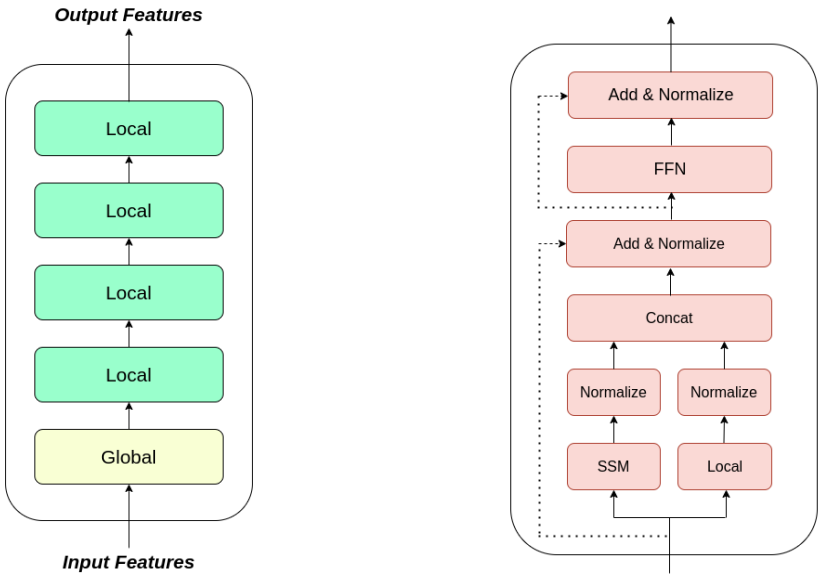

Figure 1. Illustration of a 4-layered SPADE: on the left: an outline of the model; on the right: intricacies of the global layer.

It was shown in [27] that SPADE turns out to be efficient and effective in various natural language processing tasks. It outperforms existing methods in the long-range arena benchmark [37]. It also exhibits significantly improved speed and attains superior performance over the vanilla transformer [38] in autoregressive language modeling.

3.5. Proposed Approach: DSAT-ECG

The proposed model architecture DSAT (Diffusion State Space Augmented Transformer)-ECG takes inspiration from the SSSD-ECG [29] architecture, which is based on DiffWave, i.e., a diffusion-based model for audio synthesis [39]. In this improved design, the S4 [35] has been substituted with SPADE [27] layers. These SPADE layers now function as diffu-

sion layers within every residual block, seamlessly integrating the diffusion embedding. Furthermore, the conventional second S4 layer has been replaced by an additional SPADE layer, which is seamlessly integrated after the addition assignment, along with conditional information. The model architecture is depicted in Figure 2.

