

Review

Unraveling Arrhythmias with Graph-Based Analysis: A Survey of the MIT-BIH Database

Sadiq Alinsaif 

College of Computer Science and Engineering, University of Hafr Al Batin, Al Jamiah,
Hafar Al Batin 39524, Saudi Arabia; alinsaif@uhb.edu.sa

Abstract: Cardiac arrhythmias, characterized by deviations from the normal rhythmic contractions of the heart, pose a formidable diagnostic challenge. Early and accurate detection remains an integral component of effective diagnosis, informing critical decisions made by cardiologists. This review paper surveys diverse computational intelligence methodologies employed for arrhythmia analysis within the context of the widely utilized MIT-BIH dataset. The paucity of adequately annotated medical datasets significantly impedes advancements in various healthcare domains. Publicly accessible resources such as the MIT-BIH Arrhythmia Database serve as invaluable tools for evaluating and refining computer-assisted diagnosis (CAD) techniques specifically targeted toward arrhythmia detection. However, even this established dataset grapples with the challenge of class imbalance, further complicating its effective analysis. This review explores the current research landscape surrounding the application of graph-based approaches for both anomaly detection and classification within the MIT-BIH database. By analyzing diverse methodologies and their respective accuracies, this investigation aims to empower researchers and practitioners in the field of ECG signal analysis. The ultimate objective is to refine and optimize CAD algorithms, ultimately culminating in improved patient care outcomes.

Keywords: deep learning; fine-tuning; CNN-based features; classification; anomaly detection; cardiac arrhythmia; ECG; graph theory



Citation: Alinsaif, S. Unraveling Arrhythmias with Graph-Based Analysis: A Survey of the MIT-BIH Database. *Computation* **2024**, *12*, 21. <https://doi.org/10.3390/computation12020021>

Academic Editors: Akbar Ali, Guojun Li, Mingchu Li, Rao Li, Colton Magnant and Madhumangal Pal

Received: 3 January 2024

Revised: 22 January 2024

Accepted: 23 January 2024

Published: 25 January 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Despite its origin in 1980, the Massachusetts Institute of Technology–Beth Israel Hospital (MIT-BIH) electrocardiogram (ECG) Arrhythmia Database retains a substantial and enduring influence on the field of arrhythmia characterization and detection, arguably surpassing initial expectations [1]. This longevity coincides with advancements in artificial and computational intelligence, enabling new analytical and interpretive models for ECG-based diagnoses of arrhythmias and other cardiac pathologies [2–6]. These techniques can be broadly categorized into:

1. **Traditional learning-based approaches:** employing classical machine learning (CML) algorithms and established feature extraction/selection methods.
2. **Deep learning (DL) approaches:** leveraging deep features obtained through training from scratch, model fine-tuning, or hybrid configurations combining traditional descriptors with deep-feature representations.

The application of machine learning (ML) for ECG data analysis holds significant promise in the development of prognostic and diagnostic CAD systems. ECG CAD systems can serve as a valuable tool for medical professionals, facilitating objective diagnosis [3]. The association between different ECG records can be established through supervised [7], semi-supervised [8], or unsupervised [9] ML approaches. Supervised learning entails training a model on a labeled dataset where ground-truth labels are known for each record. In the context of medical data classification, prominent supervised learning algorithms employed include multilayer perceptrons (MLPs) [10] and support vector machines (SVMs) [11].

Conversely, in situations where ground-truth labels are unavailable, unsupervised learning can be utilized to discover latent patterns within the data. Examples of such algorithms include k-means clustering (k-means) [12] and principal component analysis (PCA) [13].

Traditional ML approaches rely on the extraction of informative features that effectively represent the underlying disease [14]. Success in this endeavor hinges on the extraction of numerical measurements that inherently manifest the disease characteristics [15]. When a chosen feature extraction method effectively captures the pathological signatures of the desired phenomenon (disease), the subsequent application of an ML algorithm is more likely to yield accurate disease prediction outcomes [15].

The recent paradigm shift in ECG data analysis has gravitated towards the application of DL techniques. Unlike traditional methods, DL offers generic, non-domain-specific operation sequences directly applicable to raw input signals, including ECG records [3–5,16–20]. A prominent example of DL architecture is the convolutional neural network (CNN), which demonstrably exhibits efficacy in ECG data analysis [3–5,16–20]. The inherent strength of DL models lies in their ability to learn and discover multilevel representations from data. Lower-level layers typically extract fundamental features like edges and color, while higher layers progressively abstract these features into semantically meaningful representations of the input [21]. This characteristic has spurred an active research field exploring the transferability of knowledge gained from pretrained models to the domain of ECG arrhythmia detection [8,22]. Alternatively, pretrained CNN models can be utilized as unsupervised feature extractors, bypassing the need for fine-tuning [23]. Moreover, hybrid approaches combining hand-crafted features with deep features extracted from pretrained models are also being explored [24].

A proliferating body of research in ECG data classification and anomaly detection has employed the MIT-BIH database as a foundational resource. Analyzing the ECG signals is crucial for diagnosing and managing cardiovascular diseases. Traditional approaches often rely on preprocessing steps and feature engineering, but these can be cumbersome and susceptible to error. However, graph-based methods have emerged as a powerful tool for uncovering complex relationships within ECG data, offering a potential paradigm shift in ECG analysis. Additionally, existing reviews often lack a granular focus on graph-based analysis and modeling. This paper addresses this gap by providing a survey of recent advances in feature extraction and DL methods applied to arrhythmia detection within the MIT-BIH ECG dataset.

Noteworthy, the applications of graph-based theory in biomedicine extend far beyond the confines of the MIT-BIH database, captivating researchers across diverse domains. Its utility has been demonstrated, e.g., in: (1) epilepsy detection: utilizing electroencephalogram (EEG) data, graph-based approaches effectively distinguish epileptic seizures from normal brain activity [25]; (2) functional brain connectivity: delineating the complex relationship within the brain, a graph analysis of EEG data sheds light on functional connectivity patterns [26]; (3) essential gene identification: within protein–protein interaction networks, graph-based algorithms efficiently pinpoint essential genes [27]; (4) biological network analysis: studying the dynamics of biological networks, graph-based methods offer powerful tools for understanding complex biological systems [28].

For the sake of completeness, we outline the fundamental concepts of graph theory that may be encountered throughout this survey, along with their respective mathematical formulations as presented in the literature [29–31]:

1. **Undirected graph:** A graph, denoted by G , is typically defined as a tuple (V, E) , where V represents the set of vertices (nodes) and E represents the set of edges, denoting the interactions between nodes. Furthermore, an edge e can be defined as $(u, v) \in E$, which signifies a connection between the two nodes u and v . Therefore, we can state that u and v are neighbors. Notably, nodes can also possess a multiedge connection, wherein two or more edges exist between them with identical endpoints. Multiedge connections hold importance as they indicate that nodes are linked by more than one path, each potentially conveying distinct information.

2. **Directed graph:** A directed graph, denoted by G' , can be defined as (V, A) , where the function f maps each member of A to an ordered pair of nodes in V . These ordered pairs constitute directed edges, also known as arcs or arrows. Moreover, a directed graph exhibits asymmetry, such that if an edge $e = (u, v)$ has a direction from u to v , it does not necessarily imply the existence of an edge with a direction from v to u .
3. **Weighted graph:** A weighted graph, denoted by G'' , is defined as (V, E, ω) , where V and E retain their previous meanings, and ω represents a weight function that assigns a score to each edge, denoted as $\omega(e) \in \mathbb{R}$. Typically, the weight assigned to an edge (u, v) reflects the relevance between nodes u and v , with higher weights signifying greater relevance.

