ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 6 Issue 2 April to June 2026
Paper Id: JETA-V6I2P111/ Page No: 96-105

Review Article

Automated Ecg Classification Using Transfer Learning From Vision Models To Medical Signal Domains

FNU Sudhakar Abhijeet
Northeastern University, Boston

Received Date: 20 March 2026 Revised Date: 28 March 2026 Accepted Date: 17 April 2026

Abstract: *Electrocardiogram (ECG) classification is important for initial detection and diagnosis of cardiovascular diseases, but it remains challenging without human interpretation, as the signals vary and labeled datasets are limited. The latest developments in deep learning have shown excellent performance, but established models still tend to require large amounts of domain-specific training data. The review examines the new paradigm of leveraging transfer learning from vision-based models, such as convolutional neural networks and transformer models, for ECG classification tasks. By converting one-dimensional ECGs into two-dimensional images such as spectrograms and scalograms, vision models that have been trained can be successfully customized to leverage the powerful attributes of medical signals. The paper critically examines the strategies of representation, transfer learning methods, and transfer learning benchmark databases, and highlights performance gains and generalization analyses. Also, major issues such as domain mismatch, interpretability, and constraints are addressed. The review concludes with research directions for the future, including self-supervised learning, multimodal integration, and real-time clinical applications, as well as the possibility of cross-domain knowledge transfer in the development of intelligent healthcare systems.*

Keywords: *ECG Classification, Transfer Learning, Vision Transformers, Convolutional Neural Networks, Biomedical Signal Processing, Domain Adaptation*

I. INTRODUCTION

The analysis of medical signals and images is an essential component of modern healthcare systems, helping clinicians interpret complex physiological and anatomical data. Electrocardiograms (ECG), magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound are modalities that offer complementary approaches to human health and are high-dimensional, noisy, and variable. In ECG analysis, the signal is a time series, and the clinically important features include the P wave, QRS complex, and T wave, which represent cardiac electrical activity. On the other hand, spatial structures, textures, and intensity variations associated with pathological states are encoded in medical imaging modalities. The classical diagnostic algorithms are based on domain-knowledge procedures, such as signal preprocessing, noise reduction, signal segmentation, and manual feature extraction, followed by a statistical/rule-based classification step. Although these methods have been used to aid clinical decision-making, they are often unable to capture the complex, non-linear relationships present in large-scale biomedical data. Pattern recognition has therefore become a major paradigm for the discovery of latent structures and correlations in heterogeneous datasets.

Over the past few years, this field has been transforming with artificial intelligence (AI), especially deep learning, which enables end-to-end learning schemes that automatically derive hierarchical representations of raw inputs. Convolutional architectures and deep neural networks have proven very successful at capturing temporal and spatial dependencies, greatly improving diagnostic accuracy and efficiency and reducing the need for manual feature engineering [1]-[3].Despite these achievements, significant issues remain in clinical practice regarding the manual interpretation of medical signals and images. The diagnostic evaluation process is both time-consuming and requires extensive training and experience, predisposing it to fatigue and cognitive bias. Also, inter-observer and intra-observer variability have always been a problem, and in some cases, they can lead to inconsistent diagnoses, especially when dealing with subtle abnormalities or vague patterns [4].

These limitations are exacerbated by the growing volume of clinical and medical information generated by advanced imaging technology and continuous monitoring systems, which impedes clinical processes and delays decision-making. Deep learning has therefore presented an innovative solution in this regard, providing automated, scalable, and repeatable analysis solutions. Due to their ability to process large volumes of data and complex architectures, deep learning systems have demonstrated performance equal to or superior to that of expert clinicians on tasks such as arrhythmia detection, tumor classification, and disease prediction [5]. Additionally, such models learn discriminative, task-specific features directly from the data, meaning they do not require handcrafted feature design and can generalize better across patient populations and clinical environments [6].

This has led to an increasing use of deep learning in computer-aided diagnosis systems to improve efficiency and reliability. One of the main challenges for deep learning applications in the medical field is the scarcity of large, well-labeled datasets. Medical data annotation needs specialized clinical knowledge, a lot of time, and strict control over regulations and ethics, whereas natural image datasets need none of these (as opposed to natural image datasets). The limited number of labeled data can easily cause overfitting during the training of deep models, directly through initialization, thereby limiting their generalization ability.

