

Deep Learning for Myocardial Infarction Detection Using Electrocardiogram Images: A Systematic Review

J. Octavio Gutierrez-Garcia ^{1,†} , Edgar Roman-Rangel ^{1,*,†}  and Juan Manuel Rendón-Mancha ^{2,*,†} 

¹ Department of Computer Science, ITAM, Rio Hondo 1, Mexico City 01080, Mexico; octavio.gutierrez@itam.mx

² Centro de Investigación en Ciencias, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Cuernavaca 62209, Mexico

* Correspondence: edgar.roman@itam.mx (E.R.-R.); rendon@uaem.mx (J.M.R.-M.)

† These authors contributed equally to this work.

Abstract

Myocardial infarction is a leading cause of global mortality. Electrocardiograms (ECG) are the standard diagnostic tool; however, even among medical experts, the accuracy of ECG-supported diagnoses varies. To improve diagnostic accuracy, deep learning has emerged as a promising approach. Although numerous studies demonstrate its potential, the field lacks a unified characterization of state-of-the-art methods using ECG images. This systematic review, following PRISMA guidelines, addresses this gap by analyzing studies presenting deep learning models trained on ECG images for myocardial infarction detection. Searches across six scientific databases yielded 361 records, which were filtered to 47 articles using inclusion, exclusion, and quality criteria. Guided by 12 research questions, this review contributes (i) a characterization of deep learning architectures for myocardial infarction detection using ECG images; (ii) an assessment of training and evaluation practices of deep learning models; (iii) a description of state-of-the-art results in terms of machine learning metrics; (iv) a pioneering exploration of the use of vision transformers for myocardial infarction detection; (v) a compilation of ECG databases; and (vi) future research directions aimed at advancing deep learning approaches for myocardial infarction detection, e.g., involving domain experts in evaluating deep learning models to guarantee their safe deployment in clinical settings.

Keywords: deep learning; myocardial infarction; electrocardiogram; ECG; ECG classification; computer vision; machine learning

MSC: 68T01; 68T45



Academic Editor: Lingfeng Liu

Received: 17 December 2025

Revised: 28 January 2026

Accepted: 6 February 2026

Published: 10 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Cardiovascular diseases, in particular myocardial infarction, remain the leading cause of global mortality [1] and a major contributor to the burden on the health system worldwide [2]. Accurate early diagnosis of myocardial infarction is essential to prevent catastrophic outcomes, a task for which the electrocardiogram (ECG) serves as the standard diagnostic tool [3]. Indeed, the ECG provides data on cardiac rhythm, electrical conduction, and morphology [4].

However, even among medical experts, ECG-supported diagnoses and, consequently, diagnostic accuracy vary due to differences in interpretation and diagnostic skills [5]. This problem is exacerbated by the global disparity in access to specialized cardiologists, a disparity that persists even within developed countries [6]. In contrast, there is a rapidly

growing volume of ECG data generated by modern electronic health devices, e.g., [7], that can be used to improve diagnosis. These scenarios create an urgent imperative for advanced analytical tools to standardize the interpretation of ECGs and provide diagnostic support to overburdened or non-specialized physicians. Consequently, artificial intelligence techniques, particularly those combined with computer vision models, have emerged as one of the most investigated solutions for automated ECG analysis and myocardial infarction detection [8,9].

Due to the importance of myocardial infarction for population health and the growing use of artificial intelligence techniques in this domain, numerous reviews have been conducted on related topics. In general, those reviews differ in both methodology and scope. Regarding methodology, there are narrative reviews [10–12], bibliometric analysis [13], scoping reviews [14–16], and systematic reviews [8,17–25].

Focusing on systematic reviews (comparable to the present work) and analyzing their scope, some are highly specialized. For example, certain reviews concentrate on arrhythmia associated with myocardial infarction (e.g., [24]), ECG wearables for cardiac health monitoring (e.g., [25]), or machine learning models trained exclusively on 12-lead ECGs (e.g., [20]). Alternatively, some systematic reviews adopt a broader scope; for instance, Petmezas et al. [17] examined deep learning models trained on ECG data published within a two-year range (2020–2021), categorizing them by application domain, e.g., cardiovascular disease diagnosis, sleep analysis, biometric recognition, among others. Also broad in scope, Musa et al. [8] analyzed scientific articles presenting deep learning models trained on ECGs regardless of their application domain (as in [17]). Likewise, Wu and Guo [21] conducted a systematic review focused on deep learning and electrocardiography to determine which cardiovascular diseases are being diagnosed using deep learning models trained on ECGs. Analogously, application-domain agnostic, the systematic survey by Khalid et al. [18] focused on ECG classification, searching Scopus, PubMed, and IEEE Xplore, and selecting 90 articles. However, their search string included only “ECG classification”, “electrocardiogram classification”, and “deep learning”, which may have excluded relevant articles (e.g., those using the phrase “classification of ECGs”). Additionally, their analysis was guided by only two specific research questions. Oke and Cavus [19] searched for articles presenting machine learning models trained on ECG data for detecting and predicting heart conditions (e.g., arrhythmia). However, their review aimed to assess the impact of artificial intelligence on cardiology rather than recent advances in deep learning for myocardial infarction detection. Similarly comprehensive in scope, Fuadah and Lim [22] and Pandey et al. [23] examined work on machine learning algorithms (including deep learning) for classifying cardiovascular diseases in general.

The main differences of this work from other systematic reviews are as follows. This work focuses on deep learning models trained exclusively on ECG images, whereas other reviews considered models trained on either ECG images or time series signals (a.k.a., leads or derivations), which may require different algorithms. This review is the only systematic review involving vision transformers (as the search string included related terms). Furthermore, this review posed a set of research questions emphasizing evaluation aspects, such as whether domain experts were involved. Moreover, unlike Zworth et al. [20], Fuadah and Lim [22], Pandey et al. [23], who focused on machine learning models (e.g., random forest and gradient boosting), the present systematic review concentrates exclusively on deep learning models, and in contrast to Musa et al. [8], Petmezas et al. [17], Khalid et al. [18], Wu and Guo [21], who analyzed works using deep learning models trained on ECGs regardless of their application domain, this review focuses specifically on myocardial infarction detection. To the best of the authors’ knowledge, it is the first systematic review to discuss uncertainty quan-

tification in myocardial infarction diagnosis. Finally, unlike in the work of Petmezas et al. [17], Oke and Cavus [19], Wu and Guo [21], Fuadah and Lim [22], Pandey et al. [23], no time constraints were applied to the searches for scientific articles; however, a publication period of relevant articles emerged as one of the findings of this systematic review.

The objective of this systematic review is to characterize the state of the art in myocardial infarction detection using deep learning models trained on ECG images. To achieve this objective, the methodology adopted for this review was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [26], which are widely used for conducting systematic surveys across diverse domains, e.g., [27]. Accordingly, systematic and reproducible searches were performed in six scientific databases: ACM Digital Library, IEEE Xplore, PubMed, ScienceDirect, Scopus, and Web of Science. The consolidated search results yielded 361 article records, which were filtered using inclusion, exclusion, and quality criteria to ensure the analysis of relevant scientific articles. This process resulted in 47 selected articles proposing deep learning models trained on ECG images for myocardial infarction detection. Subsequently, a bibliometric analysis was conducted using the selected articles. Afterward, a comprehensive analysis and synthesis of the selected articles were performed and guided by a set of research questions designed to examine myocardial infarction detection using deep learning models trained on ECG images, considering aspects such as deep learning architectures, available databases, target classes, model training and evaluation practices, and quantitative results (see Section 2.1 for details).

By addressing the proposed research questions, this systematic review makes the following contributions:

- A characterization of deep learning architectures for myocardial infarction detection using ECG images exclusively.
- An assessment of training and evaluation practices of deep learning models trained on ECG images for myocardial infarction detection.
- A description of state-of-the-art results in terms of machine learning metrics for deep learning models detecting myocardial infarction.
- A pioneering exploration of the use of vision transformers trained on ECG images for myocardial infarction detection.
- A compilation of available ECG databases that can be used to train deep learning models to detect cardiac conditions.
- A set of future research directions aimed at advancing deep learning approaches for myocardial infarction detection and addressing existing research gaps (identified from the analysis of the selected articles).

This article is structured as follows. Section 2 describes the methodology used in the present systematic review and lists the research questions that guided this work. Section 3 presents the search results and the answers to the proposed research questions. Section 4 discusses the results derived from the research questions posed. Section 5 includes a set of future research directions to advance the field of myocardial infarction detection using deep learning. Finally, Section 6 presents concluding remarks.

2. Method

The methodology for conducting this review was based on the PRISMA guidelines [26] for reporting systematic reviews. The protocol for this systematic review has been registered with the Open Science Framework (OSF) of the Center for Open Science (see [28]).

This section presents (i) the research questions that guided the analysis of the scientific literature; (ii) the search strategy used to identify related research articles; (iii) the inclusion, exclusion, and quality criteria used to screen relevant research articles; (iv) the procedure

followed to extract data from the selected articles; and (v) the synthesis method to answer the research questions.

2.1. Research Questions

The research questions (RQs) posed in this systematic review are as follows:

- RQ1. What deep learning architectures are commonly used to detect myocardial infarction on ECG images?
- RQ2. Is transfer learning utilized in research on myocardial infarction detection using deep learning? If so, which techniques are applied?
- RQ3. What class labels are used by deep learning models for the detection of myocardial infarction?
- RQ4. At what level of detail do works on myocardial infarction detection describe their deep learning models?
- RQ5. What metrics are used to evaluate deep learning models for myocardial infarction detection?
- RQ6. What are the best reported results for the detection of myocardial infarction supported by deep learning?
- RQ7. What ECG datasets are used to train and test the deep learning models for the detection of myocardial infarction?
- RQ8. How many ECG leads, and which ones, are used to train deep learning models for myocardial infarction detection?
- RQ9. Are ECG datasets used to train deep learning models for the detection of myocardial infarction imbalanced? If so, how do research efforts tackle class imbalance?
- RQ10. Do works on myocardial infarction detection generate synthetic data for training their models?
- RQ11. What preprocessing techniques are used for the detection of myocardial infarction supported by deep learning?
- RQ12. What future work directions are proposed by research efforts focused on myocardial infarction detection supported by deep learning?

It should be noted that some of the above research questions have more than one reason for being proposed. Overall, the rationale behind their formulation is as follows. Research questions RQ1 and RQ2 were posed to explore what deep learning architectures, either well-known or customized, are used to detect myocardial infarction and related factors. Research question RQ3 was posed to identify the scope and specific objectives (in terms of target classes) of deep learning models in the context of cardiovascular conditions related to myocardial infarction. Research questions RQ2, RQ4, RQ7, RQ8, RQ10, and RQ11 were posed to explore aspects related to reproducibility. Research questions RQ5 and RQ6 were designed to explore performance evaluation practices and characterize state-of-the-art quantitative results, respectively. Research question RQ7 was posed to identify relevant ECG datasets (and their characteristics) that can be used to train deep learning models. Research question RQ8 was posed to determine what ECG leads are used to train deep learning models within this application domain. Research questions RQ2, RQ8, RQ9, RQ10, and RQ11 were designed to explore model training practices. Finally, research question RQ12 was posed to explore possible research gaps within the field identified by authors of the selected articles.

2.2. Search Strategy

To focus the analysis on scientific articles related to the detection of myocardial infarction and related factors using ECG images supported by deep learning, the structure of the search string (Figure 1) consisted of keywords related to (i) ECGs; (ii) deep learning

and computer vision; (iii) images; and (iv) myocardial infarction. Relevant terms were included either as acronyms or in their full form: convolutional neural network (CNN), vision transformer (ViT), myocardial infarction (MI), electrocardiogram (ECG and EKG), acute coronary syndrome (ACS), acute myocardial infarction (AMI), segment elevation (STE), and segment elevation myocardial infarction (STEMI).

In addition, based on a round of preliminary searches and in order to retrieve a list of potentially relevant scientific articles, the search string (Table 1) also included a set of exclusion terms. These terms (e.g., “image registration” and “electroencephalogram”) appeared in the titles and abstracts of several articles unrelated to the objective of this systematic review and were therefore excluded from the searches.

The keywords were selected based on (i) their relevance to the objective of this systematic review and (ii) an exploration of suitable synonyms and related terms. Whereas some keywords were specific, e.g., ViT and STEMI, others were general, e.g., deep learning and electrocardiogram, which resulted in a representative list of contextually significant scientific articles. It is worth mentioning that the search results were not restricted to a given time period to include both recent and pioneering scientific articles. The full search string is reported in Table 1.