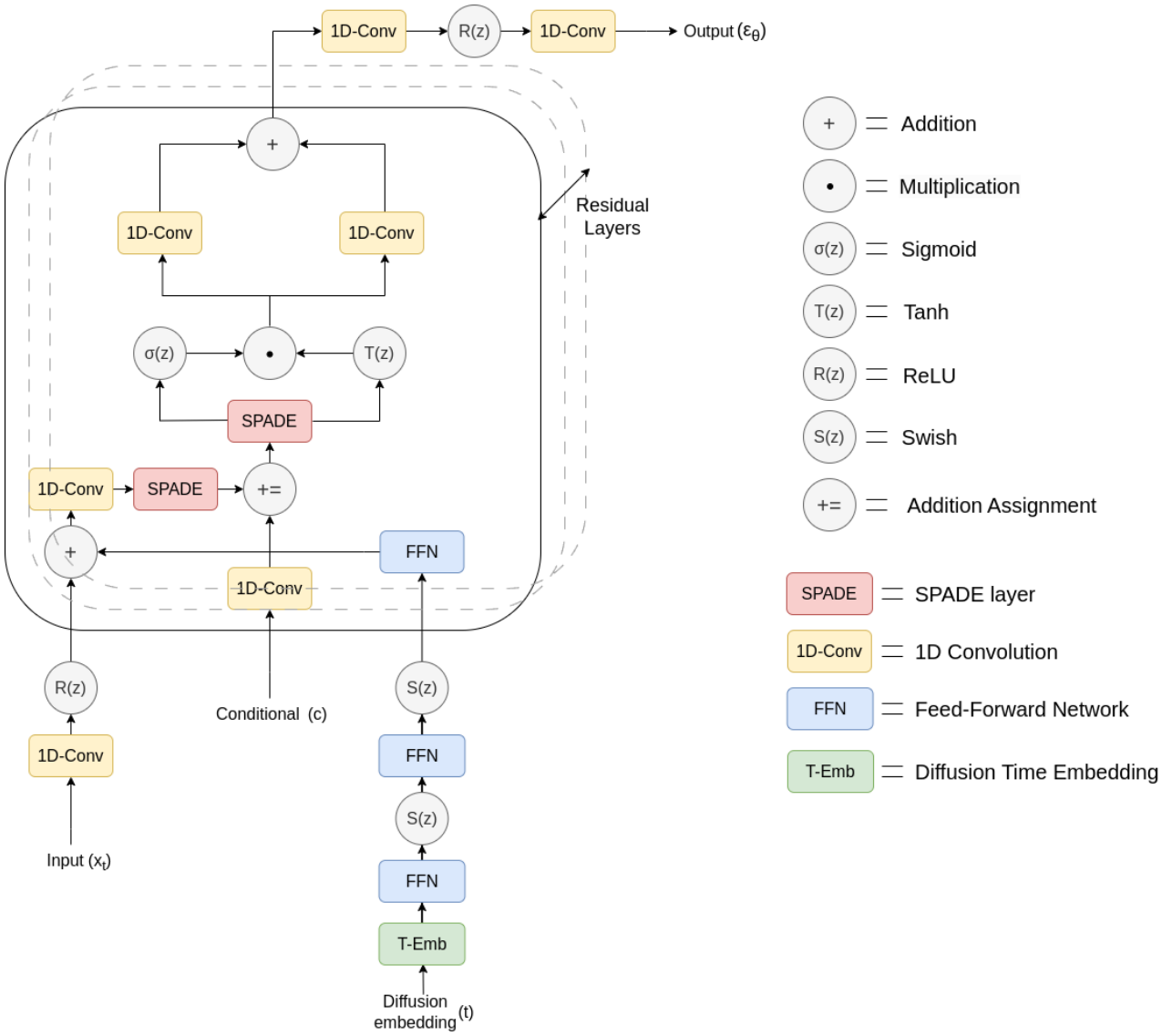

Figure 2. Proposed DSAT-ECG model architecture.

During the training phase, the input samples 'x_0' are noised as a part of the forward diffusion process (see Section 3.2) at different diffusion time steps 't' sampled uniformly from 1 to T transforming the samples to 'x_t'. Next, the model receives 'x_t', 't', and 'c' as input and outputs 'ϵ_0' , where 'c' is the class conditional; in this work, it is a multilabeled vector of length 12 representing 46 unique heart rhythm labels and 'ϵ_0' is the predicted noise in the samples. For the model to converge, gradient descent is applied to MSE regression loss between the actual noise and the predicted noise. It is mathematically represented as

$$\nabla_\theta \left|\left|\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha_t}}x_0 - \sqrt{1-\overline{\alpha_t}}\epsilon, t, c)\right|\right|^2 \tag{11}$$

During the inference phase, x_T is sampled as a pure Gaussian noise, i.e., $x_T \sim \mathcal{N}(0,1)$, and as a part of the reverse diffusion process (see Section 3.2) noise is iteratively removed

from it, starting from diffusion time step T to 0 and transforming it to x_0. It is mathematically represented as

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t, c)\right) + \sigma_t z \tag{12}$$

where σ_t is the fixed variance schedule and $z \sim \mathcal{N}(0,1)$ *if* $t > 1$, *else* z is set to 0.

Following [21,29] we also focused on training generative models to synthesize a set of eight leads. This set includes the six precordial leads as well as two limb leads (namely, leads I and aVF in this work). Following this, we employ sampling from the two limb leads, I and aVF, to reconstruct the remaining four leads using Einthoven's law and Goldberger's equation as

$$\begin{aligned} III &= II - I, \\ aVL &= (I - III)/2, \\ aVF &= (II + III)/2, \\ -aVR &= (I + II)/2. \end{aligned} \tag{13}$$

4. Results

In this section, we describe the process through which we conducted a comprehensive assessment of the proposed model's quality and authenticity, drawing inspiration from the evaluation approach used in [32]. The evaluation involved a direct comparison between the proposed model and several other existing models that were proposed in [29], namely SSSD-ECG, WaveGAN, and P2PGAN. We sought to offer a comprehensive perspective on how our model showcases advancements over existing methodologies, highlighting its promising implications, particularly in the realm of ECG synthesis. Figure 3 shows a 12-lead ECG sample generated by different models.