Transfer learning has become one of the most effective approaches to overcoming these limitations by leveraging knowledge learned by models trained on large-scale datasets, such as natural images. Transfer learning enables the manipulation of learned representations for task-specific medical tasks (e.g., fine-tuning and feature extraction), achieving high performance even with sparse training data [7]. Transfer learning is also useful in domain adaptation, enabling models to extrapolate to changes in data distributions, acquisition devices, and patient demographics [8]. As a result, it has become a staple practice in current medical AI, filling the data sparsity/deep neural network capacity gap.

The combination of vision-based deep learning systems, especially convolutional neural networks (CNNs) and Vision Transformers (ViTs), has further advanced automated medical processing. Such models are optimally constructed to represent spatial hierarchies and long-range dependence, which have been found to be very useful in medical imaging and transformed signal representation. In ECG analysis, a one-dimensional signal can be transformed into a two-dimensional signal, e.g., a spectrogram or scalogram, enabling the use of vision models for feature extraction and classification tasks [9], [10]. CNNs are very good at refining local patterns and textures with the help of hierarchical feature maps, whereas ViTs use self-attention to form relationships with contextual information on the global scale of data. The reasons that prompt the adoption of such models are set by several fundamental benefits:

- Powerful hierarchical and multi-scale feature extraction on complicated medical data.
- Capability to exploit large-scale, ready-trained architectures, improving the performance using small labeled datasets.
- Stability to noise, artifacts, and inter-patient variation of actual clinical data.
- Multi-modal diagnostic modality, such as ECG, MRI, CT, and multimodal.
- Greater contextual knowledge with the help of attention mechanisms, which allows better interpretability and global feature modeling.

II. BACKGROUND AND RELATED WORK

Diagnostic Medicine Medical signals and images are a multi-disciplinary science that applies the principles of biomedical engineering, signal processing, and computational intelligence to derive clinically valuable information amid the swamp of imaging physiological surges. The Electrocardiogram (ECG) signal is a time series representing the electrical activity of the heart and has a waveform with repetitive components, including the P wave, QRS complex, and T wave.

These sections are very important for providing details about the heart rhythm, conduction patterns, and defects that may occur, such as arrhythmias. Conversely, medical imaging technologies, such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound, provide high-dimensional spatial information on the structural and functional characteristics of tissues and organs. Noises, motion artifacts, inter-patient variability, and fluctuations in acquisition protocols are examples of issues that may be encountered when analyzing such data. The classic pipeline used in analysis generally comprises the following procedures: data capture, pre-processing, segmentation, feature acquisition, and classification. Preprocessing is also important for quality-enhancing methods for such data, including filtering, normalization, denoising, and baseline correction. Segmentation algorithms, i.e., finding cardiac cycles in ECG signals or lesions and anatomic boundaries in medical images, do isolation of areas of interest.

In the past, feature extraction has been performed using hand-crafted features (such as statistical features (mean, variance), morphological features, and frequency-domain features obtained through a Fourier or a wavelet transform). These methods also have a limitation in their ability to generalize to non-homogeneous data, as they are premised on assumptions in the field. With the introduction of machine learning and then deep learning, this paradigm has changed. At this point, automated feature learning is possible, and the model can learn complex, non-linear features directly from raw data. This move has been invaluable in advancing the scalability, flexibility, and diagnostics of medical information analysis architecture, and in the basis of intelligent and automated health care solutions [11]-[13].