Our preference for the MIT-BIH dataset aligns with its widespread adoption and facilitates methodological comparisons by minimizing the influence of dataset-specific biases. This survey's primary objective is to contribute to the understanding of the interplay between ML/DL and ECG arrhythmia detection by focusing on the following key areas:

- Summarizing graph-based techniques tailored for anomaly detection and classification within the MIT-BIH database.
- Comparing the performance of these techniques on the aforementioned dataset.
- Bridging a gap in the existing literature: while a review addressing techniques utilizing the MIT-BIH database exists [32], it does not prioritize graph-based representations. Therefore, this survey offers a dedicated exploration of graph-based methods for both classification and anomaly detection.
- Delineating the key contributions made by graph-based techniques within the last five years (2019–2023).

Consequently, we anticipate that both researchers and practitioners will find this review instrumental in:

- Evaluating the current landscape of techniques for ECG arrhythmia detection within the MIT-BIH dataset.
- Informing the development of novel methodologies by highlighting the strengths and limitations of existing approaches.
- Providing a benchmark for performance assessment, enabling the objective comparison of newly proposed methods against established standards.

The remainder of this review is structured as follows. Section 2 offers a foundational introduction to the MIT-BIH database, a widely recognized benchmark for evaluating proposed frameworks, and details the approach employed for article selection, ensuring the reviewed studies' relevance to the chosen topic. Section 3 establishes a theoretical foundation by presenting an ML pipeline framework for ECG data analysis. Building upon this framework, Section 4 delves into the core of this review, dividing the selected articles based on their application of graph theory for ECG data analysis. This section categorizes the reviewed studies into anomaly detection and classification tasks, and provides remarks and observations. Finally, Section 5 concludes this survey paper.

2. The MIT-BIH Arrhythmia Dataset

Extensively utilized within the domain of arrhythmia research, the publicly accessible Massachusetts Institute of Technology–Beth Israel Hospital (MIT-BIH) Arrhythmia Database constitutes a comprehensive and continuously expanding repository of well-annotated digital physiological recordings and associated data accessible to the biomedical research community [33]. Comprised of 48 thirty-minute segments of dual-channel, twenty-four-hour ECG recordings, the MIT-BIH Arrhythmia Database serves as a benchmark dataset for arrhythmia-related investigations [1]. Each recording is accompanied by an annotation file (.atr) wherein each heartbeat is categorized according to its specific type. Adhering to the classifications established by the Association for the Advancement of Medical Instrumentation (AAMI), the initial eighteen heartbeat categories within the MIT-BIH arrhythmia data can be consolidated into five primary types: normal (N), supraventricular

ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q) [34]. The specific details of this grouping are delineated in Table 1.

Table 1. Aligning the beats of AAMI standard and beat types in the MIT-BIH database.

AAMI Heartbeat Types	MIT-BIH Heartbeat Types
Nonectopic beats	Normal beat
	Left bundle branch block
	Right bundle branch block
	Nodal (junctional) escape beat
	Atrial escape beat
Supraventricular ectopic beats	Atrial premature beat
	Aberrated atrial premature beat
	Supraventricular premature beat
	Nodal (junctional) premature Beat
Ventricular ectopic beats	Ventricular flutter beat
	Premature ventricular contraction
	Ventricular escape beat
	Start of ventricular flutter fibrillation
Fusion beats	End of ventricular flutter fibrillation
	Fusion of ventricular and normal beats
Unknown beats	Fusion of paced and normal beats
	Paced beat
	Unclassifiable beats

The Procedure for Selecting Articles

Extensive literature retrieval efforts were undertaken to identify articles employing CML/DL for both classification and anomaly detection methodologies within the domain of ECG research [35–39].

A range of prominent scientific search engines, including Scopus and PubMed, were considered. However, Google Scholar, recognized for its robust search capabilities, was ultimately chosen for the comprehensive literature search in this paper. Acknowledging the distinct, yet interrelated, nature of classification and anomaly detection, our search strategy employed targeted queries focused on the specific domain of “graph-based MIT-BIH arrhythmia database”. Separate searches were conducted for classification and anomaly detection within the temporal window of 2019–2023. These initial queries yielded a total of 250 articles (classification: 150; anomaly detection: 100). Given the primary focus of this survey on graph-based techniques evaluated on the MIT-BIH database for both classification and anomaly detection, a further refinement of the search criteria was implemented. This more stringent approach ultimately identified a set of 10 articles for each category, representing the core literature for our analysis of the MIT-BIH database analysis using graph-based techniques.

3. Arrhythmia Association Using Machine Learning

Building on the foundational framework established by Mitchell [40], ML can be conceptually understood through the interplay of three key components:

- **Task (T):** The specific problem or objective the ML model is designed to tackle. In the context of this work, the task (T) would be classifying and identifying various arrhythmias within labeled ECG recordings.

- **Experience (E):** The training data, a collection of labeled examples, serve as the basis for the model's learning and knowledge acquisition. For arrhythmia classification, the experience (E) would comprise a labeled dataset of ECG recordings, with each recording assigned to a specific arrhythmia type.
- **Performance (P):** The effectiveness of the ML model on the designated task, typically measured by metrics such as accuracy, precision, recall, and other relevant evaluation criteria. The ideal model exhibits a strong generalizability, performing accurately on unseen data beyond the training set.

Within this framework, both hand-crafted and DL models for ECG analysis are constructed. The model ingests an input ECG record from the experience (E) and maps it to an output label representing the detected arrhythmia type. Optimization algorithms refine the model's internal parameters based on its performance on the training data, aiming to achieve optimal accuracy and generalizability. Ultimately, the ideal ML model demonstrates the ability to accurately classify unseen ECG recordings, potentially assisting cardiologists in diagnostic and treatment decisions.

Cross-validation (CV) serves as a widely employed technique for evaluating the performance of both CML and DL models, providing statistically rigorous results. This method assesses the efficacy of a classifier system by partitioning the dataset into training and testing subsets. The testing data remain hidden during the training process, ensuring an unbiased evaluation. The dataset is divided into K folds, with $K - 1$ folds utilized for training and the remaining fold used for testing. This process is iterated K times, each time employing a different fold for testing. In clinical settings, researchers often investigate the statistical associations between symptoms (represented by test samples) and the presence of disease. Identifying significant associations necessitates expressing data in clinically meaningful ways. To evaluate the performance of different classifiers within each fold, several common metrics [41] are employed:

- **Accuracy:** the proportion of correctly classified samples.
- **Sensitivity:** the ability of the model to correctly identify true positive cases (i.e., identifying diseased patients who truly have the disease).
- **Specificity:** the ability of the model to correctly identify true negative cases (i.e., identifying healthy patients who truly do not have the disease).
- **Area Under the Curve (AUC):** A graphical plot of the model's performance, showing the relationship between true positive rate (sensitivity) and false positive rate ($1 - \text{specificity}$).

By employing these metrics in conjunction with CV, researchers gain insights into the generalizability and clinical relevance of their proposed models within the context of ECG data analysis.