The searches were conducted on 21 February 2025 using the following scientific databases: ACM Digital Library, IEEE Xplore, PubMed, ScienceDirect, Scopus, and Web of Science. Databases that may include gray literature and non-peer-reviewed articles such as arXiv and Google Scholar were not included because this systematic review is focused on peer-reviewed scientific articles and supported by reproducible search results.



Figure 1. Structure of the search string.

Table 1. Search string.

Keyword Category	Keywords
Electrocardiogram	“electrocardiog*” OR “ecg*” OR “ekg*”
Deep Learning and Computer Vision	“convolutional” OR “cnn*” OR “convolution*” OR “transformer*” OR “deep learning” OR “computer vision” OR “ViT”
Image	“imag*”
Myocardial Infarction	“myocardial infarct*” OR “heart attack*” OR “heart arrest*” OR “ste” OR “stemi” OR “ami” OR “segment elevation” OR “mi” OR “acs” OR “acute” OR “coronary syndrome” OR “st-elevation” OR “st elevation”
Terms to be excluded	“image registration” OR “EEG*” OR “tomograph*” OR “tomography” OR “mri*” OR “non-ECG-gated” OR “expression recognition” OR “echo*” OR “ct” OR “emotion*” OR “psych*” OR “apnea” OR “flu” OR “electroencephalogram*” OR “multiomic*” OR “x-ray” OR “ica” OR “angiography” OR “cta” OR “instead of ECG*”

2.3. Selection Criteria

The selected articles analyzed in this systematic review were screened using exclusion, inclusion, and quality criteria (Figure 2).

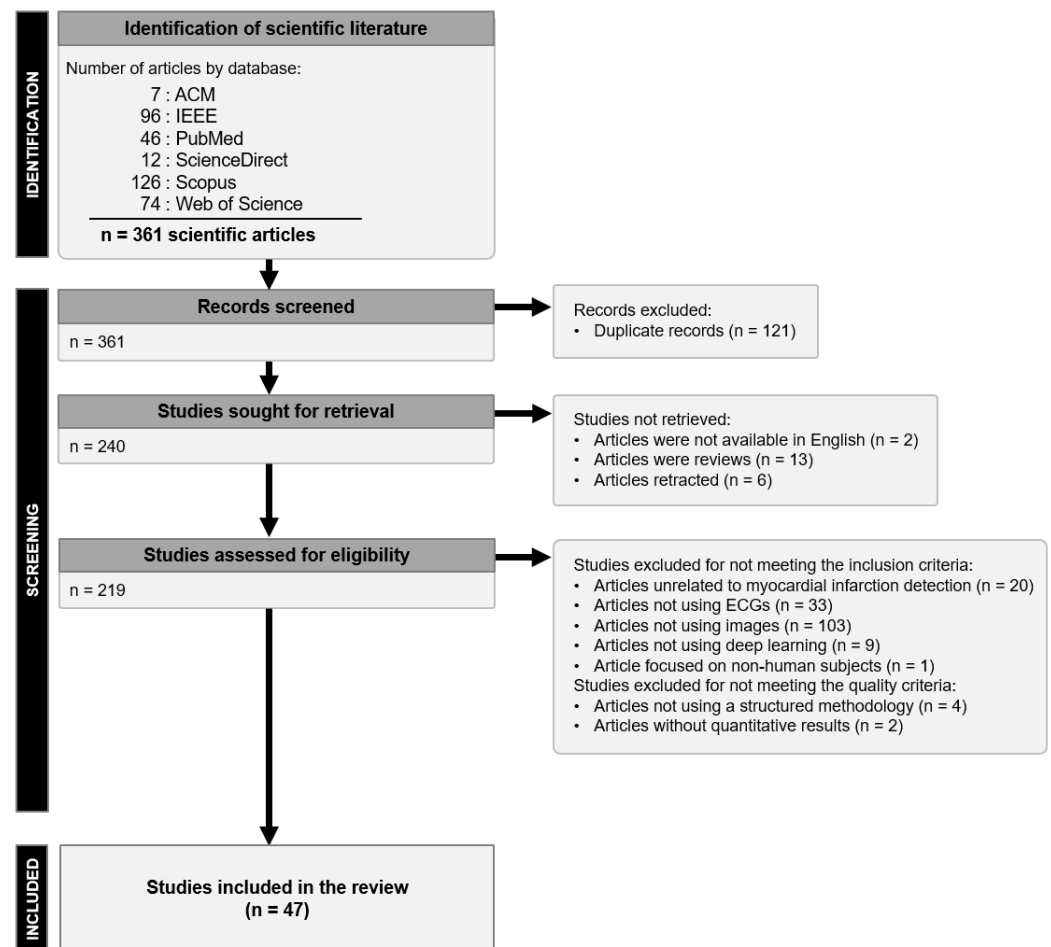


Figure 2. Article screening process.

The exclusion criteria were as follows:

- Articles written in languages other than English.
- Articles presenting reviews, overviews or surveys.
- Retracted articles.

The inclusion criteria were as follows:

- Articles related to the detection of myocardial infarction (and associated factors).
- Articles proposing systems whose input is an ECG image.
- Articles using images with plotted derivations.
- Articles proposing deep learning models.
- Articles using ECGs of human subjects.

The quality criteria were as follows:

- Articles presenting quantitative performance evaluations of their deep learning models.
- Articles following a structured research methodology, e.g., articles reporting a pipeline for training and evaluating their deep learning models as well as discussing their results.

The selection criteria were applied in the following order. Firstly, the raw search results obtained from the six scientific databases were combined into a single file and duplicates

were removed. Afterward, the exclusion criteria were applied. It should be noted that even though some of the search engines allowed filtering articles based on some of the exclusion criteria, their search results included a few rejected articles, articles not written in English, or articles presenting reviews, overviews or surveys. Then, the exclusion criteria were applied manually by skimming through article titles and abstracts, as well as by visiting their official websites to look for rejection notices. Subsequently, the inclusion criteria were applied, first by reading article titles and abstracts, and when it was infeasible to make a decision based on these, the full text of the article was read. Once the inclusion criteria were applied, the remaining articles were carefully read in full to assess them against the quality criteria. It should also be noted that the three authors of this systematic review independently applied both the inclusion and quality criteria. In cases where complete agreement was not reached, the inclusion of the articles was discussed until a consensus was reached.

2.4. Data Collection Process and Data Extraction Strategy

To collect and consolidate relevant data items associated with the research questions posed in this systematic review, a database consisting of a single table (representing a data extraction form) was created. Whereas a record was associated with a selected article, the fields were associated with the research questions. Afterward, a couple of selected studies were used to pilot (and adjust) the data extraction form. Subsequently, the selected articles were read in full to identify data items (e.g., deep learning architectures) relevant to answering the posed research questions.

2.5. Synthesis Method

The data extracted from the selected articles were carefully examined by means of interpretation, abstraction, categorization, and integration. The synthesis process was performed collaboratively by the authors of this systematic review.

3. Results

This section presents the search results (Section 3.1), descriptive statistics on the selected articles (Section 3.2), and the answers to the proposed research questions (Section 3.3).

3.1. Search Results

Figure 2 summarizes the article identification and screening process. A total of 361 article records were retrieved from the six scientific databases selected for this systematic review. Afterward, the search results were consolidated, and 121 duplicate records were removed. Then, the selection criteria (defined in Section 2.3) were applied to the remaining 240 articles. As a result of applying the exclusion, inclusion, and quality criteria, 21, 166, 6 articles were removed, respectively. The screening process identified 47 articles focused on myocardial detection using deep learning models, with ECG images as input.

3.2. Selected Studies Statistics

The selected studies were published as journal articles (51.06%), conference papers (46.81%), and book chapters (2.13%). For these 47 manuscripts, Table 2 presents the distribution of impact factor quantiles and rankings for journals and conferences where the selected articles were published, respectively. Whereas the quantiles were extracted from the most recent Journal Citation Reports by Clarivate at the time of writing this present article, the conference rankings were obtained from the CORE rankings published by the Computing Research and Education Association of Australasia.

Table 2. Distribution of impact factor quantiles and rankings for journals and conferences where the selected articles were published. Only 46 papers are listed here, as 1 of the 47 selected articles corresponds to a book chapter.

Frequency	
Journals	
Q1	13
Q2	6
Q3	2
Q4	1
Not indexed	2
Conferences	
A+	0
A	1
B	0
C	3
Not ranked	18

The most frequent publication sources were Scientific Reports, IEEE Access, and International Conference on Computing Communication and Networking Technologies, with three, two, and two selected articles, respectively. This suggests that no publication source concentrated research studies on deep learning for myocardial infarction detection using ECG images. However, IEEE journals and IEEE conferences were prevalent as can be observed in Figure 3 showing a word cloud constructed from the publication sources of the selected articles. The word cloud of publication sources suggests that the scope of the journals and conferences (where the selected articles were published) is mostly related to computer science and, in particular, related to artificial intelligence applications. Figure 3 also shows that journals or conferences focused on medical advances are not predominant among the publication sources of the selected articles.

As guided by the search string (reported in Table 1), in general, the domain of the selected articles (as defined by their titles, see Figure 4) is characterized by topics related to ECG images, deep learning models for detection and classification, and myocardial infarction.

Although no time window was defined in the search string, the retrieved articles were published between 2018 and 2025, which emerged as an outcome of the search process. In addition, as shown in Figure 5, there is a positive trend in the number of publications by year (except for 2025, since the searches of this systematic review were carried out in February 2025).



Figure 3. A word cloud generated from the publication sources of the selected articles.

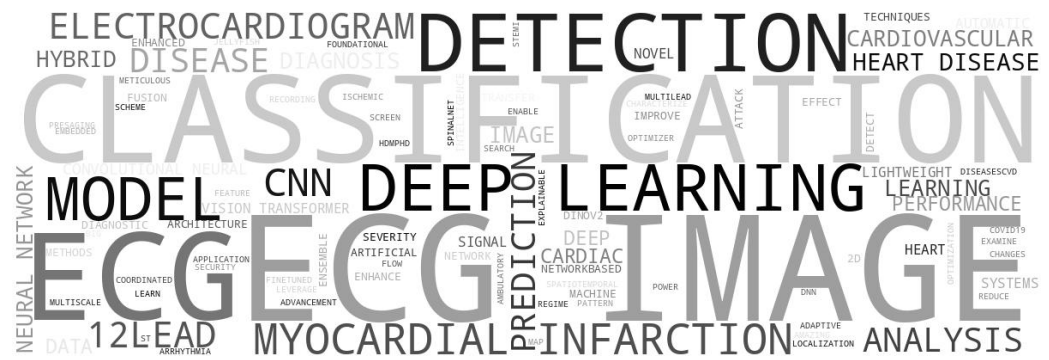


Figure 4. A word cloud generated from the titles of the selected articles.

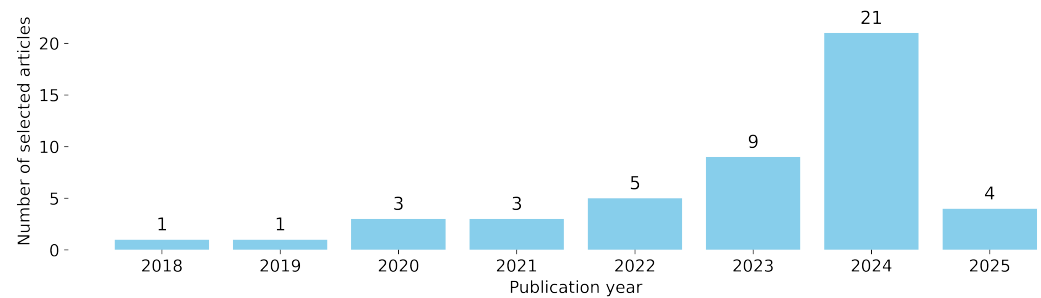


Figure 5. Selected articles by publication year.

3.3. Data Synthesis: Responses to Research Questions

This section presents the responses to the research question posed in this systematic review, which aims to characterize how myocardial infarction detection supported by ECG images is performed using deep learning.

3.3.1. RQ1: What Deep Learning Architectures Are Commonly Used to Detect Myocardial Infarction on ECG Images?

The most commonly used neural network architectures in the selected articles are, firstly, customized ad hoc CNNs (see Figure 6), where authors define the number and type of layers in the model. In second place, well-known architectures are employed. The most widely used among these are EfficientNet [29], ResNet [30], DenseNet [31], VGG [32], and InceptionNet [33]. Moreover, in [34] a siamese network is presented and in [35] an ensemble model is used. Figure 6 shows the frequency distribution of all architectures used in the selected articles. In addition, the deep learning models reported in the selected articles fall into three architecture types: CNNs, transformers, and recurrent neural networks (RNNs). Figures 6 and 7 list all the models found for each architecture type and present the proportion of models corresponding to each type, respectively. It should be noted that the sum of the deep learning models presented in Figures 6 and 7 exceeds the number of selected articles because some authors (of the selected articles) included more than one model.