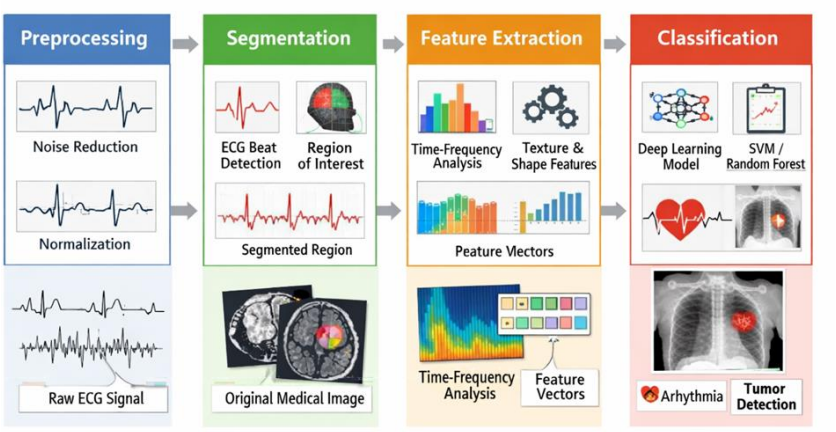

Figure 1: Fundamental Pipeline of Medical Signal and Image Analysis

It is against these fundamental processes that more contemporary methodologies of medical signal and image analysis are emerging, and a wide array of advanced computational methods that significantly improve the quality of analyses and their clinical translation. Signal processing methods, such as the Fourier transform, the short-time Fourier transform (STFT), and the wavelet transform, allow decomposition of ECG signals into time-frequency representations, thereby enabling identification of temporal abnormalities and minor variations in cardiac activity. Similarly, image processing algorithms, e.g., edge detection, histogram equalization, region growing, and morphological operations, are applied to enhance the structural information and the precision of segmentation in medical images. Classical machine learning algorithms widely applied to classification tasks based on engineered features include support vector machines (SVMs), k-nearest neighbors (k-NN), and random forests. These techniques, though, are weak in their use of quality characteristics and in their ability to elucidate complex data distributions. Deep learning has addressed these weaknesses by developing architectures that learn features in a hierarchical manner. Spatial features are particularly well-suited to CNNs because they are designed with stacked convolutional filters and therefore are optimally applied to medical imaging problems. At the same time, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are specifically designed to model temporal dependencies, making them well-suited for modeling temporal dependencies in sequence data such as ECG signals. More recently, transformer-based architectures have also been found to better represent long-range dependencies through attention mechanisms and have been significantly better on both imaging and signal-based tasks. All these advances represent the next step in integrating end-to-end learning systems that combine preprocessing, feature extraction, and, in a single optimized system, thereby bringing robustness, scalability, and diagnostic accuracy to existing healthcare systems [14], [15].

III. VISION MODELS AND THEIR RELEVANCE TO ECG

Vision-based deep learning models have become a productive paradigm for analyzing medical data, particularly in fields where more intuitively derived one-dimensional indicators, such as electrocardiograms (ECGs), are employed. Although ECG signals are temporal, recent advances have enabled their transformation into two-dimensional representations, such as spectrograms, scalograms, and recurrence plots, thereby enabling the effective utilization of vision models. Convolutional neural networks (CNNs) and Vision Transformers (ViTs), designed to solve large-scale image recognition tasks, can learn hierarchical and context-based features in representations [9]. Such models are very effective at identifying both local and global correlations in spatial variation, which are significant for detecting minor alterations in cardiac activity.

Their application to ECG classification through transfer learning has also been accelerated by the success of pre-trained vision architectures, particularly those trained on massive datasets such as ImageNet [10]. Such models can generalize successfully even when data are limited in a medical setting, owing to the representations learned during training. In addition, due to the lack of sensitivity to noise and signal variation characteristic of real-life ECG data, vision models are insensitive to these issues. Their flexibility across a range of representation and diagnostic processes underscores their importance in automated cardiac analysis systems, serving as an intermediate between signal processing and visual learning [9]-[11].

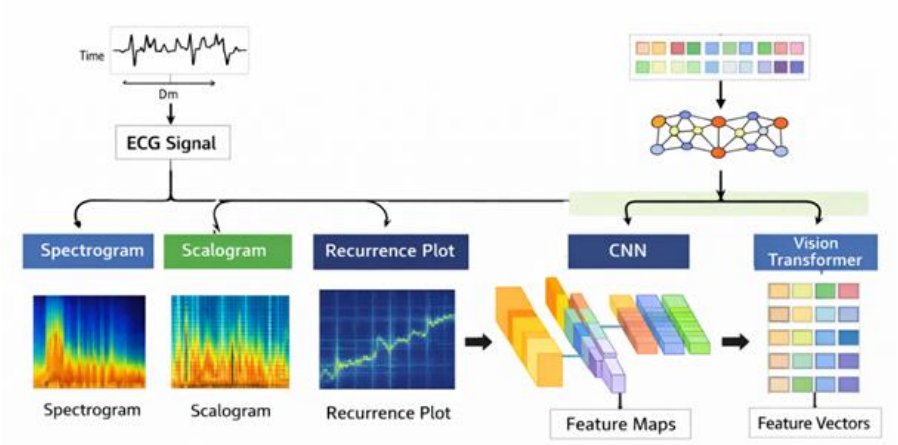

Figure 2: Application of Vision-Based Deep Learning Models To ECG Signal Representations