Extracting informative features from ECG signals represents a crucial step in constructing robust classification models for CML-based approaches. This process involves identifying and quantifying relevant characteristics that discriminate between different arrhythmias. Common examples of such features include:

- **Morphological and positional features:** These features capture the shape characteristics of the ECG signal. Examples include the amplitudes and widths of peaks and valleys (e.g., R and P waves), interwave distances (e.g., R-R interval), and other relevant shape descriptors [7,42,43].
- **Spectral methods:** This category encompasses frequency-domain representations of the ECG signal obtained through transformation techniques. A prominent example is the wavelet transform. This technique decomposes the signal into frequency subbands, enabling the analysis of its underlying components at different scales and orientations. Recent research has demonstrated the efficacy of wavelet-based features in ECG classification [7,44,45]. Furthermore, hybrid approaches combining DL with wavelet transforms have emerged to leverage the strengths of both methods [17,19,46].

These extracted feature vectors serve as the basis for the subsequent analysis of the ECG signals within CML-based approaches. In contrast, the defining advantage of DL resides in its inherent ability to directly learn features from the raw input ECG data end-to-end [16], effectively bypassing the dedicated feature extraction step. Essentially, the construction of a CAD system employing ML algorithms for distinguishing normal and abnormal samples involves the following core stages:

- Stage 1: data acquisition: relevant ECG-based arrhythmia datasets are procured.
- Stage 2: preprocessing: the data undergo a series of preprocessing steps, including:
 - Denoising to remove unwanted noise artifacts.
 - Peak detection to identify key signal components.
 - Signal segmentation to partition the data into meaningful segments.
- Feature engineering (CML)/feature learning (DL): For CML models, extracting and selecting informative features from the preprocessed data. Conversely, DL approaches learn features directly from the raw input data during model training.
- Model training and evaluation: the application of the chosen classification algorithm (CML or DL) to the prepared data, followed by a rigorous evaluation to assess its performance and generalizability.

The readily available MIT-BIH dataset completes stage 1 (data acquisition) within the construction process, allowing practitioners to seamlessly transition to stage 2 (preprocessing). This stage focuses on preparing the ECG data for subsequent analysis and model training. A common preprocessing step, employed by both CML and DL approaches, involves denoising the ECG signal to remove unwanted artifacts [45,47–49]. Denoising aims to mitigate or eliminate the distorting influence of artifacts, which can originate from diverse sources such as respiration, body movements, electrode contact issues, and skin-electrode impedance. This purification step enhances the overall quality of the ECG signal, thereby facilitating the extraction of its inherent and pertinent characteristics. Moreover, the availability of R-peak annotations within the MIT-BIH database offers a robust ground truth for segmenting ECG signals. These readily available annotations enable the delineation of individual cardiac beats [50], facilitating further analysis and model development based on segmented data. Consequently, the processed signals lead to building an effective and accurate model in the subsequent anomaly detection/classification model.

Following the preprocessing stage, feature extraction commences, aiming to identify a robust and informative set of descriptors from the preprocessed ECG signal. The literature [51,52] presents a diverse array of feature extraction techniques for differentiating normal and abnormal ECG signals. Texture-based approaches (e.g., local binary pattern, structural co-occurrence matrix), morphological-based methods, and wavelet-transform-based algorithms are utilized for ECG data analysis. Each approach leverages different discriminative properties of the ECG signal to generate a set of features suitable for the subsequent analysis and model training.

Beyond approaches based on texture analysis or wavelet transforms, the literature explores alternative feature extraction methods that deviate conceptually. One such example is the utilization of visual-perception-inspired features proposed by Anand et al. [53]. These features aim to emulate the human visual system's ability to discern patterns within the ECG signal. Following the successful extraction of robust and discriminative features through hand-engineering, the choice of the specific CML technique (e.g., SVM or MLP) ought to have a minimal impact on the accuracy and efficiency of ECG anomaly detection/classification. This implies that a well-constructed feature set can mitigate the influence of the chosen CML algorithm on the overall performance of the model.

In contrast to CML-based approaches, the alternative pipeline leverages DL, which has emerged as the prevailing paradigm for tackling machine vision tasks [54]. CNNs constitute a prevalent DL architecture frequently employed for ECG data analysis, e.g., in [3–5,8,16–20,55,56]. Unlike CML methods that rely on hand-engineered features, CNNs directly process raw ECG data as input. Within these architectures, a series of operations,

such as convolution, pooling, and batch normalization, play a pivotal role in extracting discriminative features and ultimately contribute to the robustness of the model. The learned feature maps are then fed to a final classification layer, typically employing a softmax activation function. However, certain studies advocate for utilizing DL models as feature extractors instead of end-to-end classifiers [57]. In such cases, the extracted feature vectors serve as input to subsequent CML models.

DL models exhibit inherent vulnerabilities when trained on limited datasets, particularly when initializing training from scratch [58]. This vulnerability is especially pertinent within the context of specific arrhythmia classes within the MIT-BIH database, where the availability of well-annotated samples is scarce. Consequently, DL models in such scenarios are susceptible to generalization issues and overfitting. Overfitting implies that the model memorizes the intricacies of the training data, hindering its ability to generalize and accurately classify unseen samples. To address this challenge, numerous studies propose various techniques for mitigating the impact of imbalanced datasets within the context of ECG data analysis [59–61]. Several studies, such as those by Rai et al. [59] and Shoughi et al. [61], propose the application of synthetic minority oversampling techniques (SMOTEs) to address the issue of imbalanced datasets in ECG data. SMOTE aims to augment the under-represented classes within the training data by generating synthetic samples that share characteristics with the existing minority samples. This approach effectively balances the class distribution, mitigating the susceptibility of DL models to overfitting and ultimately strengthening their generalizability. By reducing the bias towards the majority class, SMOTE enables the model to learn a more comprehensive representation of the underlying data distribution, leading to improved performance on unseen examples.

4. Literature Review: A Graph-Centric Exploration of the MIT-BIH Database

ECG signals extracted from the MIT-BIH database, encompassing recordings from 47 individuals, constitute a pivotal benchmark for evaluating algorithms designed to detect cardiac arrhythmias. Building CAD systems that accurately analyze these irregular rhythms plays a crucial role in facilitating timely diagnoses and interventions, with the potential to alleviate the burden of cardiovascular disease.

Graph-based approaches have emerged as a potent paradigm for a deeper examination into the intricacies of ECG data. Unlike traditional feature-based methods, these approaches leverage network representations to capture the intricate interplay and temporal dynamics inherent in arrhythmias, transcending the limitations of an isolated feature analysis.

This literature review embarks on a multifaceted exploration of the current landscape of research utilizing graph-based approaches within the context of the MIT-BIH database. We delve into the domain of anomaly detection, where algorithms strive to identify deviations from the expected rhythm, potentially uncovering hidden pathologies. Subsequently, we navigate the realm of classification, where the focus shifts towards accurately categorizing arrhythmia types.

Through this multifaceted lens, we aim to illuminate the potential of graph-based approaches in unraveling the complexities of arrhythmias within the MIT-BIH database. This exploration not only equips researchers with the tools to develop more accurate and robust arrhythmia detection algorithms but also contributes to the advancement of the field by highlighting the hidden potential of less-explored graph-based methods.

4.1. Graph-Based Techniques: Anomaly Detection in MIT-BIH Database

4.1.1. Anomaly Detection Landscape

The quest for detecting anomalous patterns in data permeates numerous disciplines, as evidenced by extensive research in various domains [62,63]. Within the lens of ML, anomaly detection methods can be broadly categorized as unsupervised, semi-supervised, or supervised approaches. Yet, when considering time series data, like ECG signals, further nuances emerge, prompting the exploration of distance-based, density-based, and forecasting-based techniques.