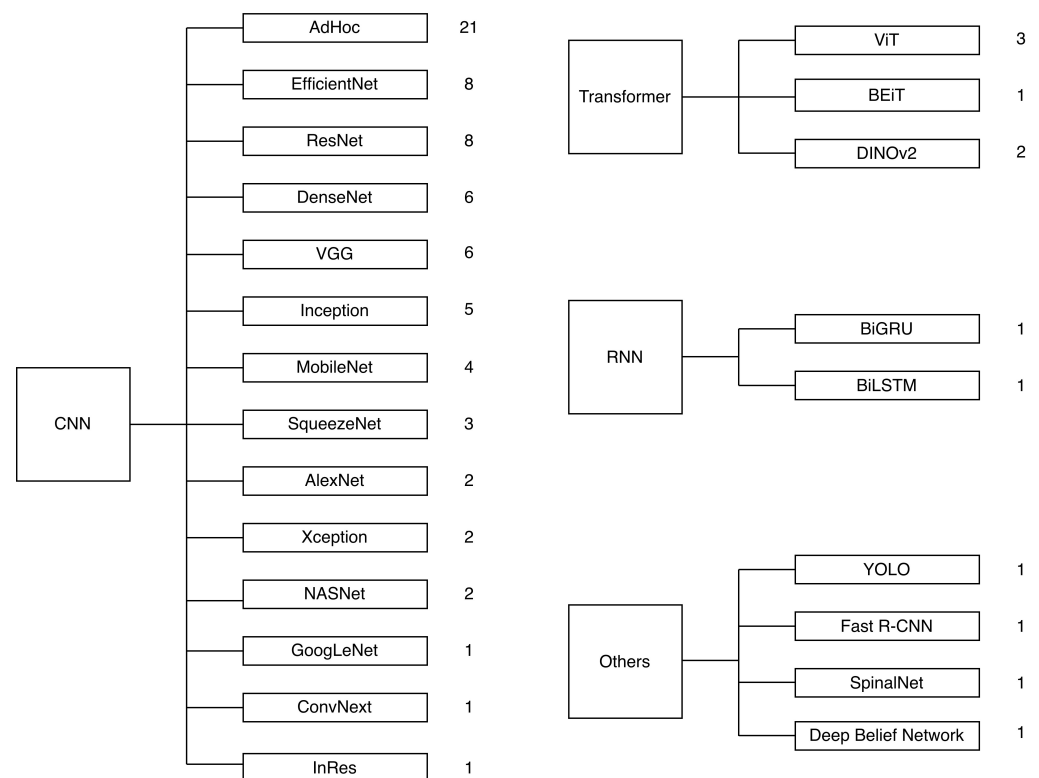


Figure 6. Taxonomy of deep neural network architectures for myocardial infarction. Numbers outside the boxes indicate the frequency of each model within the selected articles.

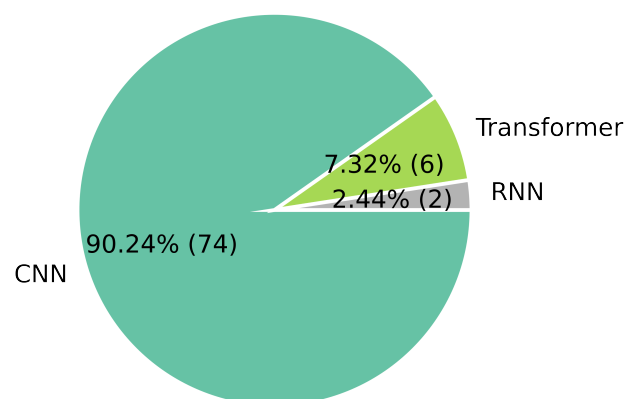


Figure 7. Proportion of deep learning models grouped by architecture type for myocardial infarction detection.

3.3.2. RQ2: Is Transfer Learning Utilized in Research on Myocardial Infarction Detection Using Deep Learning? If So, Which Techniques Are Applied?

As shown in Figure 8a, 25 out of the 47 selected articles made use of transfer learning, most of which employed fine-tuning without warm-up (see Figure 8b) to adapt their models for myocardial infarction detection. This result aligns with the use of well-known deep neural network architectures reported in Figures 6 and 7.

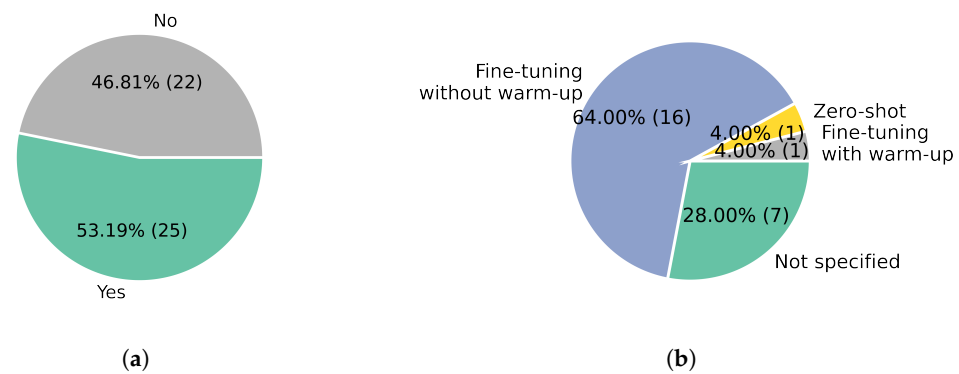


Figure 8. Use of transfer learning in the selected articles. (a) Percentage of selected articles using transfer learning. (b) Usage proportion of transfer learning techniques.

Notice that 24 out of the 25 articles that use transfer learning were pre-trained on the ImageNet dataset [36], one of which was jointly pre-trained with the COCO dataset [37], and only one of these 25 articles reported been pre-trained directly using ECG images. Nonetheless, all of them were fine-tuned using ECG images.

3.3.3. RQ3: What Class Labels Are Used by Deep Learning Models for the Detection of Myocardial Infarction?

The most used target class labels (of the models presented in the selected articles) are (i) normal, (ii) myocardial infarction, (iii) abnormal heartbeats, and (iv) history of myocardial infarction; see Figure 9 for a complete list of target classes and their frequency. In addition, each deep learning model for myocardial infarction detection grouped these classes into target class sets. Among the most frequent sets are {normal, abnormal heartbeat, myocardial infarction, history of myocardial infarction} and {normal, myocardial infarction}. See Figure 10 for a list of common target class sets and their usage proportion.

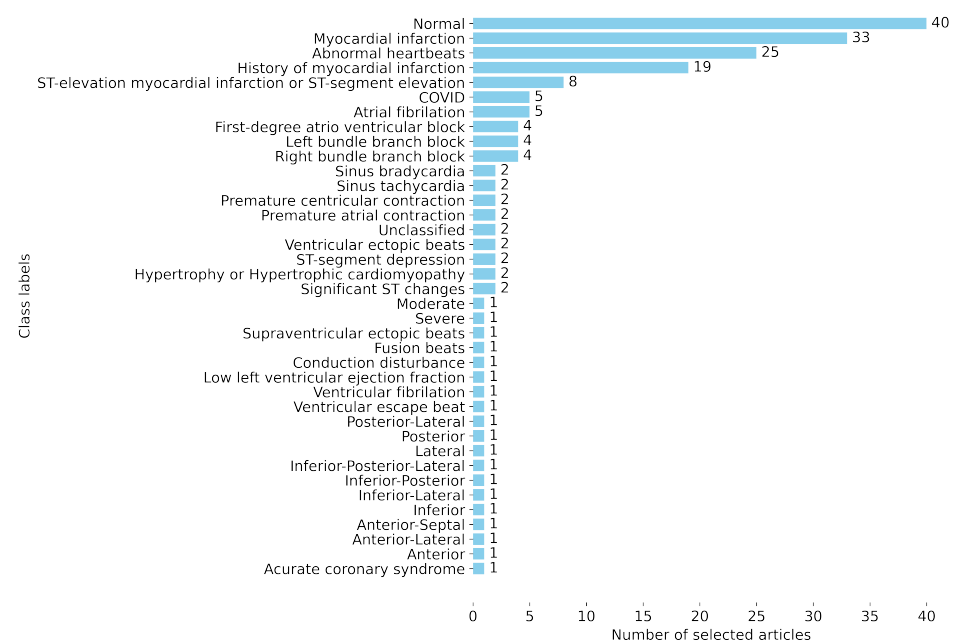


Figure 9. Frequency of target classes used by deep learning models for myocardial infarction detection.

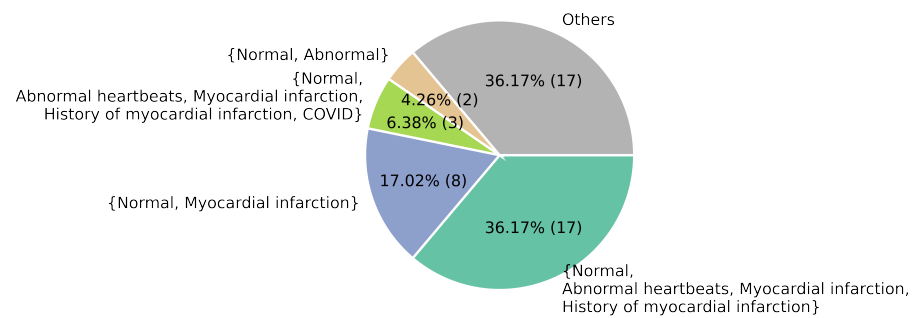


Figure 10. Proportion of target class sets used by deep learning models for myocardial infarction detection.

3.3.4. RQ4: At What Level of Detail Do Works on Myocardial Infarction Detection Describe Their Deep Learning Models?

As shown in Figure 11, 59.57% of the deep learning models were completely described in the selected articles. However, 12 out of the 47 articles provide only a partial description, which makes it difficult to ensure reproducibility. Unfortunately, seven selected articles provided no description of their models.

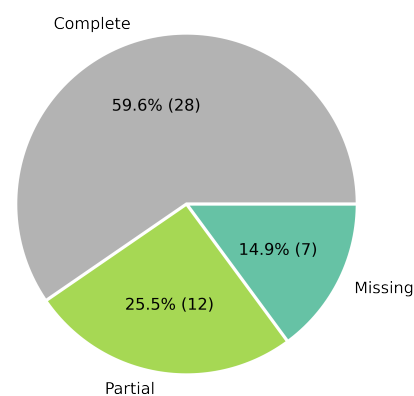


Figure 11. Proportion of selected articles providing a complete, partial, and no model description.

3.3.5. RQ5: What Metrics Are Used to Evaluate Deep Learning Models for Myocardial Infarction Detection?

As observed in Figure 12, the most common performance metrics used to evaluate deep learning models for myocardial infarction detection were accuracy, sensitivity, precision, F1 score, and specificity, which were used by 80.85%, 80.85%, 65.96%, 59.57%, and 38.30% of the selected articles, respectively. Also, it should be mentioned that the median number of metrics reported by the selected articles was 4. However, as indicated in Section 3.3.9, most of the selected articles reported working with imbalanced datasets, and some performance metrics of machine learning models trained on those datasets frequently mask poor performance [38], e.g., accuracy. Then, metrics such as the area under the precision–recall curve (only used by 8.51% of the selected articles) are among the most suitable performance metrics for imbalanced datasets [38,39].

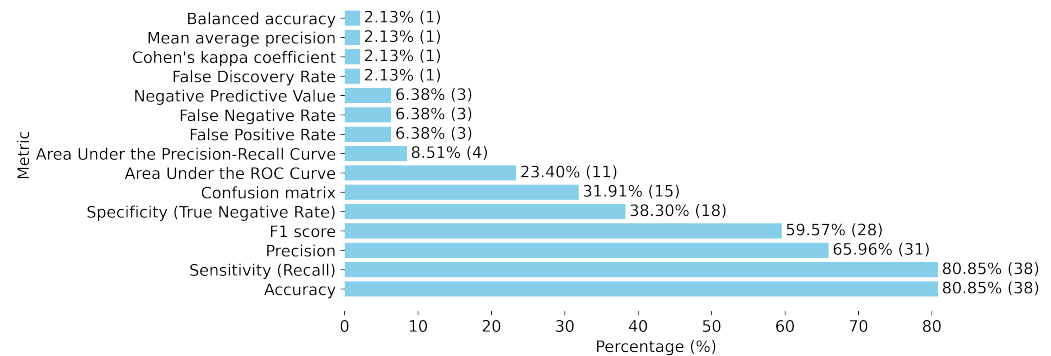


Figure 12. Distribution of the performance metrics used by the selected articles.