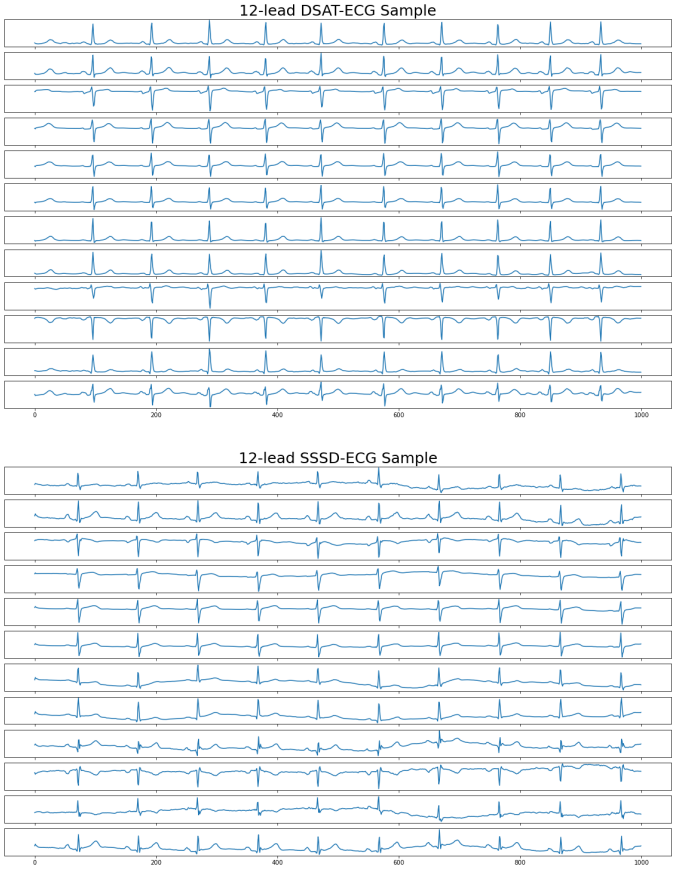

Figure 3. *Cont.*

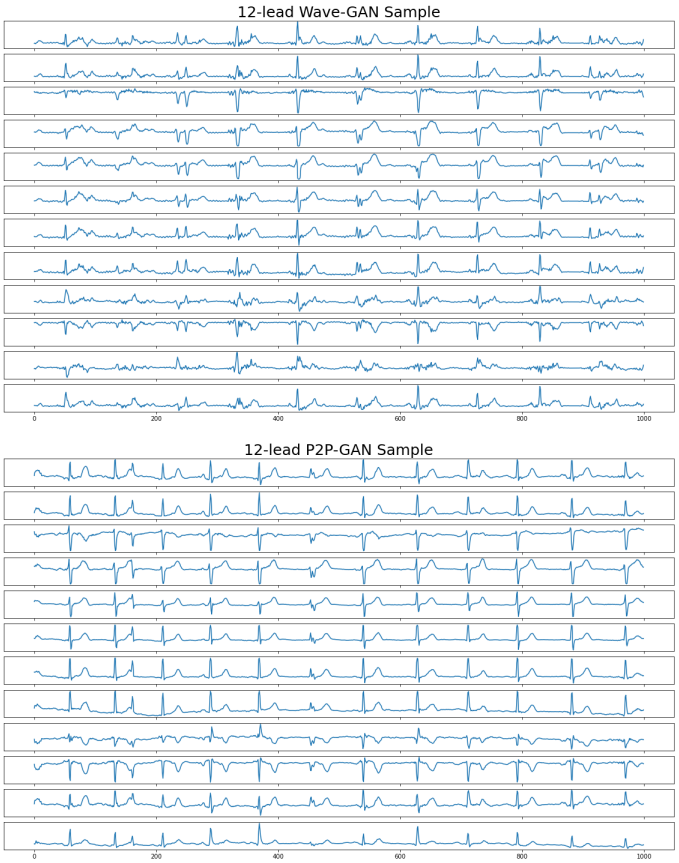

Figure 3. Twelve-lead ECG sample generated by different models.