A. Convolutional Neural Networks (CNNs) in ECG Analysis

The fact that convolutional neural networks can be successfully trained to obtain hierarchical spatial representations from structured inputs is a key factor in the application of vision-based deep learning algorithms to analyze ECG as a core element of their implementation. CNNs can use their convolutional filters to isolate local pattern components (e.g., edges, textures, waveform morphologies) that correlate with specific cardiac pathology by converting ECG signals into two-dimensional representations, such as spectrograms, scalograms, and recurrence plots. This capability is also particularly helpful for observing arrhythmias, ischemic changes, and minimal deviations in cardiac cycles that may be difficult to detect using traditional methods. Other popular architectures, such as AlexNet, VGGNet, and ResNet, have also been shown to perform well for large-scale image classification and have been successfully applied to ECG classification using transfer learning. Even in situations where little labelled data is available, they can still perform feature extraction.

CNNs are a layered construction that enable gradual feature abstraction: lower levels learn low-level features, such as signal gradients and edges; middle layers learn higher-level features, such as waveforms; and higher levels learn more abstract features, such as semantic representations of pathological conditions. They are strong and versatile and have been extensively applied in automated diagnostic pipelines. The fact that their receptive field is small and, consequently, they cannot capture long-range temporal dependencies with long ECG sequences, is one of the major weaknesses of CNNs. This has been the catalyst for the finding of hybrid architectures and transformer-based models that may be more suited to capture global contextual relationships [3], [10].

B. Vision Transformers (ViTs) for ECG Classification

A new system in deep learning is the Vision Transformer (ViT): it does not rely on convolutional operations but rather on self-attention to model data relationships. Unlike CNNs that operate on localized receptive fields, ViTs process each input image as a sequence of fixed-size patches, enabling the model to capture long-range dependencies and global contextual information. This international modeling may be particularly useful in ECG analysis, where a signal is converted into a two-dimensional representation, e.g., a spectrogram or a wavelet scalogram. There may be numerous temporal divisions involved in CAD disorders, and ViTs' ability to establish relationships between unrelated areas improves recognition of more complex structures, such as atrial fibrillation and ventricular arrhythmias.

The next significant ViT's strength is its scalability. ViTs also demonstrated exceptional performance of conventional CNNs as model and dataset sizes increase; hence, they are being scaled to large medical datasets. The attention mechanism also has inherent interpretability, as it indicates which portions of the input contribute most to the classification decision, which is especially important in a clinical environment where explainability is paramount. Transfer learning can also be used to address this issue, as it enables the use of pre-trained transformer models, which can be refined on smaller ECG datasets. The recent objectives indicated that ViTs with appropriate representation strategies and transfer learning patterns can achieve high performance compared to CNN-based models, with a specific emphasis on simple and long-range dependencies for ECG categories [9].

C. Why Vision Models Work for ECG

The reason why vision-based models are very effective in analyzing ECG is that converting the time-series signals, which are one-dimensional, into two-dimensional representations preserves both temporal and frequency-domain information. Techniques such as the short-time Fourier transform (STFT), continuous wavelet transform (CWT), Gramian Angular Fields (GAF), and recurrence plots permit encoding ECG signals as high-structure images. These representations

reveal complex trends, including changes in frequency, temporal dependencies, and nonlinear dynamics, which are crucial for accurate heart diagnosis. These methods enable vision models to realize their inherent potential to extract spatial features and perform pattern recognition by converting ECG signals into image data.

The work is particularly well-suited to CNNs and ViTs, as they are trained to capture spatial correlations and hierarchical relationships in data. CNNs are better at local features and texture detection than ViTs, which have a complementary advantage in global dependence, as shown by attention checks. In addition, transfer learning on large image datasets provides these models with a powerful initialization point, enabling them to learn strong, generalizable features from small medical datasets. Advanced vision architectures have evolved by integrating signal processing techniques, yielding major improvements in classification accuracy, robustness, and scalability. This intersection of biomedical signal processing and computer vision has thus rendered vision models an influential contributor to the development of ECG analysis frameworks to support clinical decision-making and real-time healthcare systems [7], [12].