3.3.6. RQ6: What Are the Best Reported Results for the Detection of Myocardial Infarction Supported by Deep Learning?

To compare, to some extent, the quantitative results reported by the selected articles, Table 3 and Figure 13 report their performance measures and distribution for the six most used performance metrics (see Section 3.3.5), respectively. The metrics most used were selected because more than 80% of the selected articles reported at least one of them. In fact, only two selected articles [40,41] did not report the most used metrics, and reported the area under the precision–recall curve. Hence, for the sake of completeness, such metric was also included in the comparison. Also, to synthesize and ease performance comparison, Table 3 reports macro performance measures, and when the selected articles proposed more than one deep learning model (as in [42–44]), the best result reported for each metric was included regardless of the model. In this article, a macro performance measure is defined as the arithmetic mean of the performance metric computed independently for each class. For example, macro-precision is calculated as the average of the precision values obtained for the normal class, the myocardial infarction class, and the history of myocardial infarction class.

As shown in Figure 13, the median measure for all the performance metrics is higher than 0.95 suggesting relatively outstanding performances for a large share of the deep learning models reported in the selected articles. In fact, only a few selected articles reported performance measures lower than 0.8 for each performance metric. Although, it should be highlighted that very few selected articles reported the area under the precision–recall curve, a robust metric for imbalanced datasets [38,39]. In addition, the absence of negative results in the selected articles may be attributed to a well-documented publication bias, where only research efforts reporting positive results are published. However, within a health domain such as myocardial infarction detection, providing evidence of what does not work may help to improve or accelerate future research efforts, particularly because training deep learning models is computationally intensive and time-consuming.

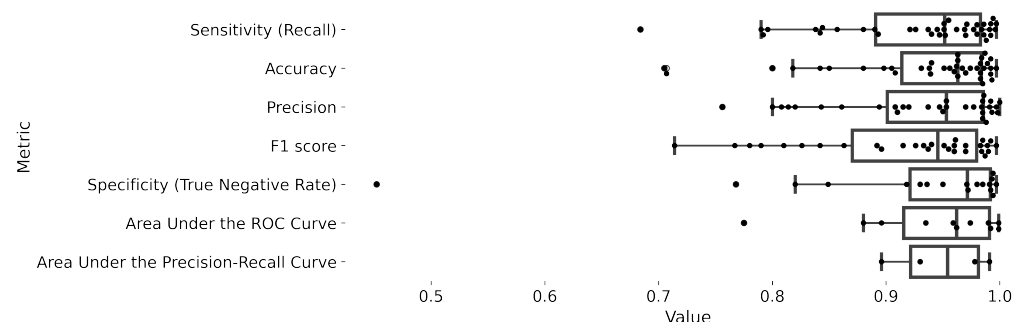


Figure 13. Distribution of performance values by metric, as reported in the selected articles.

Table 3. Macro performance measures reported in the selected articles. Column CV/HO indicates whether a cross-validation or hold-out approach was used. The column ‘Split’ lists the proportional split size for the train/validation/test sets. N: no validation set was used, only test set. E: used an external set of data for testing. U: no details are provided.

Ref.	Rec.	Acc.	Prec.	F1 Score	Specif.	AUROC	AUPRC	CV/HO	Split
[45]	0.952	0.908	-	-	0.936	-	-	HO	90/N/10
[42]	0.796	0.880	0.756	0.767	-	0.935	-	HO	90/N/10
[46]	0.844	-	-	0.892	0.849	0.896	-	HO	57/N/43
[47]	0.857	0.705	-	0.863	0.452	-	-	HO	80/N/20
[40]	-	-	-	-	-	-	0.93	HO	100/N/E
[48]	0.842	0.842	0.843	0.842	-	-	-	HO	U
[49]	0.969	0.983	0.977	0.961	0.980	-	-	HO	70/20/10
[50]	-	0.931	-	-	-	-	-	HO	60/N/40
[43]	0.996	0.993	0.998	-	0.997	-	-	CV	5 fold
[51]	0.951	0.951	0.953	0.951	-	-	-	CV	10 fold
[52]	-	0.956	-	-	-	-	-	U	U
[53]	0.955	0.963	0.953	0.955	0.993	-	-	HO	80/10/10
[54]	0.992	0.992	0.993	0.992	-	0.999	-	HO	70/20/10
[55]	0.926	0.938	0.947	0.933	-	-	-	CV	5 fold
[56]	0.962	0.963	0.894	0.926	-	0.962	-	HO	48/5/47
[57]	0.940	-	0.950	0.940	-	-	-	HO	80/N/20
[58]	0.937	0.940	0.937	0.937	-	-	-	HO	80/N/20
[59]	0.684	0.960	0.861	0.714	0.971	-	-	HO	79/4/17
[44]	0.921	0.963	0.908	0.915	0.991	-	-	HO	80/10/10
[60]	0.890	0.970	0.920	0.960	-	-	-	U	U
[61]	0.994	-	0.986	0.990	-	-	-	HO	80/N/10
[62]	-	0.850	-	-	-	-	-	CV	5 fold
[63]	0.79	0.984	0.800	0.790	-	-	-	HO	80/N/20
[64]	-	0.980	-	-	-	-	-	HO	60/20/20
[65]	0.971	0.986	1.000	0.985	-	-	-	HO	60/N/40
[66]	0.977	-	0.991	0.970	-	-	0.978	CV	5 fold
[67]	0.997	0.997	0.997	0.997	-	-	-	HO	60/30/10
[68]	0.991	0.992	-	-	0.994	-	-	HO	80/10/10
[69]	-	0.967	-	0.960	-	-	-	HO	70/N/30
[70]	0.951	0.974	0.985	-	0.985	-	-	HO	60/20/20
[71]	0.988	0.985	0.985	0.987	-	-	-	CV	U
[72]	0.792	0.818	0.808	0.780	-	-	-	HO	80/N/20
[73]	0.980	0.983	0.985	-	-	-	-	CV	U
[74]	0.980	0.939	-	-	0.768	-	-	CV	U
[35]	0.970	-	0.970	0.970	-	0.775	-	HO	80/10/10
[75]	0.986	0.989	0.992	0.989	-	0.999	-	HO	80/10/10
[76]	0.945	0.962	0.915	-	0.972	0.990	-	HO	80/10/10
[77]	-	-	-	-	-	0.974	-	HO	70/20/10
[34]	0.893	-	0.910	0.896	0.950	-	-	CV	U
[78]	0.992	0.992	-	-	0.992	-	-	HO	66/N/33
[79]	0.880	0.800	0.820	0.810	0.820	0.880	-	HO	68/16/16
[80]	0.947	0.898	-	-	0.918	-	-	HO	90/N/10
[81]	0.986	0.987	0.988	0.984	-	0.992	-	HO	U
[82]	-	0.707	-	-	-	-	-	HO	80/N/20
[83]	0.984	0.983	0.984	0.983	0.994	-	-	HO	80/N/20
[84]	0.838	0.905	0.814	0.826	0.930	0.959	0.896	CV	U
[41]	-	-	-	-	-	-	0.991	U	U

Table 3 also shows the validation approach of the selected papers, either cross-validation or hold-out, along with the size of the databases splits when available. We can see that 34 works relied on a hold-out approach. However, 17 of them neglected the use of a validation set, and only trained and tested their models. Additionally, only 11 articles used cross-validation, although five of them did not provide details about the number of folds. Furthermore, there are even three articles that provided no details at all. This lack of standard procedure for validation of deep learning models presents both a serious methodological flaw but also an area of opportunity for the research community.

The performance results presented in Table 3 are informative, but they should not be used for exact comparisons since the deep learning models, among other aspects, were trained on different datasets and using different experimental settings, e.g., a different proportion of training, test, and validation sets.

It should be noted that hypothesis tests were conducted to assess potential differences in performance among groupings of the selected articles, for example, based on impact factor quantiles, evaluation methodology (i.e., cross-validation versus train/validation/test splits), the use of imbalanced datasets, and the number of ECG leads used. However, no statistically significant differences were detected.

3.3.7. RQ7: What ECG Datasets Are Used to Train and Test the Deep Learning Models for the Detection of Myocardial Infarction?

Out of the 47 selected articles, 7 used two datasets, while the remaining 40 used only one dataset. The ECG image datasets used in the selected articles are reported in Table 4.

Table 4. List of ECG image datasets used in the selected articles and their frequency (** indicates proprietary, private dataset).

ECG Image Dataset	Frequency
Khan et al., 2021 [85]	14
Khan et al., 2020 [86]	6
PTB Diagnostic ECG database [87]	5
PTB-XL [88]	5
Collected by the authors of the selected article	5
Not indicated	4
China physiological signal challenge (CPSC) [89]	3
CODE15 [90]	2
The European ST-T database (EU-ST-T) [91]	2
LTST-Physionet database [92]	1
Choi et al., 2023 [93]	1
UCI Cleveland heart disease [94]	1
MIT-BIH arrhythmia database [95]	1
** Mount Sinai Health System, USA	1
** Medanta Hospital, India	1
** Zhejiang Second People's Hospital, China	1
** Hualien Tzu Chi Hospital, Taiwan	1

As reported in Table 4, the records collected by Khan et al. in 2020 [86] and 2021 [85] are the two most popular datasets, which were used 6 and 14 times, respectively. The PTB [87] and PTB-XL [88] datasets were used five times each, and five selected articles collected and used their own datasets. This unbalance in the popularity of datasets might be a consequence of the fact that the data from Khan et al. [85,86] is already in image format, aligning the type of format with the scope of this review. Furthermore, the data set in [85] presents a fair balance across its classes. Additionally, the PTB-XL dataset [88] is also well suited given its size and possibility of rebalancing by means of subsampling. However, it might not be as popular given that it is relatively new compared to Khan's sets, and given that it contains time signals instead of images, thus not providing realistic examples for most physicians. Additionally, the CPSC dataset [89] was used three times. Meanwhile, the CODE15 [90] and European ST-T [91] datasets were used twice each, and four public datasets [92–95] were used only once. The last four datasets (marked with ** in Table 4) are mentioned in the selected articles, but their data is proprietary and private. Finally, four articles (referred to as “Not indicated”) do not mention the origin of the data used to train their models.

Regarding specific population groups present in the datasets, most of the selected articles focus on detecting cardiac conditions without considering any particular group of people, or do not explicitly state such considerations. In fact, only three selected articles

addressed specific groups, one of which focuses on the Japanese population and two on populations affected by COVID-19.

3.3.8. RQ8: How Many ECG Leads, and Which Ones, Are Used to Train Deep Learning Models for Myocardial Infarction Detection?

Details regarding the number of derivations (a.k.a., leads) in the image data analyzed by the selected articles are presented in Figure 14. As shown, eight articles do not specifically mention the leads used, indicated by the labels «unknown» and «single lead (not specified)». Concretely, «unknown» corresponds to five articles that provide no details, while «single lead (not specified)» refers to three articles that mention the use of a single derivation without specifying its name.

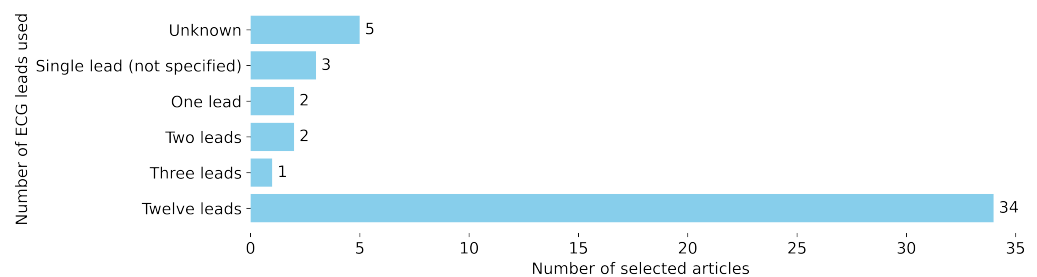


Figure 14. Distribution of the number of leads used by the selected articles.

Additionally, two articles used a single derivation: lead 2 and lead 8, respectively. Two other articles used derivations in pairs (referred to as «two leads»), namely lead 2 with lead 3, and lead 9 with lead 12. One article made use of three derivations (referred to as «three leads»): leads 7, 9, and 12. Finally, 34 articles used the common format of all 12 leads together.