4.1. Training Settings

The implementation of all models was carried out utilizing the PyTorch Library, and the training process took place on Kaggle's P100 GPU featuring 16 GB of RAM. For both the diffusion-based models, namely DSAT-ECG (ours) and SSSD-ECG, the configuration included four stacked residual layers each. These layers consisted of 256 residual channels and 256 skip channels, incorporated with a three-level diffusion embedding across dimensions of 128, 256, and 256. Both models employed a linear schedule involving 200 diffusion time steps, and the noise scheduler beta was adjusted from 0.0001 to 0.02. For optimization, the Adam optimizer was chosen with a learning rate of 2×10^{-4} . The GAN-based model closely followed the configuration outlined in [29]. This encompassed a generator model size of 50, utilizing five deconvolutional blocks, and having 1000 latent dimensions. The discriminator, on the other hand, featured a model size of 50, incorporating six convolutional blocks. The optimization process employed the Adam optimizer with a learning rate of 1×10^{-4}.

4.2. Quality Evaluation

While evaluating the quality of the ECG samples, we essentially measure how much the generated synthetic ECG samples resemble real ECG samples, encompassing visual traits and morphological attributes. To quantitatively measure the quality of these generated ECG samples, we make use of two distinct distance functions: Dynamic Time Warping (DTW) [40] and Maximum Mean Discrepancy (MMD) [41]. DTW is a method utilized to determine the similarity between two time series sequences. It finds the optimal alignment between two sequences while considering temporal distortions. It is mathematically expressed as

$$D_{i,j} = f(x_i, y_j) + \min\{D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1}\} \tag{14}$$

The iteration is conducted for values of i ranging from 1 to N and values of j ranging from 1 to M, where N and M represent the lengths of the series x and y, respectively. Typically, the function $f(x_i, y_j)$ is defined as the square of the difference between x_i and y_j, that is, $(x_i - y_j)^2$.

MMD is another technique employed to measure the dissimilarity between two sets of data, which could be probability distributions or samples. It aims to distinguish differences in distribution characteristics and can effectively detect variations in data points. It is mathematically expressed as

$$MMD(X,Y) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k(x_i, x_j) + \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j \neq i}^{m} k(y_i, y_j) - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} k(x_i, y_j), \tag{15}$$

In the context of the provided equation, the symbols hold the following interpretations: n represents the number of samples in the set X, while m corresponds to the number of samples in the set Y. The variables x_i and y_i denote individual samples taken from sets X and Y, respectively. Furthermore, the expression $k(x, y)$ signifies the Gaussian radial basis function (RBF) kernel, specifically defined as $k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$, where σ stands for the kernel bandwidth parameter.

The outcomes of both distance metrics consistently demonstrate that the ECG samples produced by the proposed DSAT-ECG model are substantially more similar to the actual samples in terms of their quality (see Table 1 below).

Table 1. Quality evaluation of ECG samples.

Model	Avg DTW (Lower is Better)	Avg MMD (Lower is Better)
WaveGAN	8.057	6.418×10^{-3}
P2PGAN	8.430	6.381×10^{-3}
SSSD-ECG	7.481	4.784×10^{-3}
DSAT-ECG (Ours)	**7.174**	$\mathbf{4.603 \times 10^{-3}}$

Best results are shown in bold. All the models were evaluated with a batch size of 16.

4.3. Authenticity Evaluation

While evaluating the authenticity of the ECG samples, we assess the similarity in performance of a pre-trained classifier, which has been trained on actual ECG data when applied to both synthetic test samples and real ECG test samples. To accurately gauge the authenticity of these generated ECG samples, we employ two quantitative metrics: multilabel accuracy and multilabel AUROC (Area Under the Receiver Operating Characteristic Curve).

Multilabel accuracy is calculated as the ratio of the number of correctly predicted labels to the total number of labels, expressed as

$$\text{Multilabel Accuracy} = \frac{\text{Number of Correctly Predicted Labels}}{\text{Total Number of Labels}} \tag{16}$$

In multilabel AUROC, for each label, the ROC curve is created by plotting the true-positive rate *TPR* against the false-positive rate *FPR* at different classification thresholds. The multilabel AUROC is the average of the individual AUROC values for each label, expressed as

$$\text{Multilabel AUROC} = \frac{1}{\text{Number of Labels}} \left(\sum_{i=1}^{\text{Number of Labels}} \text{AUROC}_i \right) \tag{17}$$

where AUROC_i is the AUROC value for the ith label. The AUROC is calculated using the integral under the ROC curve.