IV. ECG REPRESENTATION TECHNIQUES FOR VISION TRANSFER

The successful implementation of vision-based deep learning models for electrocardiogram (ECG)-based analysis depends entirely on converting one-dimensional signals into two-dimensional signals. The spatial features can be extracted from these changes, and these changes do not affect significant temporal and frequency-related information inherent in cardiac signals. Time-frequency representations, such as spectrograms generated by the short-time Fourier transform (STFT), are among the most popular techniques for revealing details in local frequency over time [13, 14]. Similarly, the continuous wavelet transform (CWT) offers both multi-resolution analysis and high-frequency transient properties, as well as low-frequency tendencies, which render it particularly useful for detecting arrhythmia. Other encoding techniques besides these have also been explored, such as Gramian Angular Fields (GAF) and Recurrence Plots (RP), as means of encoding temporal dependencies with regular image patterns [15]. These processes allow vision models to leverage their inherent capacity to seek textures, forms, and spatial relationships. The type of representation used can significantly influence the quality of features learned during training and ensure they are readable. Moreover, the representations enable the use of transfer learning to train the ECG data to match image-based model architectures. As a result, representation learning will be a central interface between biomedical signal processing and computer vision, delivering potent, scalable, and high-performance diagnostic systems in contemporary healthcare [16]-[20].

Table 1: ECG Representation Techniques for Vision Transfer

Technique	Method Description	Domain	Key Features Captured	Advantages	Limitations	Common Use Cases
Raw Signal Plot (2D Mapping)	Converts 1D ECG waveform into 2D image plots	Time-domain	Amplitude variations, waveform morphology	Simple implementation, preserves original signal structure	Limited frequency information	Basic ECG classification, visualization
Spectrogram (STFT)	Applies the short-time Fourier transform to generate a time-frequency representation	Time-Frequency	Frequency variations over time	Captures localized frequency features, widely used	Fixed resolution trade-off	Arrhythmia detection, abnormal rhythm analysis
Scalogram (CWT)	Uses the continuous wavelet transform for multi-resolution analysis	Time-Frequency	Multi-scale temporal and frequency features	High-resolution captures transient patterns effectively	Computationally intensive	Complex cardiac abnormality detection

Gramian Angular Field (GAF)	Encodes time series into polar coordinates and computes angular relationships	Spatial Encoding	Temporal correlations, global structure	Converts time dependencies into visual textures	May lose fine temporal granularity	Pattern recognition, deep learning models
Recurrence Plot (RP)	Visualizes recurrence of states in phase space	Non-linear Dynamics	Repetitive patterns, system dynamics	Captures non-linear relationships	Sensitive to parameter selection	ECG anomaly detection, chaos analysis
Markov Transition Field (MTF)	Encodes transition probabilities of time series values	Probabilistic	State transitions over time	Preserves temporal dynamics in image form	Complex computation	Advanced classification tasks
Hybrid Representations	Combines multiple techniques (e.g., STFT + GAF)	Multi-domain	Combined temporal, frequency, and spatial features	Improves model robustness and accuracy	Increased computational cost	High-performance deep learning pipelines

V. TRANSFER LEARNING STRATEGIES

Transfer learning is also a fundamental approach in modern deep learning systems, particularly when expensive and limited labelled data are available. In practice, in most real-world tasks, it is not possible to train models on raw data due to limited data, costly annotation, and overtraining. Transfer learning addresses the issues by leveraging information from models trained on large-scale datasets and applying the resulting representations to new, domain-specific problems. The given approach will dramatically reduce the amount of labelled data and improve model performance and training rates. Premade models can handle rich hierarchical features, from low-level features such as edge patterns and textures to high-level semantics. Transfer learning can be applied across a broad spectrum because the characteristics can be effectively applied in other fields. The most commonly used are feature extraction, where the already trained model is used as a fixed backbone, and fine-tuning, where a small or all the layers of the model are retrained to fit the new application. With the adoption of hybrid solutions, performance is further enhanced by incorporating diverse architectures or learning paradigms to capture diverse patterns in the data.

In addition to these conventional approaches, superior methods such as domain adaptation, self-supervised learning, and multi-task learning have expanded the scope of transfer learning. Domain adaptation aims to reduce the gap between the source and target data distributions, and self-supervised learning helps models learn to use beneficial representations from unlabeled data. In its turn, multi-task learning improves generalization by learning similar goals simultaneously. Combined, these ensure that it is more powerful, convergent, and scalable. Transfer learning has therefore emerged as a basis for developing effective, general, and powerful deep learning systems across a wide range of applications.