In addition to using various numbers of leads from ECG records, one of the selected articles reports making use of a complementary external source of information: a 9-lead signal recorded with a smartwatch.

3.3.9. RQ9: Are ECG Datasets Used to Train Deep Learning Models for the Detection of Myocardial Infarction Imbalanced? If So, How Do Research Efforts Tackle Class Imbalance?

Figure 15a shows that, out of the 47 selected articles, only 8 (17.02%) report having a balanced dataset in terms of the number of target classes, while 4 articles (8.51%) do not provide this information. Moreover, 35 articles (74.47%) had to deal with unbalanced datasets, which represents the most realistic scenario given the nature of the data.

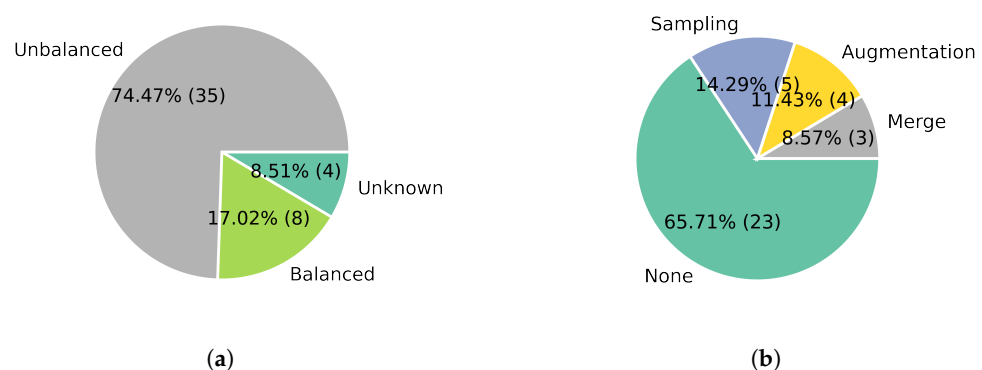


Figure 15. Target class balance within each dataset: (a) distribution of balanced and unbalanced datasets and (b) distribution of balancing methods.

Figure 15b shows the methods used by the authors of the selected articles to balance the datasets. Notably, the majority (65.71%) did not utilize any rebalancing method, while five articles used resampling, four applied data augmentation, and three merged records from different datasets.

3.3.10. RQ10: Do Works on Myocardial Infarction Detection Generate Synthetic Data for Training Their Models?

As shown in Figure 16, 65.96% of the works did not generate synthetic data to train their deep learning models, while the remaining works generated synthetic data using standard augmentation techniques (e.g., scaling ECG images).

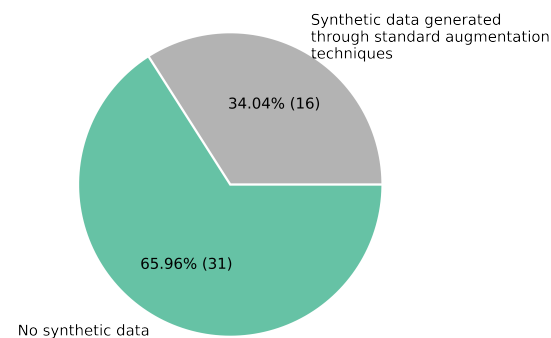


Figure 16. Proportion of deep learning models proposed in the selected articles whose training involved synthetic data generation.

3.3.11. RQ11: What Preprocessing Techniques Are Used for the Detection of Myocardial Infarction Supported by Deep Learning?

To train their deep models, most of the works preprocessed the ECG images using a myriad of techniques (Figure 17); in fact, only six selected articles trained their models using raw ECG images, i.e., without preprocessing them. Figure 17 lists the preprocessing techniques used and their frequencies.

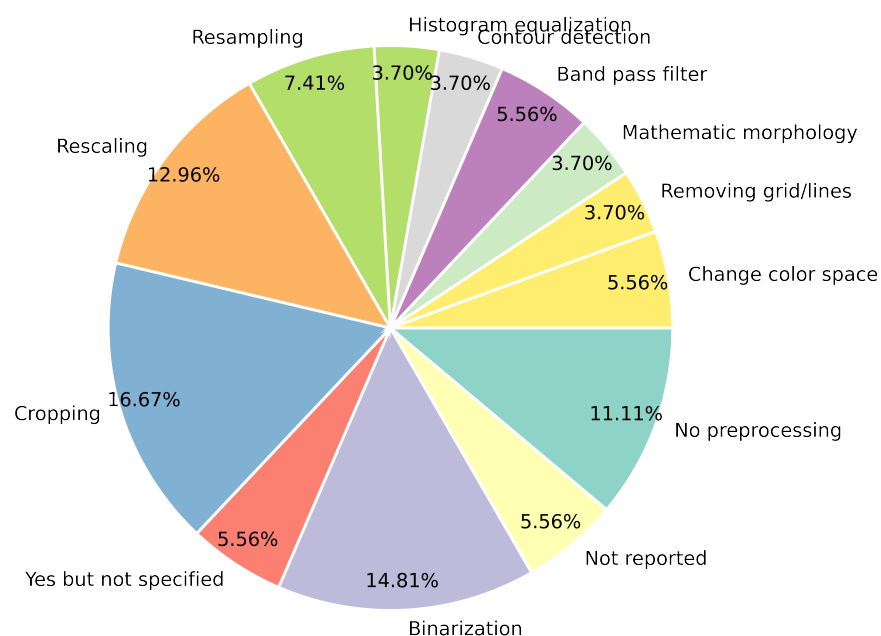


Figure 17. Preprocessing techniques used in the selected articles.

From the selected articles, it seems that the decision regarding which preprocessing technique must be applied depends on the state of the data and the expertise of the

researchers. Therefore, given the wide variety of source data and preprocessing approaches, it becomes rather difficult to make fair comparisons across their strategies and results.

3.3.12. RQ12: What Future Work Directions Are Proposed by Research Efforts Focused on Myocardial Infarction Detection Supported by Deep Learning?

As shown in Figure 18, most of the selected articles proposed future work directions related to refining their deep learning models by (i) collecting additional data to retrain them (the most frequent future work direction); (ii) further evaluating their models; (iii) performing hyperparameter optimization; and (iv) exploring other deep learning architectures or techniques. In addition, a relatively large percentage of the selected articles proposed extending the scope of the work by addressing other cardiovascular conditions (the second most frequent future work direction) or introducing more classes into the target variable. It is worth highlighting that the use of explainability techniques, that might be of relevance to the deployment of deep learning models in clinical settings, was listed as the fourth most common future work direction. See Figure 18 for the complete list of future work directions proposed by the authors of the selected articles.

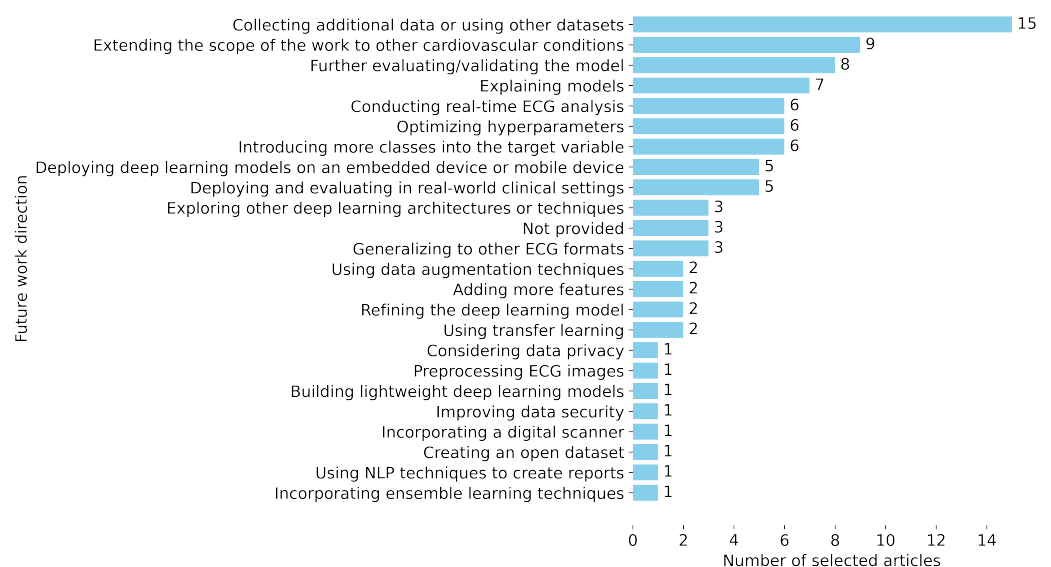


Figure 18. Future work directions proposed by the authors of the selected articles.

4. Discussion

The responses to the research questions presented in Section 2.1 shed light on the challenges that must be addressed to fully leverage deep learning for myocardial infarction detection.

The response to RQ1 shows that a wide range of deep learning architectures have been employed for myocardial infarction detection. CNN-based models predominate over Transformer- and RNN-based approaches, which were implemented only in a few studies. Consequently, due to the relatively small sample sizes and additional factors (e.g., the use of different datasets for training), there is insufficient evidence to draw statistically valid conclusions regarding neither performance differences across deep learning architectures nor correlations among them.

As discussed in the response to RQ2, most of the deep learning models presented in the selected articles relied on transfer learning followed by fine-tuning. This strategy likely reflects the relatively small size of the ECG datasets (Table 4) used in the majority of selected articles. However, even when 24 out of the 25 selected works that used transfer learning were pretrained using the ImageNet dataset, it remains difficult to assess the

extent to which fine-tuning pretrained models improve myocardial infarction classification when compared with ad hoc deep learning models trained from scratch. Furthermore, this widely used transfer learning approach adds to the discussion of whether, or not, CNN models that are pretrained on the ImageNet dataset are biased towards texture features observed during such pretraining procedure [96,97].

According to the response to RQ3, the models were predominantly trained to discriminate between two target class sets, namely, {normal, myocardial infarction} and {normal, myocardial infarction, abnormal heartbeats, history of myocardial infarction}, although several other class target sets were reported. This diversity in target class sets complicates fair performance comparisons among the deep learning models proposed in the selected articles. Nevertheless, it is expected that these models were designed to address specific research objectives; thus, the different target classes are justified. Additionally, as indicated by the responses to RQ4 and RQ5, only 28 of the 47 selected articles fully described their classification models, and source code was generally not provided. Moreover, a total of 15 different performance metrics were reported across the studies. This highlights the need for standardized benchmarking procedures and evaluation guidelines to facilitate comparison and support the safe deployment of deep learning models in clinical settings.

In relation to data organization and splitting, we observed that the hold-out approach is far more popular than cross-validation, with 72.3% (34 papers) and 21.3% (10 papers), respectively, and even having 3 papers with no details, as shown in Table 3. Having hold-out as the dominant approach suggests that the research community trusts the amount of data available for experimentation, which leads to having fixed splits of data. Although this is a desirable scenario, the selected works fail to show evidence about whether the distribution of variables holds across the train, validation, and test splits. Performing such an evaluation as part of the Exploratory Data Analysis (EDA) must become a standard practice in data mining and machine learning research. Furthermore, 17 out of the 34 papers that use hold-out limited the corresponding data split to only two subsets, leaving out the validation process. In addition, none of these 34 papers report specifics about the train-validation-test split other than percentages. Concretely, from the manuscripts, it is impossible to know which images belong to each subset, thus limiting reproducibility.

Regarding data imbalance, the response to RQ9 reveals that 35 of the 47 selected articles addressed imbalanced datasets, yet only 12 applied explicit balancing methods. Despite this, many studies reported high accuracy and sensitivity values (see the response to RQ6), while only a limited number used robust metrics suitable for imbalanced datasets, such as the area under the precision–recall curve or the area under the ROC curve. More importantly, these works do not report the impact of implementing balancing procedures on their datasets, which would help assess the performance of the deep learning models before varying condition of data, which seems to be of the highest relevance for such a health-related research area. These findings, together with the fact that most studies did not make their curated ECG image datasets publicly available, suggest the possibility of data leakage in some cases, for example, when ECG images from the same patient are included in both training and test sets.

Furthermore, as noted in the response to RQ7, two predominant datasets [85,86] were used across multiple studies. There seems to be two possible reasons for the high frequency of using these datasets: the fact that their records are already in image format and the fair balance among their classes. However, these datasets were compiled under the auspices of a Pakistani Institute of Cardiology, which may indicate that the ECG recordings correspond to a specific population and possibly to particular ECG recording devices. This raises concerns about potential population and device-related biases. In an attempt to verify this

hypothesis, the dataset descriptions in [85,86] were examined; however, no information regarding participants' nationality or the recording devices was reported.