For the classification task, we adopt the ResNet-1D model proposed in [42], which represents a 1D variant of the ResNet architecture. We trained the ResNet-1D model on the

train-set of the PTB-XL dataset (see Section 3.1) and evaluated the classifier on the real test set and on the test sets generated by the models. The result of the classifier on different samples is shown in Table 2 below. The test shows that the classifier's performance on samples generated by DSAT-ECG is most closest to its performance on real ECG samples compared to other models, affirming the fidelity of the DSAT-ECG model in synthesizing authentic ECG characteristics.

Table 2. Authenticity evaluation of ECG samples.

Test Sample	Accuracy (Higher is Better)	AUROC (Higher is Better)
WaveGAN	89.32	88.34
P2PGAN	89.67	89.58
SSSD-ECG	95.01	93.98
DSAT-ECG (Ours)	**95.84**	**94.56**
Real	***96.38***	***94.67***

Best results on the synthesized sample are shown in bold and results on the real sample are shown in bold italics. All samples were evaluated in batches of batch size of 16.

5. Discussion

The results obtained from the evaluation of our DSAT-ECG model reveal compelling insights into the efficacy and potential impact of our approach. The performance of our model, both in terms of quality and authenticity assessment, indicates notable progress in the field of ECG data synthesis.

Our quality assessment, employing the metrics Dynamic Time Warping (DTW) and Maximum Mean Discrepancy (MMD), illuminates DSAT-ECG's ability to emulate intricate ECG patterns. The notably lower DTW and MMD scores, in contrast to competing models, highlight our model's capability to capture nuanced temporal dynamics and the ability to preserve essential statistical properties inherent in real ECG data.

The assessment of authenticity, a key determinant in the data synthesis landscape, produced notable results for DSAT-ECG. The attainment of higher accuracy and AUROC metrics than the competing models signifies our model's proficiency in generating data that resonate with the authenticity of genuine ECG data. This alignment is further substantiated by the observation that classifier performance on our model's generated samples closely resembles that of real samples. These outcomes collectively reinforce the genuine nature of the synthesized data and accentuate the model's suitability for applications requiring reliable and authentic data representation.

The progress in ECG data synthesis holds significance for both research and practical uses. The idea of improving data quality opens doors for better medical research, more accurate diagnostics, and improved clinical applications. This aligns with the need to handle private health data carefully. This is where creating artificial data becomes important, offering a way to balance data usefulness with privacy concerns.

For future research, we will focus on shifting from fixed to trainable variances for enhanced adaptability in capturing data intricacies, exploring alternative schedules beyond linear trajectories to optimize model convergence patterns, Furthermore, we plan to expand the loss function beyond MSE for better results and we will also try other variants of diffusion, like latent diffusion.

6. Conclusions

In conclusion, our study introduces DSAT-ECG, a State Space Augmented Transformer model trained using the diffusion technique. This approach addresses ECG data synthesis across a set of 12 distinct heart rhythms in a multilabel context, signifying the synthesis of ECG on a particular condition. The performance of the generated samples excelled across qualitative and authenticity assessments, outperforming the competitor models SSSD-ECG, WaveGAN, and P2PGAN employed for evaluation. Nevertheless, it is worth acknowledging that the DSAT-ECG samples do not completely substitute real samples

for classifier training. This difference presents an opportunity for future advancements, demonstrating the potential for progress in the field of ECG synthesis in the near future. While the superiority of the proposed diffusion model over GAN-based models is evident, it is important to note that the computational time required for both training and inference is considerably longer. This observation serves as motivation for us to explore and develop faster diffusion-based models, such as the latent diffusion model, for ECG synthesis in our future work.