The inherent nature of time series data presents unique challenges during the development of CML and DL techniques for ECG analysis. A multitude of influential factors contribute to the construction of such automated frameworks, rendering direct comparisons between studies intricate. Notably, these factors encompass the selection of ECG data and specific signals, applied preprocessing techniques, and the chosen data split for model training. Consequently, objectively assessing and comparing the performance of different methods becomes a formidable task, as pinpointing the precise element(s) responsible for any performance gains proves elusive. Furthermore, certain challenges specific to time series data pose additional complexities, necessitating further exploration (discussed subsequently).

4.1.2. Challenges in Subsequence Anomaly Detection

Many existing anomaly detection approaches are plagued by limitations that hinder their potential. Many require a visionary understanding of the anomaly's length and frequency. Others are confined to detecting local echoes, thus missing recurring patterns. Some studies depend on prior domain knowledge for designing anomaly discovery algorithms or cumbersome implementations, proving inefficient in the face of recurrent anomalies of the same type.

In the burgeoning era of big data, accurate and efficient anomaly detection in multivariate time series data assumes a paramount importance. However, achieving this goal presents a complex puzzle: balancing fast model inference for real-time analysis, navigating unlabeled datasets for unsupervised learning, and effectively handling excessively long time series. Overcoming these hurdles requires innovative paradigms that bridge the gap between accuracy, speed, and scalability.

These limitations necessitate a novel movement in anomaly detection, one that embraces the inherent complexities of time series data and facilitates robust anomaly detection, paving the way for diverse solutions of real-world applications.

4.1.3. Anomaly Detection in Biomedicine: A Critical Precursor

Accurately identifying anomalies in time series data, particularly in biomedical modalities like ECGs, plays a pivotal role in both data preprocessing and postprocessing [64]. Unearthing these deviations from normalcy often signifies the presence of underlying disorders, demanding prompt identification and intervention. Therefore, robust anomaly detection methods serve as crucial tools for enhancing the efficacy of biomedical data analysis, ultimately leading to improved diagnosis and personalized patient care. Thus, this section examines five specific realms of graph-based methods, highlighting their unique capabilities in uncovering hidden anomalies nestled within complex data structures.

1. **Boniol and Palpanas [65,66]: unveiling hidden anomalies through a graph's lens: introducing the Series2Graph (s2g) approach.**

Boniol and Palpanas illuminate a novel path for unsupervised subsequence anomaly detection with their Series2Graph (S2G) technique. S2G bypasses the requirement for labeled instances or anomaly-free data, offering domain-agnostic flexibility and adaptability to anomalies of varying lengths. At the heart of S2G lies a unique graph-based representation of time series subsequences. It masterfully unfolds in three interconnected steps: (1) **Embedding subsequences in shape-preserving space:** S2G embeds subsequences into a vector space, delicately preserving their essential shapes, paving the way for subsequent pattern discovery. (2) **Unraveling recurrent patterns through overlapping trajectories:** within this shape-centric space, S2G identifies overlapping trajectories, revealing recurrent patterns embedded within the data that serve as subtle markers of normalcy. (3) **Constructing a graph of normality:** S2G builds a graph where nodes embody these overlapping trajectories, and edges represent transitions between subsequences observed in the original series. This graph elegantly encodes both recurring patterns and their interrelationships, serving as a blueprint of normality.

The meticulously constructed graph empowers S2G to discern anomalies—subsequences that stray from the well-trodden paths of normalcy. These deviations, manifested as infrequent or absent patterns within the graph, stand exposed, revealing their anomalous nature. When tested on the MIT-BIH Supraventricular Arrhythmia Database (MBA), S2G demonstrated its prowess, achieving top-k accuracies ranging from 20% to 100%. Notably, its performance peaked when the input length exceeded the expected anomaly length, showcasing its adaptability to diverse anomaly patterns.

S2G's remarkable capabilities, unburdened by the need for labeled data or prior domain knowledge, herald a promising advancement in unsupervised anomaly detection across a spectrum of domains. Its potential to unveil hidden anomalies within complex time series data, including biomedical signals like ECGs, holds significant promise for the early detection of health abnormalities and improved clinical decision-making.

2. **Schneider et al. [67]: unveiling hidden anomalies at scale: dads takes series2graph to new heights.**

Schneider et al. push the boundaries of anomaly detection with a distributed anomaly detection system (DADS), an innovative system that catapults the effectiveness of S2G to new heights. S2G, an unsupervised anomaly detection method, excels at pinpointing hidden dissonances within time series, regardless of their length or recurring nature. However, its single-threaded architecture limits its ability to grapple with truly massive datasets. Enter DADS, which is built upon the foundations of S2G, empowered by the principles of the actor programming model.

DADS engineers a distributed processing framework, seamlessly dividing the data, intermediate states, and computations across multiple processors within a cluster. This minimizes communication overhead and synchronization barriers. DADS outpaces S2G by orders of magnitude, exhibiting near-perfect linear scaling with the number of processors employed. This efficiency opens doors to analyzing much larger sequences, unfurling secrets within big data.

Schneider et al.'s DADS work transcends the limitations of S2G, not only in terms of speed but also in its scalability to tackle unprecedented data volumes. This opens doors to exciting possibilities in diverse fields not only in ECG anomaly detection, e.g., monitoring complex systems to detect financial fraud [68].

3. **Ma et al. [69]: unveiling hidden patterns in multivariate time series: a deep learning Bi-Transformer engineers unsupervised anomaly detection.**

Ma et al. propose an approach in the realm of anomaly detection with an unsupervised Bi-Transformer anomaly detection method (BTAD), an unsupervised DL method that navigates the complexities of multivariate time series data. At the essence of BTAD lies a crafted Bi-Transformer architecture, two parallel dimensions that extract and analyze features. The Bi-Transformer's prowess stems from its adaptive multihead attention mechanism, which attunes to the nuances of each dimension within the multivariate data, capturing their unique patterns. The researchers amplify BTAD's versatility through an ensemble of auxiliary techniques: (1) **An alternating update strategy:** A generative adversarial training framework ensuring the model is sharpening its focus on anomalies and minimizing the distractions of false positives. As such, magnifying the anomalous patterns allows the model to identify even the faintest deviations from normalcy. (2) **A dataset division method:** Inspired by model-agnostic metalearning (MAML), which transcends mere universality, it empowers the model to rapidly grasp the nuances of novel anomaly types and perform efficient detection with limited data, making it a versatile anomaly detector for uncharted datasets. MAML empowers BTAD to generalize across diverse multivariate time series, extending its reach to a broader spectrum of applications. (3) **A modified decoder structure:** this module disentangles itself from direct input inference, instead harnessing its knowledge of the latent space's intricate structure to produce faithful reconstructions, even within the complexities of high-dimensional datasets.