The response to RQ8 indicates that most studies trained their models using 12-lead ECGs, while a smaller subset (eight articles) used between one and three leads. It should be mentioned that none of the selected articles evaluated the impact of systematically removing or adding ECG leads during training. This leaves open the research question of whether deep learning models can learn nonlinear transformations from a subset of leads while achieving performance comparable to models trained on full 12-lead ECGs.

In this regard, it seems also relevant to compare the use of ECG images vs ECG signals, i.e., leads. Although the selected articles focus merely on images, we anticipate that some of the advantages of using voltage signals include, fine resolutions for sampling and data quality, and control over visual variations observed in post-digitized images. On the other hand, using signals might have a few limitations, including the need for sophisticated devices not available in certain regions and the potential need for calibrating them, which might lead to possible biases in terms of the population. Nonetheless, since this review focuses on computer vision techniques, the aforementioned analysis remains an open question and out of scope in this present work.

In relation to RQ10, most articles did not generate synthetic data or relied solely on traditional data augmentation techniques, such as image scaling. In the presence of imbalanced datasets, in addition to conventional balancing methods (Figure 15), synthetic data generation could help mitigate class imbalance. Future research may benefit from synthetic ECG image generation approaches, such as the method proposed in [98], to improve myocardial infarction detection using deep learning.

With respect to data preprocessing techniques, the response to RQ11 indicates that rescaling, cropping, and binarization were commonly applied. However, the effectiveness of these preprocessing steps remains unclear, as the selected articles did not report comparative results with and without preprocessing.

Finally, as indicated in the response to RQ12, the authors of the selected articles outlined several future research directions, including the collection of additional data, further evaluation of models, and the incorporation of explainability techniques. In this regard, despite the relatively high performance measures reported (Figure 13), most authors emphasized the need for larger datasets to support robust model training. This entails additional experimentation and thorough evaluation prior to clinical deployment. Regarding explainability, a key implication is that model outputs should be interpreted by clinical specialists, as deep learning systems are intended to support decision making. Although five studies suggested clinical evaluation of their models, none provided a detailed methodology for conducting such evaluations in real-world clinical settings.

5. Future Research Directions

In addition to answering the proposed research questions, the analysis conducted in this systematic review has yielded a set of future research directions that could be explored to (i) advance research on deep learning models for myocardial infarction trained on ECG images and/or (ii) address existing research gaps. The proposed future research directions are as follows.

5.1. Quantifying Diagnostic Uncertainty

Deep learning is an emerging and growing field of practice for the detection of myocardial infarction and related medical conditions. Although deep learning models have shown promising results in diagnosing myocardial infarction (see Table 3), even comparable to those of domain experts (see [41,79]), their effectiveness is not fully guaranteed, i.e., their

classifications are subject to error. Moreover, traditional machine learning metrics such as precision, accuracy, or recall, while valid performance indicators, do not quantify the uncertainty in the predictions of a deep learning model. In this regard, the conformal prediction framework [99] quantifies the uncertainty of predictions by creating prediction sets containing the true diagnosis with a predefined level of confidence (e.g., 99%). Regardless of whether deep learning models are used solely as decision support tools, models for myocardial detection should be conformalized to quantify their uncertainty. More importantly, this quantification of diagnostic uncertainty must be disclosed to users. However, it should be mentioned that none of the selected articles made use of a framework to quantify uncertainty. Hence, a standard practice in the domain of deep learning models for the detection of myocardial infarction, should involve conformalizing the models to quantify the uncertainty of their predictions. Consequently, conformal prediction (or other uncertainty quantification methods) should be used to quantify the uncertainty of model predictions and to inform model training and evaluation. In addition, uncertainty metrics should be reported. Thus, a future research direction should be to assess to what extent these conformalized deep learning models for the detection of myocardial infarction can inform decision making.

5.2. Explaining Deep Learning Models

In general, deep learning models are not interpretable [100], i.e., models by themselves do not provide an explanation as to why a particular prediction was made [101] (unlike logistic regression models, where the relationship between input features and outcome is explicitly stated). Hence, a myriad of techniques have been developed to explain deep models. However, only a few of the selected articles analyzed in this systematic review made use of explainable machine learning techniques, and since the most predominant deep learning architecture was CNNs (see Figure 6), the majority used class activation maps (as in [62]) and its generalizations HiResCAM (as in [77]) and Grad-CAM (as in [40,55,56,76,79]). Class activation maps and its variants were designed for CNNs and create a heatmap highlighting what features (e.g., ECG image regions) were the most relevant to predict a given class. It is worth mentioning that only one of the selected articles [76] also made use of other explainable techniques (namely LIME, SHAP, and LRP-epsilon) to highlight the relevance of ECG segments in predicting a given class. However, within this application domain, the need for explanation of deep learning model predictions is of paramount importance because (i) from an ethical perspective, domain experts must be able to interpret and verify the reasons behind a given prediction; and (ii) from a legal perspective, artificial intelligence regulation frameworks such as the EU Artificial Intelligence Act [102] have incorporated individual rights to explanations for why an artificial intelligence model has made a particular decision. In addition, some of the authors of the selected articles (7 out of 47) have detected this need for explanation and listed as one of their future work directions explaining their deep learning models. Hence, a future research direction of this research area should be to develop a robust explainable deep learning framework in the context of myocardial infarction detection indicating what explainable techniques should be used to interpret and verify the reasons behind an outcome and comply with ethical and legal standards.

5.3. Improving the Evaluation of Deep Learning Models by Involving Domain Experts

Among the 47 selected articles, only 4 research efforts involved domain experts in evaluating their deep learning models, see [41,44,79,84]. In [41,79], the evaluation by domain experts (e.g., physicians and cardiologists) served as a baseline to compare the classification performance of CNNs for the detection of myocardial infarction (in [79]) and

acute coronary syndrome (in [41]). In [44], domain experts, namely cardiologists, reviewed ECGs misclassified by the deep learning model to confirm or refute its results. In [84], the authors evaluated the help provided by their deep learning model to 51 physicians in the interpretation of ECG for the detection of myocardial infarction. Either to serve as an evaluation baseline or to confirm or refute the model results, the evaluation of deep learning models for myocardial infarction must involve domain experts before deploying them in clinical settings. Moreover, a system for myocardial detection supported by deep learning is not assumed to be fully autonomous in diagnosing cardiovascular diseases, but to serve as a decision support system designed to assist domain experts such as physicians and cardiologists in the diagnosis. Hence, as in [84], domain experts should evaluate deep learning models for the detection of cardiovascular diseases in terms of the assistance they provide regardless of their performance as measured by machine learning metrics. For this reason, guidelines and metrics for the evaluation of deep learning models by domain experts should be developed in order to guarantee a safe deployment and use of systems supported by deep learning in the domain of myocardial infarction detection.

5.4. Developing Benchmarks and Guidelines for Reproducibility

Although most of the authors of the selected articles disclosed the ECG data sources used to train their deep learning models (Table 4), for a myriad of reasons, not all data sets are publicly available (see [56,84]). In addition, in general, when the ECG image datasets are available, the authors of the selected articles curated and pre-processed ECG images using multiple parametrized techniques (see Figure 17). In this regard, only [35,66,76] made publicly available their curated ECG image dataset. However, a significant share of the selected articles did not sufficiently describe their deep learning architectures or provide details on training hyperparameters, which complicates the reproducibility of model training. Moreover, only 2 (out of 47) selected articles provided source code, see [40,72]. Also, the authors of the selected articles reported different sets of performance metrics and in some cases disjoint sets across research efforts, see Figure 12 and Table 3. The evidence collected in this systematic review highlights the importance of a scientific effort toward a standardized protocol for reproducibility and definition of benchmarks that allow for a comparable and transparent evaluation of deep learning models for myocardial infarction detection.

6. Conclusions

Using a methodology based on PRISMA guidelines, this systematic review characterizes the current state of the art in myocardial infarction detection with deep learning models trained on ECG images. This characterization can introduce both researchers and practitioners to the application of deep learning in this domain, while experienced researchers may benefit from the proposed future directions to define new projects.

The novelty of this work derives from the systematic analysis, guided by a set of 12 research questions, of recent and relevant scientific articles on myocardial infarction detection using deep learning models trained on ECG images. In addition, to the best of the authors' knowledge, this is the first systematic review to explore the use of vision transformers for myocardial infarction detection.

The significance of this work stems from (i) the characterization of deep learning architectures for myocardial infarction detection; (ii) the assessment of training and evaluation practices for deep learning models trained on ECG images; (iii) the identification of macro-level performance measures achieved by selected deep learning models; (iv) the compilation of available ECG image datasets for future research; and (v) the proposal of future research directions aimed at advancing deep learning approaches and addressing existing gaps in automated myocardial infarction detection. The proposed future directions

include: (i) quantifying diagnostic uncertainty using conformal prediction; (ii) explaining deep learning models through explainable machine learning techniques; (iii) improving evaluation practices by involving domain experts; and (iv) developing benchmarks and guidelines for reproducibility.

From the responses to the research questions posed in this systematic review, it can be concluded that customized, ad hoc CNNs are the most commonly used neural network architectures for detecting myocardial infarction from ECG images. However, well-known architectures such as EfficientNet and ResNet were also fine-tuned to improve myocardial infarction detection. To train or fine-tune these deep learning models, the majority of the selected articles relied on 12-lead ECGs; in this regard, the ECG image dataset compiled by Khan et al., (2021) [85] was the most popular. It should be noted that most of the selected articles preprocessed ECG images using a wide range of techniques, from binarization to contour detection. Additionally, some models were trained with synthetic data, primarily generated using standard augmentation techniques such as image scaling. Moreover, most of the deep learning models reported in the selected articles classified ECG images into four classes: (i) normal, (ii) myocardial infarction, (iii) abnormal heartbeats, and (iv) history of myocardial infarction. Based on this common target class set and due to the nature of the data in this application domain, most articles dealt with imbalanced datasets when training their models. However, a relatively large share of works did not apply any balancing methods. In fact, while most articles reported common performance metrics such as accuracy, sensitivity, precision, F1 score, specificity, and area under the ROC curve, only a few made use of the area under the precision–recall curve, a robust metric for imbalanced datasets. Regarding these metrics, the median macro measures exceeded 0.95, suggesting relatively outstanding performance for a large share of the reported models. With respect to reproducibility, although most articles provided complete descriptions of their models, a non-negligible proportion omitted details or offered only partial descriptions. As a future work direction, most authors of the selected articles agreed that their models should be trained on larger datasets.

From the proposed future directions, it can be concluded that the safe and successful deployment of deep learning models for myocardial infarction in clinical settings requires the involvement of domain experts in the evaluation and validation of the models, in addition to the quantification of their diagnostic uncertainty. Explainable machine learning techniques should be employed to inform domain experts why, for example, an ECG is classified as normal or abnormal. Consequently, the involvement of domain experts throughout the training, evaluation, and deployment phases of deep learning models is vital. However, the results of this systematic review indicate that only a limited number of studies have adopted this practice.

It is acknowledged that there is a possibility of selection bias in the selected articles. Specifically, the search string included terms related to deep learning and myocardial infarction detection and was applied to the title, abstract, and keywords of articles indexed in scientific databases; however, cardiology-focused articles may not have included deep learning-related terms in their titles, abstracts, or keywords. As a consequence, such studies may have been excluded from this systematic review, whose domain focuses on the interaction between deep learning and myocardial infarction detection. It is also acknowledged that this review focuses exclusively on deep learning models trained on ECG images. However, relevant research also exists on models trained using other features or ECG signals rather than images. Future work will focus on analyzing and characterizing these efforts to complete the systematic assessment of the state of the art and contrast those findings with the present results. An additional future research direction is to conduct a meta-analysis to

assess the heterogeneity of deep learning approaches for myocardial infarction detection and to analyze variations in their performance across specific populations.

Author Contributions: Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, visualization, project administration: J.O.G.-G., E.R.-R. and J.M.R.-M. All authors have read and agreed to the published version of the manuscript.