Author Contributions: Conceptualization, M.H.Z.; methodology, M.H.Z.; software, M.H.Z.; validation, M.H.Z. and F.S.; data curation, M.H.Z.; writing—original draft preparation, M.H.Z.; writing—review and editing, M.H.Z. and F.S.; visualization, M.H.Z.; supervision, F.S.; project administration, F.S.; funding acquisition, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hong, S.; Zhou, Y.; Shang, J.; Xiao, C.; Sun, J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Comp. Biol. Med.* **2020**, *122*, 103801. [CrossRef] [PubMed]
2. Petmezas, G.; Stefanopoulos, L.; Kilintzis, V.; Tzavelis, A.; Rogers, J.A.; Katsaggelos, A.K.; Maglaveras, N. State-of-the-Art Deep Learning Methods on Electrocardiogram Data: Systematic Review. *JMIR Med. Inform.* **2022**, *10*, e38454. [CrossRef]
3. Wagner, P.; Strodthoff, N.; Bousseljot, R.D.; Samek, W.; Schaeffter, T. PTB-XL, a Large Publicly Available Electrocardiography Dataset. (Version 1.0.3). 2022. Available online: https://physionet.org/content/ptb-xl/1.0.3/ (accessed on 8 September 2023).
4. Wagner, P.; Strodthoff, N.; Bousseljot, R.D.; Kreiseler, D.; Lunze, F.I.; Samek, W.; Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *Sci. Data* **2020**, *7*, 154 . [CrossRef] [PubMed]
5. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **2000**, *101*, E215–E220. [CrossRef] [PubMed]
6. Mehari, T.; Strodthoff, N. Self-supervised representation learning from 12-lead ECG data. *Comp. Biol. Med.* **2021**, *141*, 105114. [CrossRef]
7. Mehari, T.; Strodthoff, N. Advancing the State-of-the-Art for ECG Analysis through Structured State Space Models. *arXiv* **2022**, arXiv:2211.07579.
8. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated Learning for Healthcare Informatics. *J. Healthcare Inform. Res.* **2020**, *5*, 1–19. [CrossRef]
9. Yin, H.; Mallya, A.; Vahdat, A.; Álvarez, J.M.; Kautz, J.; Molchanov, P. See through Gradients: Image Batch Recovery via GradInversion. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 16332–16341.
10. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:stat.ML/1406.2661.
11. Delaney, A.M.; Brophy, E.; Ward, T.E. Synthesis of Realistic ECG using Generative Adversarial Networks. *arXiv* **2019**, arXiv:eess.SP/1909.09150.
12. Zhu, F.; Fei, Y.; Fu, Y.; Liu, Q.; Shen, B. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Sci. Rep.* **2019**, *9*, 6734. [CrossRef]
13. Golany, T.; Lavee, G.; Yarden, S.; Radinsky, K. Improving ECG Classification Using Generative Adversarial Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13280–13285. [CrossRef]
14. Rafi, T.H.; Woong Ko, Y. HeartNet: Self Multihead Attention Mechanism via Convolutional Network With Adversarial Data Synthesis for ECG-Based Arrhythmia Classification. *IEEE Access* **2022**, *10*, 100501–100512. [CrossRef]
15. Li, X.; Metsis, V.; Wang, H.; Ngu, A. *TTS-GAN: A Transformer-Based Time-Series Generative Adversarial Network*; Springer: Berlin, Germnay, 2022; pp. 133–143. [CrossRef]
16. Golany, T.; Freedman, D.; Radinsky, K. ECG ODE-GAN: Learning Ordinary Differential Equations of ECG Dynamics via Generative Adversarial Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2021; Volume 35, pp. 134–141. [CrossRef]