Experiments on the MBA dataset showcase BTAD's virtuosity, achieving a precision of 0.9548, recall of 0.9999, AUC of 0.9879, and F1-score of 0.9769

4. **Zarei et al. [70]: GraphTS weaves a new path for subsequence anomaly detection.** Zarei et al. unveil graph-based time series (GraphTS), a technique that combines graph theory and visual representation to capture hidden anomalies within time series data. GraphTS's approach consists of the following key steps: (1) **Visualizing time series:** GraphTS commences with a mesmerizing 2D visualization technique, 2Dviz, which projects the time series onto a spatial-temporal plane. This arranges subsequence patterns into a high-quality visualization that potentially allows an improved detection of anomalies. (2) **Time series into a graph:** Inspired by this visual representation, GraphTS constructs a graph where nodes embody subsequence patterns, and edges chronicle the frequency of their successive encounters in the original time series. This graph captures both the harmonious normal and anomalous patterns. (3) **Unmasking anomalies through weighted paths:** GraphTS reveals anomalies through their distinct pathways within the graph. Normal patterns move along paths with high-weighted edges, while anomalies tread upon paths of lesser weight, enabling their identification with clarity and precision. The GraphTS approach's advantages resonate with the following: (1) GraphTS embraces both recurrent and anomalous patterns, unveiling the full spectrum of anomalies within the time series, unlike methods that focus solely on individual notes; (2) it constructs its graph without prior knowledge of anomaly length, effortlessly detecting anomaly patterns of any duration; (3) GraphTS transforms raw time series data into a graph, rendering anomaly detection as effortless as tracing paths of diminished weight. Finally, experiments on the MBA dataset showcase GraphTS's virtuosity, surpassing the renowned Series2Graph algorithm in terms of top-k accuracy.
5. **Liu et al. [71]: a topological technique for ECG anomaly detection.** Liu et al. propose a fusion of topological data analysis (TDA) and nonlinear feature extraction to unveil hidden patterns within the intricate rhythms of ECG signals, paving the way for robust anomaly detection and personalized healthcare. The proposed approach consists of the following key steps: (1) **Mapping the ECG's topological space:** They embark on transforming the ECG time series into a topological space, akin to an intricate map of its dynamic landscape. This is achieved through time-delay embedding, a technique that unfolds the ECG's rhythmic patterns to obtain the ECG point cloud. (2) **Constructing a topological imprint:** Within the topological space, they assemble a point cloud representing the ECG data. Persistent homology, a powerful tool for unraveling relationships between data points within complex structures, is then employed to forge a topological imprint—a fingerprint capturing the essence of the ECG's intricate dynamics. (3) **Extracting persistent landscapes:** From this imprint, they extract the persistence landscape, a mathematical model that unveils the persistent topological features of the ECG signal. These features reveal both the heart's regular rhythms and its discordant anomalies.

Using the Physionet MIT-BIH dataset categorized according to AAMI standards, the authors trained their model with only 20% of the total data. Yet, it achieved accuracies of 100% for normal heartbeats, 98.75% for ventricular beats, 95.88% for supraventricular beats, and 91.97% for fusion beats. The model's remarkable performance with limited training data suggests its potential for addressing the issue of data scarcity in the field of ECG data analysis.

4.2. Graph-Based Techniques: Classification of the MIT-BIH Database

Within the rapidly evolving landscape of computational intelligence, a plethora of novel models have emerged for the classification and interpretation of arrhythmias and other cardiovascular pathologies via ECG signals [6,72]. Among these, feature extraction methods and DL architectures leveraging graph-based representations exhibit noteworthy promise. Their potential lies in the ability to effectively characterize the underlying disease

through the successful extraction of informative numerical features intrinsically intertwined with the pathological process. Furthermore, the robust capture of disease-specific pathological properties within a graph-based framework can empower CML and DL algorithms to achieve superior accuracy in disease prediction.

The present work focuses on specific methodologies for arrhythmia classification leveraging the MIT-BIH ECG dataset. The extensive adoption of this dataset facilitates a robust and comprehensive analysis of diverse graph-based models and classifiers. This dataset selection minimizes the influence of dataset-specific biases, enabling a more generalizable evaluation of the proposed techniques. However, a persistent challenge lies in the variability of classification tasks across studies, as evidenced by discrepancies in the number of classes employed by different authors. This heterogeneity necessitates careful consideration when comparing and drawing conclusions from the existing literature.

This section embarks on the exploration of five graph-based techniques employed for classification tasks within the domain of ECG analysis. These studies showcase the diverse applications and remarkable successes of these methods in overcoming various challenges inherent to ECG data. They encompass a spectrum of tasks, from the precise identification of peaks to the accurate classification of arrhythmias, thus exemplifying the versatility and immense potential of graph-based techniques in unraveling the hidden patterns within ECG signals. Each study is individually scrutinized, with a focus on its unique methodology, salient contributions, and achieved results within the context of the MIT-BIH Arrhythmia Database.

- 1. Fotoohinasab et al. [73]: R-peak detection with knowledge-guided graph constraints**
The first study focuses on the fundamental task of R-peak detection, which forms the bedrock of ECG analysis. Fotoohinasab et al. propose utilizing a graph-constrained change-point detection (GCCD) model. By reframing fiducial point delineation as a change-point detection challenge, the GCCD model exploits the sparsity of these changes to efficiently locate important markers within the fluctuating ECG signal. By capitalizing on the inherent sparsity of change points, the proposed model efficiently identifies abrupt transitions within the ECG signal, eliminating the need for any preprocessing steps in R-peak detection. Furthermore, this model leverages the sparsity of change points within the ECG signal and incorporates prior biological knowledge through constraint graphs. The proposed approach initializes with a simple hand-crafted constraint graph, followed by a novel graph learning algorithm that iteratively optimizes the graph structure via a greedy search. This optimization maximizes R-peak detection accuracy, resulting in a constraint graph tailored for optimal performance. The authors analyze the trade-off between manually defined and automatically learned constraint graphs by comparing their structural differences and R-peak detection accuracy. Utilizing the MIT-BIH Arrhythmia Database for evaluation, the model achieved outstanding performance, reaching a 99.64% sensitivity, 99.71% positive predictivity, and 0.19% error rate with the manual graph, and comparable results with the learned graph (99.76% sensitivity, 99.68% positive predictivity, 0.55% error rate).
- 2. Subasi et al. [74]: tower graph transformation for high-fidelity classification**
Subasi et al. (2023) introduce a tool for ECG signal classification, the “tower graph transformation”. This approach leverages a unique graph structure to generate signals enriched with essential features. Employing minimum, maximum, and average pooling techniques, the tower graph transforms the raw ECG signal into a multilayered representation, capturing both local and global variations. To further refine information extraction, Subasi et al. propose a “one-dimensional hexadecimal adaptive pattern” that efficiently identifies informative features within the transformed signals. This is followed by a rigorous feature selection process utilizing the “ReliefF and iterative Neighborhood Component Analysis (RFINCA)”, ensuring only the most discriminative features are presented to the classifier.

Before feeding the data to classifiers, each ECG signal undergoes a multistep feature extraction process. First, the tower graph transformation extracts diverse local and global information through pooling techniques. This generates a richer representation within each node of the graph. Subsequently, the one-dimensional hexadecimal adaptive pattern efficiently uncovers 1536 features per node, leading to a comprehensive pool of 15,360 candidate features. Finally, employing the RFINCA selection approach, the model identifies the 142 most discriminative features.

The authors demonstrated the performance of their method by achieving remarkable classification accuracy, reaching 95.70% and 97.10% with artificial neural networks and deep neural networks, respectively.

3. **Jiang et al. [75]: unveiling the multilabel dependencies of ECGs with graph-powered deep learning**

Jiang et al. craft a DL architecture that embraces the intricate reality of multiple concurrent cardiac conditions within 12-lead ECGs. Their model transcends conventional approaches by integrating various modules to achieve this effort. (1) **Residual blocks**: these robust units enhance information flow within the network, preserving crucial details for accurate classification. (2) **Bidirectional gated recurrent unit (Bi-GRU)**: this powerful tool captures the sequential nature of ECG signals, ensuring a context-aware analysis of the dynamic cardiac landscape. (3) **Graph convolutional network (GCN)**: This component considers the inherent interdependent relationships between different cardiac diseases, allowing the model to decipher the intricate interplay of coexisting conditions. This module is trained to exploit the authors' custom-designed class-aware binary cross-entropy loss function.