Funding: J. O. Gutierrez-Garcia and E. Roman-Rangel gratefully acknowledge the financial support from the Asociación Mexicana de Cultura, A.C. (Grant No. NA).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

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

Abbreviations

The following abbreviations are used in this manuscript:

ACS	Acute coronary syndrome
AMI	Acute myocardial infarction
CNN	Convolutional neural network
ECG	Electrocardiogram
EKG	Electrocardiogram
MI	Myocardial infarction
MRI	Magnetic resonance imaging
RNN	Recurrent neural network
ST	Segment
STE	ST elevation
STEMI	ST-elevation myocardial infarction
ViT	Vision transformer

References

1. Ren, J.; Chen, X.; Wang, T.; Liu, C.; Wang, K. Regenerative therapies for myocardial infarction: Exploring the critical role of energy metabolism in achieving cardiac repair. *Front. Cardiovasc. Med.* **2025**, *12*, 1533105. [\[CrossRef\]](#)
2. Victor, G.; Shishani, K.; Vellone, E.; Froelicher, E.S. The Global Burden of Cardiovascular Disease in Adults: A Mapping Review. *J. Cardiovasc. Nurs.* **2024**, *40*, 523–537. [\[CrossRef\]](#)
3. Tsutsui, K.; Brimer, S.B.; Ben-Moshe, N.; Sellal, J.M.; Oster, J.; Mori, H.; Ikeda, Y.; Arai, T.; Nakano, S.; Kato, R.; et al. SHDB-AF: A Japanese Holter ECG database of atrial fibrillation. *Sci. Data* **2025**, *12*, 454. [\[CrossRef\]](#)
4. Kotsialou, Z.; Makris, N.; Gall, S. Fundamentals of the electrocardiogram and common cardiac arrhythmias. *Anaesth. Intensive Care Med.* **2024**, *25*, 219–222. [\[CrossRef\]](#)
5. Attar, E.T. ECG interpretation abilities in clinical practice: Examining the role of expertise, age, and gender. *Medicine* **2025**, *104*, e42401. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Kim, J.H.; Cisneros, T.; Nguyen, A.; van Meijgaard, J.; Warraich, H.J. Geographic disparities in access to cardiologists in the United States. *J. Am. Coll. Cardiol.* **2024**, *84*, 315–316. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Zang, J.; An, Q.; Li, B.; Zhang, Z.; Gao, L.; Xue, C. A novel wearable device integrating ECG and PCG for cardiac health monitoring. *Microsyst. Nanoeng.* **2025**, *11*, 7. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Musa, N.; Gital, A.Y.; Aljojo, N.; Chiroma, H.; Adewole, K.S.; Mojeed, H.A.; Faruk, N.; Abdulkarim, A.; Emmanuel, I.; Folawiyo, Y.Y.; et al. A systematic review and Meta-data analysis on the applications of Deep Learning in Electrocardiogram. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 9677–9750. [\[CrossRef\]](#)
9. Radwa, E.; Ridha, H.; Faycal, B. Deep learning-based approaches for myocardial infarction detection: A comprehensive review recent advances and emerging challenges. *Med. Nov. Technol. Devices* **2024**, *23*, 100322. [\[CrossRef\]](#)
10. Sumalatha, U.; Prakasha, K.K.; Prabhu, S.; Nayak, V.C. Deep learning applications in ecg analysis and disease detection: An investigation study of recent advances. *IEEE Access* **2024**, *12*, 126258–126284. [\[CrossRef\]](#)

11. Han, C.; Zhou, Y.; Que, W.; Li, Z.; Shi, L. An overview of algorithms for myocardial infarction diagnostics using ecg signals: Advances and challenges. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2522713. [[CrossRef](#)]
12. Elmassaoudi, A.; Douzi, S.; Abik, M. Machine Learning Approaches for Automated Diagnosis of Cardiovascular Diseases: A Review of Electrocardiogram Data Applications. *Cardiol. Rev.* **2024**, 10-1097. [[CrossRef](#)]
13. Fang, Y.; Wu, Y.; Gao, L. Machine learning-based myocardial infarction bibliometric analysis. *Front. Med.* **2025**, *12*, 1477351. [[CrossRef](#)] [[PubMed](#)]
14. Handra, J.; James, H.; Mbilinyi, A.; Moller-Hansen, A.; O'Riley, C.; Andrade, J.; Deyell, M.; Hague, C.; Hawkins, N.; Ho, K.; et al. The Role of Machine Learning in the Detection of Cardiac Fibrosis in Electrocardiograms: Scoping Review. *JMIR Cardio* **2024**, *8*, e60697. [[CrossRef](#)]
15. Elvas, L.B.; Almeida, A.; Ferreira, J.C. The Role of AI in Cardiovascular Event Monitoring and Early Detection: Scoping Literature Review. *JMIR Med. Inform.* **2025**, *13*, e64349. [[CrossRef](#)]
16. Akouz, N.; El Ghazi, A.; Moutaouakil, W.; Hamida, S.; Cherradi, B.; Raihani, A. Predicting Cardiovascular Disease: A Scoping Survey on different Datasets and DL/ML Models using ECG. In Proceedings of the 2024 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 8–10 May 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
17. 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](#)]
18. Khalid, M.; Pluempitiwiriawej, C.; Wangsiripitak, S.; Murtaza, G.; Abdulkadhem, A.A. The applications of deep learning in ECG classification for disease diagnosis: A systematic review and meta-data analysis. *Eng. J.* **2024**, *28*, 45–77. [[CrossRef](#)]
19. Oke, O.A.; Cavus, N. A systematic review on the impact of artificial intelligence on electrocardiograms in cardiology. *Int. J. Med. Inform.* **2025**, *195*, 105753. [[CrossRef](#)]
20. Zworth, M.; Kareemi, H.; Boroumand, S.; Sikora, L.; Stiell, I.; Yadav, K. Machine learning for the diagnosis of acute coronary syndrome using a 12-lead ECG: A systematic review. *Can. J. Emerg. Med.* **2023**, *25*, 818–827. [[CrossRef](#)] [[PubMed](#)]
21. Wu, Z.; Guo, C. Deep learning and electrocardiography: Systematic review of current techniques in cardiovascular disease diagnosis and management. *BioMed. Eng. OnLine* **2025**, *24*, 23. [[CrossRef](#)]
22. Fuadah, Y.N.; Lim, K.M. Advances in cardiovascular signal analysis with future directions: A review of machine learning and deep learning models for cardiovascular disease classification based on ECG, PCG, and PPG signals. *Biomed. Eng. Lett.* **2025**, *15*, 619–660. [[CrossRef](#)] [[PubMed](#)]
23. Pandey, V.; Lilhore, U.K.; Walia, R. A systematic review on cardiovascular disease detection and classification. *Biomed. Signal Process. Control* **2025**, *102*, 107329. [[CrossRef](#)]
24. Vásquez-Iturralde, F.; Flores-Calero, M.J.; Grijalva, F.; Rosales-Acosta, A. Automatic classification of cardiac arrhythmias using deep learning techniques: A systematic review. *IEEE Access* **2024**, *12*, 118467–118492. [[CrossRef](#)]
25. Wang, R.; Veera, S.C.M.; Asan, O.; Liao, T. A systematic review on the use of consumer-based ECG wearables on cardiac health monitoring. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 6525–6537. [[CrossRef](#)]
26. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ* **2009**, *339*, b2700. [[CrossRef](#)] [[PubMed](#)]
27. Ed-Dafali, S.; Adardour, Z.; Derj, A.; Bami, A.; Hussainey, K. A PRISMA-Based Systematic Review on Economic, Social, and Governance Practices: Insights and Research Agenda. *Bus. Strategy Environ.* **2025**, *34*, 1896–1916. [[CrossRef](#)]
28. Gutierrez-Garcia, J.O.; Roman-Rangel, E.; Rendon-Mancha, J.M. Deep Learning for Myocardial Infarction Detection Using Electrocardiogram Images: A Systematic Review. *MetaArXiv* **2026**. [[CrossRef](#)]
29. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; PMLR 2019; pp. 6105–6114.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
31. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2261–2269.
32. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
33. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–9.

34. Gadag, V.; Singh, S.; Khatri, A.H.; Mishra, S.; Satapathy, S.K.; Cho, S.B.; Chowdhury, A.; Pal, A.; Mohanty, S.N. Improving myocardial infarction diagnosis with Siamese network-based ECG analysis. *PLoS ONE* **2025**, *20*, e0313390. [\[CrossRef\]](#)
35. Alsayat, A.; Mahmoud, A.A.; Alanazi, S.; Mostafa, A.M.; Alshammari, N.; Alrowaily, M.A.; Shabana, H.; Ezz, M. Enhancing cardiac diagnostics: A deep learning ensemble approach for precise ECG image classification. *J. Big Data* **2025**, *12*, 7. [\[CrossRef\]](#)
36. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 248–255.
37. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312. [\[CrossRef\]](#)
38. Jeni, L.A.; Cohn, J.F.; De La Torre, F. Facing imbalanced data—recommendations for the use of performance metrics. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 245–251.
39. McDermott, M.; Zhang, H.; Hansen, L.; Angelotti, G.; Gallifant, J. A closer look at auROC and aupRC under class imbalance. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 44102–44163.
40. Vaid, A.; Jiang, J.; Sawant, A.; Lerakis, S.; Argulian, E.; Ahuja, Y.; Lampert, J.; Charney, A.; Greenspan, H.; Narula, J.; et al. A foundational vision transformer improves diagnostic performance for electrocardiograms. *NPJ Digit. Med.* **2023**, *6*, 108. [\[CrossRef\]](#)
41. Choi, J.; Kim, J.; Spaccarotella, C.; Esposito, G.; Oh, I.Y.; Cho, Y.; Indolfi, C. Smartwatch ECG and artificial intelligence in detecting acute coronary syndrome compared to traditional 12-lead ECG. *IJC Heart Vasc.* **2025**, *56*, 101573. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Yang, Z.; Jin, A.; Li, Y.; Yu, X.; Xu, X.; Wang, J.; Li, Q.; Guo, X.; Liu, Y. A coordinated adaptive multiscale enhanced spatio-temporal fusion network for multi-lead electrocardiogram arrhythmia detection. *Sci. Rep.* **2024**, *14*, 20828. [\[CrossRef\]](#)
43. Hao, P.; Yin, X.; Wu, F.; Zhang, F. A Novel Feature Fusion Network for Myocardial Infarction Screening Based on ECG Images. In Proceedings of the International Conference on Image and Graphics, Haikou, China, 6–8 August 2021; pp. 547–558.
44. Chandra, B.; Singh, K.P.; Kalra, P.; Narang, R. Automatic diagnosis of 12-lead ECG using DINOv2. *Artif. Intell. Mach. Learn. Convolutional Neural Netw. Large Lang. Model.* **2024**, *1*, 255.
45. Jaya Mabel Rani, A.; Srivenkateswaran, C.; Vishnupriya, G.; Subramanian, N.; Ilango, P.; Jacintha, V.K. A big data scheme for heart disease classification in map reduce using jellyfish search flow regime optimization enabled Spinalnet. *Pacing Clin. Electrophysiol.* **2024**, *47*, 953–965. [\[CrossRef\]](#)
46. Xiao, R.; Xu, Y.; Pelter, M.M.; Mortara, D.W.; Hu, X. A deep learning approach to examine ischemic ST changes in ambulatory ECG recordings. *AMIA Summits Transl. Sci. Proc.* **2018**, *2018*, 256.
47. Srinivasulu, B.; Reddy, P.S.; Basha, P.H. A Deep Pattern Learning based Model for Detection of Cardiovascular Diseases (CVD). In Proceedings of the 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 3–4 May 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 191–196.
48. Aggarwal, R.; Kumar, S. A hybrid detection model for meticulous presaging of heart disease using deep learning: HDMPHD. *Int. J. Recent Innov. Trends Comput. Commun.* **2022**, *10*, 67–76. [\[CrossRef\]](#)
49. Kiran, A.; Unhelkar, B.; Shankar, S.S.; Chakrabarti, T.; Chakrabarti, P.; Sivaneasan, B.; Margala, M. A hybrid fine-tuned optimizer for enhancing ECG data security in heart attack detection systems. *J. Inf. Optim. Sci.* **2024**, *45*, 2309–2323. [\[CrossRef\]](#)
50. Rana, A.; Kim, K.K. A lightweight dnn for ecg image classification. In Proceedings of the 2020 International SoC Design Conference (ISOCC), Yeosu, Republic of Korea, 21–24 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 328–329.
51. Naidji, M.R.; Elberichi, Z. A novel hybrid vision transformer CNN for COVID-19 detection from ECG images. *Computers* **2024**, *13*, 109. [\[CrossRef\]](#)
52. Manimaran, V.; Shanthi, N.; Aravindhraj, N.; Aatarsh, K.; Adharshini, G.; Gokul, P. Advancements in heart disease classification: Leveraging deep learning techniques for ECG analysis. In Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 24–28 June 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–7.
53. Singh, K.P.; Chandra, B.; Kalra, P.K.; Narang, R. Amazing power of dinov2 for automatic diagnosis of 12-lead ecg. In Proceedings of the 2023 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 13–15 December 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1386–1391.
54. Hasan, M.N.; Hossain, M.A.; Rahman, M.A. An ensemble based lightweight deep learning model for the prediction of cardiovascular diseases from electrocardiogram images. *Eng. Appl. Artif. Intell.* **2025**, *141*, 109782. [\[CrossRef\]](#)
55. Denaro, F.; Madau, A.; Martini, C.; Pecori, R. An Explainable Approach to Characterize Heart Diseases Using ECG Images. In Proceedings of the 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE), St Albans, UK, 24 December 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 867–872.
56. Kavak, S.; Chiu, X.D.; Yen, S.J.; Chen, M.Y.C. Application of CNN for detection and localization of STEMI using 12-lead ECG images. *IEEE Access* **2022**, *10*, 38923–38930. [\[CrossRef\]](#)