17. Chen, J.; Liao, K.; Wei, K.; Ying, H.; Chen, D.Z.; Wu, J. ME-GAN: Learning Panoptic Electrocardio Representations for Multi-view ECG Synthesis Conditioned on Heart Diseases. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MA, USA, 17–23 July 2022; Volume 162, pp. 3360–3370.
18. Neifar, N.; Mdhaffar, A.; Ben-Hamadou, A.; Jmaiel, M.; Freisleben, B. Disentangling temporal and amplitude variations in ECG synthesis using anchored GANs. In Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, Tallinn, Estonia, 27–31 March 2023; pp. 645–652. [CrossRef]
19. Golany, T.; Radinsky, K.; Freedman, D. SimGANs: Simulator-Based Generative Adversarial Networks for ECG Synthesis to Improve Deep ECG Classification. In Proceedings of the 37th International Conference on Machine Learning, Virtual Event, 13–18 July 2020; pp. 3597–3606.
20. Golany, T.; Radinsky, K. PGANs: Personalized Generative Adversarial Networks for ECG Synthesis to Improve Patient-Specific Deep ECG Classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 557–564. [CrossRef]
21. Thambawita , V.; Isaksen, J.; Hicks, S.; Ghouse, J.; Ahlberg, G.; Linneberg, A.; Grarup, N.; Ellervik, C.; Olesen, M.; Hansen, T.; et al. DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Sci. Rep.* **2021**, *11*, 21896. [CrossRef] [PubMed]
22. Li, W.; Tang, Y.M.; Yu, K.M.; To, S. SLC-GAN: An automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis. *Inform. Sci.* **2022**, *589*, 738–750. [CrossRef]
23. Luo, Y.; Zhang, Y.; Cai, X.; Yuan, X. E²GAN: End-to-End Generative Adversarial Network for Multivariate Time Series Imputation. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, Macao, China, 10–16 August 2019; pp. 3094–3100. [CrossRef]
24. Sohl-Dickstein, J.; Weiss, E.A.; Maheswaranathan, N.; Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv* **2015**, arXiv:cs.LG/1503.03585.
25. Mukhopadhyay, S.; Gwilliam, M.; Agarwal, V.; Padmanabhan, N.; Swaminathan, A.; Hegde, S.; Zhou, T.; Shrivastava, A. Diffusion Models Beat GANs on Image Classification. *arXiv* **2023**, arXiv:cs.CV/2307.08702.
26. Ho, J.; Salimans, T.; Gritsenko, A.; Chan, W.; Norouzi, M.; Fleet, D.J. Video Diffusion Models. *arXiv* **2022**, arXiv:cs.CV/2204.03458.
27. Zuo, S.; Liu, X.; Jiao, J.; Charles, D.; Manavoglu, E.; Zhao, T.; Gao, J. Efficient Long Sequence Modeling via State Space Augmented Transformer. *arXiv* **2022**, arXiv:cs.CL/2212.08136.
28. Sang, Y.; Beetz, M.; Grau, V. Generation of 12-Lead Electrocardiogram with Subject-Specific, Image-Derived Characteristics Using a Conditional Variational Autoencoder. In Proceedings of the 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 28–31 March 2022; pp. 1–5. [CrossRef]
29. Alcaraz, J.M.L.; Strodthoff, N. Diffusion-based conditional ECG generation with structured state space models. *Comp. Biol. Med.* **2023**, *163*, 107115. [CrossRef]
30. Alcaraz, J.M.L.; Strodthoff, N. Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models. *arXiv* **2022**, arXiv:abs/2208.09399.
31. Neifar, N.; Ben-Hamadou, A.; Mdhaffar, A.; Jmaiel, M. DiffECG: A Generalized Probabilistic Diffusion Model for ECG Signals Synthesis. *arXiv* **2023**, arXiv:cs.CV/2306.01875.
32. Adib, E.; Fernandez, A.S.; Afghah, F.; Prevost, J.J. Synthetic ECG Signal Generation Using Probabilistic Diffusion Models. *IEEE Access* **2023**, *11*, 75818–75828. [CrossRef]
33. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 6840–6851.
34. Gu, A.; Dao, T.; Ermon, S.; Rudra, A.; Ré, C. HiPPO: Recurrent Memory with Optimal Polynomial Projections. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 1474–1487.
35. Gu, A.; Goel, K.; Ré, C. Efficiently Modeling Long Sequences with Structured State Spaces. *arXiv* **2022**, arXiv:cs.LG/2111.00396.
36. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer Normalization. *arXiv* **2016**, arXiv:stat.ML/1607.06450.
37. Tay, Y.; Dehghani, M.; Abnar, S.; Shen, Y.; Bahri, D.; Pham, P.; Rao, J.; Yang, L.; Ruder, S.; Metzler, D. Long Range Arena : A Benchmark for Efficient Transformers. In Proceedings of the International Conference on Learning Representations, Virtual Event, 3–7 May 2021.
38. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.u.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
39. Kong, Z.; Ping, W.; Huang, J.; Zhao, K.; Catanzaro, B. DiffWave: A Versatile Diffusion Model for Audio Synthesis. In Proceedings of the International Conference on Learning Representations, Virtual Event, 3–7 May 2021.
40. Müller, M. Dynamic time warping. *Inform. Retrieval Music Motion* **2007**, *2*, 69–84. [CrossRef]

41. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A Kernel Two-Sample Test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
42. Hong, S.; Xu, Y.; Khare, A.; Priambada, S.; Maher, K.; Aljiffry, A.; Sun, J.; Tumanov, A. HOLMES: Health OnLine Model Ensemble Serving for Deep Learning Models in Intensive Care Units. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, 6–10 July 2020; pp. 1614–1624.