The Jiang et al. model achieved an F1 score of 0.603 (i.e., in the context of a fivefold cross-validation scheme) in the competitive PhysioNet/Computing in Cardiology Challenge 2020.

4. **Kobat et al. [76]: a 3D prismatic pattern for intelligent ECG analysis**

Kobat et al. present an approach for arrhythmia detection, empowering intelligent assistants with the ability to interpret ECG signals. Their work leverages a novel 3D prismatic pattern, a unique graph-based representation that captures the intricate textures within these signals. Building upon a dataset of 1000 diverse ECG signals with 17 labels, the proposed architecture integrates several key components: (1) **A prismatic pattern**: this crafted 3D shape acts as a powerful feature extractor, exploring the subtle nuances of ECG signals and generating rich textural representations. (2) **A tunable Q wavelet transform**: operating at both low and high frequencies, this transform effectively captures the diverse temporal dynamics within the ECG data from 53 sub-bands. (3) **A statistical feature extractor**: this module refines the 53 sub-bands information, ensuring the computation of measurements at both low and high frequencies. (4) **Neighborhood component analysis (NCA)**: this dimensionality reduction technique carefully selects the most informative features for accurate classification. The model's performance achieved a remarkable accuracy of 97.30% using an SVM classifier with a 10-fold cross-validation scheme.

5. **He et al. [9]: a multilevel approach to conquer data variations in ECG classification**

He et al. address the limitations of traditional DL approaches in ECG classification, where subject-specific differences hinder generalizability. Their solution, the multilevel unsupervised domain adaptation framework (MLUDAF), overcomes this obstacle, enabling arrhythmia detection across diverse individuals. MLUDAF searches through the ECG data at two levels: (1) **spatio-temporal feature extraction**: leveraging the atrous spatial pyramid pooling residual (ASPP-R) module, the model captures the subtle nuances of each signal over time and space; (2) **data structure extraction**: a GCN module is then utilized to unlock the inherent relationships between different data points, enriching the feature representation with crucial structural information. However, He et al. do not stop there. To bridge the gap between subjects and achieve robust performance, they implement a three-pronged alignment strategy:

(1) **domain alignment**, minimizing discrepancies between source and target domains in the overall feature distribution; (2) **semantic alignment**, ensuring the extracted features retain relevant clinical meaning across data variations; (3) **structure alignment**, aligning the underlying data structures of both the source and target domains, further stabilizing the classification process. By integrating these alignment mechanisms, MLUDAF empowers the feature extractor to learn representations that are both domain-agnostic and semantically relevant, effectively reducing subject-specific biases. When tested on the MIT-BIH database, MLUDAF achieved an overall accuracy of 96.8% for arrhythmia detection.

4.3. Remarks

This survey provided an overview of CAD techniques leveraging graph-based representations for ECG data analysis. Conventional-based approaches have established a dominant role in differentiating various pathological conditions within ECG data. Moreover, both CML and DL methodologies have demonstrated remarkable accuracy in anomaly detection and classification of ECG data. Despite the recent surge in interest towards DL, its application in medical diagnosis remains subject to critical scrutiny [77]. This cautious stance is partly driven by the absence of standardized protocols encompassing data acquisition, as well as training and testing of models. Furthermore, the inherent data-hungry nature of DL models presents a significant challenge, as they require an abundance of annotated samples to effectively learn the patterns present in medical data. Such comprehensive annotations, crucial for supervised learning, are often time-consuming and laborious to generate. Consequently, the limited availability of large and representative ECG datasets hinders the widespread adoption and advancement of DL in this domain, posing difficulties in building generalizable models capable of accurate prediction on unseen data. In light of this data scarcity, alternative approaches offer promising avenues to overcome these limitations. These include employing shallow DL architectures, fine-tuning pretrained models [56] rather than building a model from scratch or adopting semi-supervised [8] or unsupervised learning techniques [9,36] for ECG data analysis.