57. Mhamdi, L.; Dammak, O.; Cottin, F.; Dhaou, I.B. Artificial intelligence for cardiac diseases diagnosis and prediction using ECG images on embedded systems. *Biomedicines* **2022**, *10*, 2013. [\[CrossRef\]](#)
58. Rout, M.; Nayak, S.C.; Rai, S.C. Automated Cardiovascular Disease Detection from ECG Images Using Deep Learning. In Proceedings of the 2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT), Bhubaneswar, India, 21–23 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
59. Gliner, V.; Keidar, N.; Makarov, V.; Avetisyan, A.I.; Schuster, A.; Yaniv, Y. Automatic classification of healthy and disease conditions from images or digital standard 12-lead electrocardiograms. *Sci. Rep.* **2020**, *10*, 16331. [\[CrossRef\]](#)
60. Priya, R.K.; Alias, L.; Al Salehiya, F.S.S. Cardiac Health Assessment through Advanced Computational Models for ECG Image Analysis. In Proceedings of the 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 12–14 April 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 283–288.
61. Akkuzu, N.; Ucan, M.; Kaya, M. Classification of Multi-Label Electrocardiograms Utilizing the EfficientNet CNN Model. In Proceedings of the 2023 4th International Conference on Data Analytics for Business and Industry (ICDABI), Manama, Bahrain, 25–26 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 268–272.
62. Oikawa, R.; Doi, A.; Chakraborty, B.; Itoh, T.; Nishiyama, O. Classification of prehospital-electrocardiograms taken in ambulance according to severity using deep learning neural network. In Proceedings of the 2022 IEEE 4th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), Tainan, Taiwan, 27–29 May 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 263–266.
63. Akula, C.K.; Mondal, S.; Manyam, R.; Akkineni, H.C.N.; Hemanth, B.; Appikatl, V.T.S.R.K. Customized CNN Architecture for ECG Image-Based Classification of Cardiovascular Diseases. In Proceedings of the 2024 First International Conference on Software, Systems and Information Technology (SSITCON), Tumkur, India, 18–19 October 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–7.
64. Kurian, T.; Thangam, S. Deep convolution neural network-based classification and diagnosis of heart disease using ElectroCardioGram (ECG) images. In Proceedings of the 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 7–9 April 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
65. Parvathi, R.; Pavithra, S.; Pattabiraman, V. Deep Learning Approach on Multimodal Data for Myocardial Infarction Prediction. In Proceedings of the 2024 International Conference on Computational Intelligence and Network Systems (CINS), Dubai, United Arab Emirates, 28–29 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–8.
66. Albasrawi, R.; Ilyas, M. Detecting myocardial infraction in ECG waveforms using YOLOv8. In Proceedings of the 2024 Global Digital Health Knowledge Exchange & Empowerment Conference (gDigiHealth. KEE), Abu Dhabi, United Arab Emirates, 24–26 September 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
67. 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. [\[CrossRef\]](#)
68. Alghamdi, A.; Hammad, M.; Ugail, H.; Abdel-Raheem, A.; Muhammad, K.; Khalifa, H.S.; Abd El-Latif, A.A. Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *Multimed. Tools Appl.* **2024**, *83*, 14913–14934. [\[CrossRef\]](#)
69. Amrutesh, A.; KP, A.R.; S, G. ECG image analysis for medical issue detection using deep transfer learning techniques. In Proceedings of the 2023 14th international conference on computing communication and networking technologies (ICCCNT), Delhi, India, 6–8 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7.
70. Wasimuddin, M.; Elleithy, K.; Abuzneid, A.; Faezipour, M.; Abuzagheh, O. ECG signal analysis using 2-D image classification with convolutional neural network. In Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 949–954.
71. Khalid, M.; Pluempitiwiriyaew, C.; Abdulkadhem, A.A.; Afzal, I.; Truong, T. ECGConVT: A Hybrid CNN and Vision Transformer Model for Enhanced 12-Lead ECG Images Classification. *IEEE Access* **2024**, *12*, 193043–193056. [\[CrossRef\]](#)
72. Anwar, T.; Zakir, S. Effect of image augmentation on ECG image classification using deep learning. In Proceedings of the 2021 International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, 5–7 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 182–186.
73. Sadad, T.; Safran, M.; Khan, I.; Alfarhood, S.; Khan, R.; Ashraf, I. Efficient classification of ECG images using a lightweight CNN with attention module and IoT. *Sensors* **2023**, *23*, 7697. [\[CrossRef\]](#)
74. Uchiyama, R.; Okada, Y.; Kakizaki, R.; Tomioka, S. End-to-end convolutional neural network model to detect and localize myocardial infarction using 12-Lead ECG images without preprocessing. *Bioengineering* **2022**, *9*, 430. [\[CrossRef\]](#)
75. Setiawan, A.W. Evaluation Performance of ECG Leads in Myocardial Infarction Classification Using Deep Learning. In Proceedings of the 2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom), Nara, Japan, 18–20 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.

76. Knof, H.; Bagave, P.; Boerger, M.; Tcholtchev, N.; Ding, A.Y. Exploring CNN and XAI-based approaches for accountable mi detection in the context of IOT-enabled emergency communication systems. In Proceedings of the 13th International Conference on the Internet of Things, Nagoya, Japan, 7–10 November 2023; ACM: New York, NY, USA, 2023; pp. 50–57.
77. Bellfield, R.A.; Ortega-Martorell, S.; Lip, G.Y.; Oxborough, D.; Olier, I. Impact of ECG data format on the performance of machine learning models for the prediction of myocardial infarction. *J. Electrocardiol.* **2024**, *84*, 17–26. [\[CrossRef\]](#)
78. Wasimuddin, M.; Elleithy, K.; Abuzneid, A.; Faezipour, M.; Abuzagheh, O. Multiclass ECG signal analysis using global average-based 2-D convolutional neural network modeling. *Electronics* **2021**, *10*, 170. [\[CrossRef\]](#)
79. Makimoto, H.; Höckmann, M.; Lin, T.; Glöckner, D.; Gerguri, S.; Clasen, L.; Schmidt, J.; Assadi-Schmidt, A.; Bejinariu, A.; Müller, P.; et al. Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction. *Sci. Rep.* **2020**, *10*, 8445. [\[CrossRef\]](#)
80. Khan, M.A.B.A.; Reddy, E.S. Post-COVID effect on heart after recovery based on hybrid EfficientNet-DBN with multilevel classification using ECG images. *EngMedicine* **2024**, *1*, 100021. [\[CrossRef\]](#)
81. Bharathi, R.; Neelima, E. Predicting Heart Attacks with Precision: Harnessing ECG Signals for Early Detection. *Math. Model. Eng. Probl.* **2024**, *11*, 3181. [\[CrossRef\]](#)
82. Panchal, N.; Raikar, M.M.; Baligar, V.P. Prediction of Cardiac Severity Based on ECG Images Using Deep Learning Models. In Proceedings of the 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 24–27 July 2024; IEEE: Piscataway, NJ, USA, 2024; Volume 1, pp. 1–5.
83. Mahmoud, S.; Gaber, M.; Farouk, G.; Keshk, A. Prediction of heart disease using new proposed CNN model architecture. In Proceedings of the 2023 3rd International Conference on Electronic Engineering (ICEEM), Menouf, Egypt, 7–8 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–8.
84. Park, B.E.; Shon, B.; Cho, J.; Jung, M.S.; Park, J.S.; Kim, M.S.; Lee, E.; Choi, H.; Park, H.K.; Park, Y.J.; et al. Signal-guided multitask learning for myocardial infarction classification using images of electrocardiogram. *Cardiology* **2025**, *150*, 347–356. [\[CrossRef\]](#)
85. Khan, A.H.; Hussain, M. ECG Images dataset of Cardiac Patients. Dataset. *Mendeley Data* **2021**. [\[CrossRef\]](#)
86. Khan, A.H.; Hussain, M.; Malik, M.K. ECG Images dataset of Cardiac and COVID-19 Patients. *Data Brief* **2021**, *34*, 106762. [\[CrossRef\]](#)
87. Bousseljot, R.; Kreiseler, D.; Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomed. Tech.* **1995**, *40*, 317–318. [\[CrossRef\]](#)
88. Wagner, P.; Strodthoff, N.; Bousseljot, R.D.; Samek, W.; Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *PhysioNet* **2022**, *7*, 1–15. [\[CrossRef\]](#)
89. Liu, F.; Liu, C.; Zhao, L.; Zhang, X.; Wu, X.; Xu, X.; Liu, Y.; Ma, C.; Wei, S.; He, Z.; et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. *J. Med. Imaging Health Inform.* **2018**, *8*, 1368–1373. [\[CrossRef\]](#)
90. Ribeiro, A.H.; Paixao, G.M.; Lima, E.M.; Horta Ribeiro, M.; Pinto Filho, M.M.; Gomes, P.R.; Oliveira, D.M.; Meira, W., Jr.; Schon, T.B.; Ribeiro, A.L.P. CODE-15%: A large scale annotated dataset of 12-lead ECGs. *Zenodo* **2021**. [\[CrossRef\]](#)
91. Taddei, A.; Distanti, G.; Emdin, M.; Pisani, P.; Moody, G.B.; Zeelenberg, C.; Marchesi, C. The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *Eur. Heart J.* **1992**, *13*, 1164–1172. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Jager, F.; Taddei, A.; Moody, G.B.; Emdin, M.; Antolic, G.; Dorn, R.; Smrdel, A.; Marchesi, C.; Mark, R.G. Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. *Med. Biol. Eng. Comput.* **2003**, *41*, 172–182. [\[CrossRef\]](#)
93. Choi, S.H.; Lee, H.G.; Park, S.D.; Bae, J.W.; Lee, W.; Kim, M.S.; Kim, T.H.; Lee, W.K. Electrocardiogram-based deep learning algorithm for the screening of obstructive coronary artery disease. *BMC Cardiovasc. Disord.* **2023**, *23*, 287. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Janosi, A.; Steinbrunn, W.; Pfisterer, M.; Detrano, R. Heart Disease. *UCI Mach. Learn. Repos.* **1989**. [\[CrossRef\]](#)
95. Moody, G.; Mark, R. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [\[CrossRef\]](#)
96. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.A.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019; pp. 1–22.
97. Burgert, T.; Stoll, O.; Rota, P.; Demir, B. ImageNet-trained CNNs are not biased towards texture: Revisiting feature reliance through controlled suppression. In Proceedings of the Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS), San Diego, CA, USA, 2–7 December 2025; pp. 1–32.
98. Shivashankara, K.K.; Deepanshi; Shervedani, A.M.; Clifford, G.D.; Reyna, M.A.; Sameni, R. ECG-Image-Kit: A synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization. *Physiol. Meas.* **2024**, *45*, 055019. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Zhou, X.; Chen, B.; Gui, Y.; Cheng, L. Conformal prediction: A data perspective. *ACM Comput. Surv.* **2025**, *58*, 49. [\[CrossRef\]](#)

100. Sánchez Fernández, I.; Peters, J.M. Machine learning and deep learning in medicine and neuroimaging. *Ann. Child Neurol. Soc.* **2023**, *1*, 102–122. [[CrossRef](#)]
101. Lisboa, P.J.; Saralajew, S.; Vellido, A.; Fernández-Domenech, R.; Villmann, T. The coming of age of interpretable and explainable machine learning models. *Neurocomputing* **2023**, *535*, 25–39. [[CrossRef](#)]
102. Future of Life Institute. EU Artificial Intelligence Act: Up-to-Date Developments and Analyses. 2025. Available online: <https://artificialintelligenceact.eu/> (accessed on 7 October 2025).

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.