5. Conclusions and Future Studies

Leveraging the widespread adoption and standardized format of the MIT-BIH ECG database, this survey has revealed a plethora of research endeavors for ECG arrhythmia analysis. Despite this abundance, the utilization of graph-based methodologies within the broader domain of ECG data analysis remains significantly underappreciated. Furthermore, identifying a definitive victor in the pursuit of flawless anomaly detection or classification performance proves elusive. This challenge stems from the frequent neglect of several crucial variables across diverse studies. Disparities in the split of training and testing data can significantly impact model generalizability, potentially hindering performance when applied to unseen datasets. While the prospect of training models on smaller datasets (i.e., a subset of the total dataset) holds undeniable appeal—particularly in the context of DL—ensuring the utilization of representative and challenging samples during both training and testing remains paramount for the development of robust and accurate classifiers. Our future work endeavors to expand upon this survey by incorporating a wider range of existing research, including older studies, to facilitate a comprehensive comparison and contrast of trends in feature extraction and deep learning model application within the context of ECG analysis.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Moody, G.B.; Mark, R.G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [CrossRef] [PubMed]
2. Merbouti, M.A.; Cherifi, D. Machine learning based electrocardiogram peaks analyzer for Wolff-Parkinson-White syndrome. *Biomed. Signal Process. Control* **2023**, *86*, 105302. [CrossRef]
3. Liu, Q.; Feng, Y.; Xu, H.; Li, J.; Lin, Z.; Li, S.; Qiu, S.; WU, X.; Ma, Y. PSC-Net: Integration of Convolutional Neural Networks and Transformers for Physiological Signal Classification. SSRN 4524798. Available online: <https://ouci.dntb.gov.ua/en/works/4rDDWYM9/> (accessed on 22 January 2024).
4. Shukla, N.; Pandey, A.; Shukla, A.P.; Neupane, S.C. ECG-ViT: A transformer-based ECG classifier for energy-constraint wearable devices. *J. Sens.* **2022**, *2022*, 2449956. [CrossRef]
5. Al Nazi, Z.; Biswas, A.; Rayhan, M.A.; Abir, T.A. Classification of ECG signals by dot residual LSTM network with data augmentation for anomaly detection. In Proceedings of the 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 18–20 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.
6. Clifford, G.D.; Azuaje, F.; McSharry, P. *Advanced Methods and Tools for ECG Data Analysis*; Artech House Boston: Norwood, MA, USA, 2006; Volume 10.
7. Zadeh, A.E.; Khazaei, A.; Ranaee, V. Classification of the electrocardiogram signals using supervised classifiers and efficient features. *Comput. Methods Programs Biomed.* **2010**, *99*, 179–194. [CrossRef]
8. Mohebbian, M.R.; Marateb, H.R.; Wahid, K.A. Semi-supervised active transfer learning for fetal ECG arrhythmia detection. *Comput. Methods Programs Biomed. Update* **2023**, *3*, 100096. [CrossRef]
9. He, Z.; Chen, Y.; Yuan, S.; Zhao, J.; Yuan, Z.; Polat, K.; Alhudhaif, A.; Alenezi, F.; Hamid, A. A novel unsupervised domain adaptation framework based on graph convolutional network and multi-level feature alignment for inter-subject ECG classification. *Expert Syst. Appl.* **2023**, *221*, 119711. [CrossRef]
10. Haykin, S.; Network, N. A comprehensive foundation. *Neural Netw.* **2004**, *2*, 41.
11. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: New York, NY, USA, 2013.
12. Krishna, K.; Murty, N.M. Genetic K-means algorithm. *IEEE Trans. Syst. Man -Cybern.-Part B Cybern.* **1999**, *29*, 433–439. [CrossRef]
13. Jolliffe, I. *Principal Component Analysis*; Springer: Berlin/Heidelberg, Germany, 2011.
14. Alinsaif, S.; Lang, J. Texture features in the Shearlet domain for histopathological image classification. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 312. [CrossRef]
15. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
16. Park, J.; Lee, K.; Park, N.; You, S.C.; Ko, J. Self-Attention LSTM-FCN model for arrhythmia classification and uncertainty assessment. *Artif. Intell. Med.* **2023**, *142*, 102570. [CrossRef]
17. Wang, B.; Chen, G.; Rong, L.; Liu, Y.; Yu, A.; He, X.; Wen, T.; Zhang, Y.; Hu, B. Arrhythmia Disease Diagnosis Based on ECG Time-Frequency Domain Fusion and Convolutional Neural Network. *IEEE J. Transl. Eng. Health Med.* **2022**, *11*, 116–125. [CrossRef]
18. He, X.; Shan, W.; Zhang, R.; Heidari, A.A.; Chen, H.; Zhang, Y. Improved Colony Predation Algorithm Optimized Convolutional Neural Networks for Electrocardiogram Signal Classification. *Biomimetics* **2023**, *8*, 268. [CrossRef]
19. Asif, M.S.; Faisal, M.S.; Dar, M.N.; Hamdi, M.; Elmannai, H.; Rizwan, A.; Abbas, M. Hybrid Deep Learning and Discrete Wavelet Transform-Based ECG Biometric Recognition for Arrhythmic Patients and Healthy Controls. *Sensors* **2023**, *23*, 4635. [CrossRef]
20. Liu, S.; Zhou, B.; Ding, Q.; Hooi, B.; Zhang, Z.; Shen, H.; Cheng, X. Time series anomaly detection with adversarial reconstruction networks. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 4293–4306. [CrossRef]
21. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 3320–3328.
22. Mohebbanaaz, Kumar, L.R.; Sai, Y.P. A new transfer learning approach to detect cardiac arrhythmia from ECG signals. *Signal Image Video Process.* **2022**, *16*, 1945–1953. [CrossRef]
23. Diker, A.; Engin, A. Feature extraction of ECG signal by using deep feature. In Proceedings of the 2019 7th International Symposium on Digital Forensics and Security (ISDFS), Barcelos, Portugal, 10–12 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
24. Hong, S.; Zhou, Y.; Wu, M.; Shang, J.; Wang, Q.; Li, H.; Xie, J. Combining deep neural networks and engineered features for cardiac arrhythmia detection from ECG recordings. *Physiol. Meas.* **2019**, *40*, 054009. [CrossRef]
25. Supriya, S.; Siuly, S.; Wang, H.; Zhang, Y. Epilepsy detection from EEG using complex network techniques: A review. *IEEE Rev. Biomed. Eng.* **2021**, *16*, 292–306. [CrossRef]
26. Ismail, L.E.; Karwowski, W. A graph theory-based modeling of functional brain connectivity based on EEG: A systematic review in the context of neuroergonomics. *IEEE Access* **2020**, *8*, 155103–155135. [CrossRef]
27. Alinsaif, S. Leveraging Random Forest and Graph-based Centralities to Predict Yeast Essential Genes. In Proceedings of the 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Istanbul, Turkiye, 5–8 December 2023; IEEE Computer Society: Piscataway, NJ, USA, 2023; pp. 3446–3452. [CrossRef]
28. Erciyes, K. Graph-Theoretical Analysis of Biological Networks: A Survey. *Computation* **2023**, *11*, 188. [CrossRef]
29. Junker, B.H.; Schreiber, F. *Analysis of Biological Networks*; Wiley Online Library: Hoboken, NJ, USA, 2008; Volume 2.
30. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. *Introduction to Algorithms*; MIT Press: Cambridge, MA, USA, 2022.

31. Huber, W.; Carey, V.J.; Long, L.; Falcon, S.; Gentleman, R. Graphs in molecular biology. *BMC Bioinform.* **2007**, *8*, S8. [[CrossRef](#)]
32. Apandi, Z.F.M.; Ikeura, R.; Hayakawa, S. Arrhythmia detection using MIT-BIH dataset: A review. In Proceedings of the 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA), Kuching, Malaysia, 15–17 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.
33. Goldberger, A.L.; Amaral, L.A.; 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: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [[CrossRef](#)]
34. Yang, H.; Wei, Z. Arrhythmia recognition and classification using combined parametric and visual pattern features of ECG morphology. *IEEE Access* **2020**, *8*, 47103–47117. [[CrossRef](#)]
35. Pereira, T.M.; Conceição, R.C.; Sencadas, V.; Sebastião, R. Biometric recognition: A systematic review on electrocardiogram data acquisition methods. *Sensors* **2023**, *23*, 1507. [[CrossRef](#)]
36. Nezamabadi, K.; Sardaripour, N.; Haghi, B.; Forouzanfar, M. Unsupervised ECG analysis: A review. *IEEE Rev. Biomed. Eng.* **2022**, *16*, 208–224. [[CrossRef](#)] [[PubMed](#)]
37. Deepika, S.; Jaisankar, N. Review on Machine Learning and Deep Learning-based Heart Disease Classification and Prediction. *Open Biomed. Eng. J.* **2023**, *17*, e187412072301060.
38. Annam, J.R.; Kalyanapu, S.; Ch, S.; Somala, J.; Raju, S.B. Classification of ECG heartbeat arrhythmia: A review. *Procedia Comput. Sci.* **2020**, *171*, 679–688. [[CrossRef](#)]
39. Ebrahimi, Z.; Loni, M.; Daneshlab, M.; Gharehbaghi, A. A review on deep learning methods for ECG arrhythmia classification. *Expert Syst. Appl.* **2020**, *7*, 100033. [[CrossRef](#)]
40. Mitchell, T.M. *Machine Learning*; McGraw Hill: Burr Ridge, IL, USA, 1997; Volume 45, pp. 870–877.
41. Alinsaif, S.; Lang, J. 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer’s disease based on MRI data. *Comput. Biol. Med.* **2021**, *138*, 104879. [[CrossRef](#)]
42. Lee, J.; Nam, Y.; McManus, D.D.; Chon, K.H. Time-varying coherence function for atrial fibrillation detection. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2783–2793.
43. Oresko, J.J.; Jin, Z.; Cheng, J.; Huang, S.; Sun, Y.; Duschl, H.; Cheng, A.C. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 734–740. [[CrossRef](#)]
44. Alqudah, A.M.; Albadarneh, A.; Abu-Qasmieh, I.; Alquran, H. Developing of robust and high accurate ECG beat classification by combining Gaussian mixtures and wavelets features. *Australas. Phys. Eng. Sci. Med.* **2019**, *42*, 149–157. [[CrossRef](#)] [[PubMed](#)]
45. Yu, K.; Feng, L.; Chen, Y.; Wu, M.; Zhang, Y.; Zhu, P.; Chen, W.; Wu, Q.; Hao, J. Accurate wavelet thresholding method for ECG signals. *Comput. Biol. Med.* **2023**, *169*, 107835. [[CrossRef](#)] [[PubMed](#)]
46. Yildirim, Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput. Biol. Med.* **2018**, *96*, 189–202. [[CrossRef](#)] [[PubMed](#)]
47. Cuomo, S.; De Pietro, G.; Farina, R.; Galletti, A.; Sannino, G. A revised scheme for real time ECG signal denoising based on recursive filtering. *Biomed. Signal Process. Control.* **2016**, *27*, 134–144. [[CrossRef](#)]
48. Upadhyay, P.; Upadhyay, S.; Shukla, K. Schrödinger Equation Based ECG Signal Denoising. *Chin. J. Phys.* **2022**, *77*, 2238–2257. [[CrossRef](#)]
49. Zhang, M.; Jin, H.; Zheng, B.; Luo, W. Deep Learning Modeling of Cardiac Arrhythmia Classification on Information Feature Fusion Image with Attention Mechanism. *Entropy* **2023**, *25*, 1264. [[CrossRef](#)]
50. Hassaballah, M.; Wazery, Y.M.; Ibrahim, I.E.; Farag, A. ECG heartbeat classification using machine learning and metaheuristic optimization for smart healthcare systems. *Bioengineering* **2023**, *10*, 429. [[CrossRef](#)] [[PubMed](#)]
51. Nascimento, N.M.M.; Marinho, L.B.; Peixoto, S.A.; do Vale Madeiro, J.P.; de Albuquerque, V.H.C.; Filho, P.P.R. Heart arrhythmia classification based on statistical moments and structural co-occurrence. *Circuits Syst. Signal Process.* **2020**, *39*, 631–650. [[CrossRef](#)]
52. Tuncer, T.; Dogan, S.; Pławiak, P.; Acharya, U.R. Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. *Knowl.-Based Syst.* **2019**, *186*, 104923. [[CrossRef](#)]
53. Anand, G.; Nayak, R. DeLTa: Deep local pattern representation for time-series clustering and classification using visual perception. *Knowl.-Based Syst.* **2021**, *212*, 106551. [[CrossRef](#)]
54. Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [[CrossRef](#)]
55. Shaker, A.M.; Tantawi, M.; Shedeed, H.A.; Tolba, M.F. Generalization of convolutional neural networks for ECG classification using generative adversarial networks. *IEEE Access* **2020**, *8*, 35592–35605. [[CrossRef](#)]
56. Abubaker, M.B.; Babayigit, B. Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods. *IEEE Trans. Artif. Intell.* **2022**, *4*, 373–382. [[CrossRef](#)]
57. Dhara, S.K.; Bhanja, N.; Khampariya, P. An adaptive heart disease diagnosis via ECG signal analysis with deep feature extraction and enhanced radial basis function. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2023**, 1–23. [[CrossRef](#)]
58. Wei, B.; Han, Z.; He, X.; Yin, Y. Deep learning model based breast cancer histopathological image classification. In Proceedings of the 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 28–30 April 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 348–353.
59. Rai, H.M.; Chatterjee, K.; Dashkevych, S. The prediction of cardiac abnormality and enhancement in minority class accuracy from imbalanced ECG signals using modified deep neural network models. *Comput. Biol. Med.* **2022**, *150*, 106142. [[CrossRef](#)] [[PubMed](#)]

60. Liu, Z.; Chen, Y.; Zhang, Y.; Ran, S.; Cheng, C.; Yang, G. Diagnosis of arrhythmias with few abnormal ECG samples using metric-based meta learning. *Comput. Biol. Med.* **2023**, *153*, 106465. [[CrossRef](#)]
61. Shoughi, A.; Dowlatshahi, M.B. A practical system based on CNN-BLSTM network for accurate classification of ECG heartbeats of MIT-BIH imbalanced dataset. In Proceedings of the 2021 26th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 3–4 March 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
62. Shaukat, K.; Alam, T.M.; Luo, S.; Shabbir, S.; Hameed, I.A.; Li, J.; Abbas, S.K.; Javed, U. A review of time-series anomaly detection techniques: A step to future perspectives. In Proceedings of the Advances in Information and Communication: Proceedings of the 2021 Future of Information and Communication Conference (FICC), Vancouver, BC, Canada, 29–30 April 2021; Springer: Berlin/Heidelberg, Germany, 2021; Volume 1, pp. 865–877.
63. Boniol, P.; Paparrizos, J.; Palpanas, T. New Trends in Time-Series Anomaly Detection. In Proceedings of the International Conference on Extending Database Technology (EDBT), Ioannina, Greece, 28–31 March 2023.
64. Erkuş, E.C.; Purutçuoğlu, V. A new collective anomaly detection approach using pitch frequency and dissimilarity: Pitchy anomaly detection (PAD). *J. Comput. Sci.* **2023**, *72*, 102084. [[CrossRef](#)]
65. Boniol, P.; Palpanas, T.; Meftah, M.; Remy, E. Graphan: Graph-based subsequence anomaly detection. *Proc. Vldb Endow.* **2020**, *13*, 2941–2944. [[CrossRef](#)]
66. Boniol, P.; Palpanas, T. Series2graph: Graph-based subsequence anomaly detection for time series. *arXiv* **2022**, arXiv:2207.12208.
67. Schneider, J.; Wenig, P.; Papenbrock, T. Distributed detection of sequential anomalies in univariate time series. *Vldb J.* **2021**, *30*, 579–602. [[CrossRef](#)]
68. Pourhabibi, T.; Ong, K.L.; Kam, B.H.; Boo, Y.L. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decis. Support Syst.* **2020**, *133*, 113303. [[CrossRef](#)]
69. Ma, M.; Han, L.; Zhou, C. BTAD: A binary transformer deep neural network model for anomaly detection in multivariate time series data. *Adv. Eng. Inform.* **2023**, *56*, 101949. [[CrossRef](#)]
70. Zarei, R.; Huang, G.; Wu, J. GraphTS: Graph-represented time series for subsequence anomaly detection. *PloS ONE* **2023**, *18*, e0290092. [[CrossRef](#)] [[PubMed](#)]
71. Liu, Y.; Wang, L.; Yan, Y. Persistence Landscape-based Topological Data Analysis for Personalized Arrhythmia Classification. In Proceedings of the 2023 IEEE 19th International Conference on Body Sensor Networks (BSN), Boston, MA, USA, 9–11 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
72. Nazir, S.; Dickson, D.M.; Akram, M.U. Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks. *Comput. Biol. Med.* **2023**, *156*, 106668. [[CrossRef](#)] [[PubMed](#)]
73. Fotohinasab, A.; Hocking, T.; Afghah, F. A greedy graph search algorithm based on changepoint analysis for automatic QRS complex detection. *Comput. Biol. Med.* **2021**, *130*, 104208. [[CrossRef](#)] [[PubMed](#)]
74. Subasi, A.; Dogan, S.; Tuncer, T. A novel automated tower graph based ECG signal classification method with hexadecimal local adaptive binary pattern and deep learning. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 711–725. [[CrossRef](#)]
75. Jiang, Z.; Almeida, T.P.; Schindwein, F.S.; Ng, G.A.; Zhou, H.; Li, X. Diagnostic of multiple cardiac disorders from 12-lead ECGs using graph convolutional network based multi-label classification. In Proceedings of the 2020 Computing in Cardiology, Rimini, Italy, 13–16 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–4.
76. Kobat, M.A.; Karaca, O.; Barua, P.D.; Dogan, S. PrismaToidpatnet54: An accurate ECG signal classification model using prismaToid pattern-based learning architecture. *Symmetry* **2021**, *13*, 1914. [[CrossRef](#)]
77. Faes, L.; Liu, X.; Kale, A.; Bruynseels, A.; Shamdas, M.; Moraes, G.; Fu, D.J.; Wagner, S.K.; Kern, C.; Ledsam, J.R.; et al. Deep Learning under Scrutiny: Performance against Health Care Professionals in Detecting Diseases from Medical Imaging-Systematic Review and Meta-Analysis. 2019. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3384923 (accessed on 22 January 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.