Table 2: Transfer Learning Strategies for ECG Classification

Strategy	Description	Training Approach	Advantages	Limitations	Best Use Case
Feature Extraction	Uses a pre-trained model as a fixed feature extractor; only the final classifier is trained	Freeze all backbone layers; train output layer	Fast training, low computational cost, effective with very small datasets	Limited adaptability to ECG-specific patterns	Small datasets, baseline models
Fine-Tuning (Partial)	Fine-tunes higher layers while keeping lower layers frozen	Train top layers; freeze early layers	Balances generalization and specialization	Requires careful layer selection	Moderate dataset size

Fine-Tuning (Full)	The entire model is retrained on the ECG dataset	Train all layers with a lower learning rate	High adaptability, better performance	Risk of overfitting, higher compute cost	Larger labeled datasets
Domain Adaptation	Aligns feature distributions between source (images) and target (ECG) domains	Adversarial training or feature alignment	Improves cross-domain generalization	Complex training process	Multi-source or heterogeneous datasets
Self-Supervised Pretraining	Learns representations from unlabeled ECG data before fine-tuning	Retrain on unlabeled data, then fine-tune	Reduces dependency on labeled data	Requires large unlabeled datasets	Data-scarce environments
Multi-Task Learning	Trains a model on multiple related tasks simultaneously	Shared backbone, multiple output heads	Improves generalization and robustness	Increased model complexity	Multi-label ECG classification
Layer-wise Transfer Learning	Gradually unfreezes layers during training	Progressive training of layers	Stable convergence, better adaptation	Longer training time	Fine-grained optimization
Cross-Domain Pretraining	Pretraining on related biomedical datasets instead of natural images	Train on a similar domain before the target task	Better domain relevance	Limited availability of datasets	Medical-specific applications
Hybrid Transfer Learning	Combines CNN + Transformer or multiple pre-trained models	Ensemble or hybrid architecture	High accuracy, captures local & global features	High computational cost	High-performance systems

VI. LIMITATIONS AND CHALLENGES

Despite the clear fact that much has been done to enable automated ECG classification using deep learning and transfer learning models, several limitations and challenges could be seen as obstacles to widespread adoption in clinical settings. Some of the major problems include inconsistencies and quality issues in medical data, as ECG signals may be affected by noise, motion artifacts, and defects in acquisition protocols. The aspects may degrade the model's functioning and limit its generalizability to other healthcare settings. Also, the reliance on labelled datasets, which are time-consuming and expensive to obtain and require expert clinicians to make high-quality annotations, is also a serious problem. This often results in small, nontrivial datasets that can bias model predictions. The other issue is that the dissimilarity between the trained models (trained on natural images) and the medical signal data (in a different domain) can affect the transfer of learned features. Besides, deep learning models may be considered black-box systems, which raises issues for their interpretation and use in clinical decision-making. Practical constraints can also arise from the complexity of the computation and its resource requirements, particularly when applied in real time or under extremely low-resource conditions, such as wearable devices. Finally, the objective comparison of models and the assessment of their true clinical usefulness are complicated by the lack of standardized evaluation schemes and benchmarking data. Such problems should be addressed to achieve scalable, robust, and reliable ECG classification systems.

A. Data Quality and Variability

The hardest problem in ECG analysis is that data quality is not constant due to differences in data collection tools, the patient's condition, and the surrounding environment. ECG signals are particularly prone to noise, including baseline wander and muscle artifacts, as well as power-line interference, which can mask important features in the waveforms. In addition, the datasets are not identical due to variations in electrode position, sampling, and recording period. This lack of consistency renders models difficult to generalize, especially when they are trained on small or homogenous data. This is because the model's forecasting is skewed by class proportions, with a normal heart rhythm being forecast more often than the uncommon pathological condition. Some ways to reduce variability include preprocessing (e.g., filtering and normalization), but it can never be reduced to zero. The success of data augmentation and domain adaptation schemes is

essential for making the model more resilient and for providing stable performance across a wide range of clinical conditions.

B. Domain Gap and Transferability Issues

Transfer learning has been successful in a number of applications, but the domain gap between natural image datasets and medical signal representations has been a thorn in the flesh. ImageNet and other datasets contain pre-trained models whose trained features are conditioned on visual objects, textures, and colours that are not necessarily similar to the properties of ECG-derived representations. Transforming ECG signals into image-like representations might be one way to address this gap, yet, at any rate, differences in feature distributions might continue to affect model performance. This is a significant problem, especially when transfer of knowledge between data sets or between groups of patients is involved. Adversarial learning and feature alignment are two proposed domain adaptation methods that offer benefits for this problem, but each approach typically incurs increased complexity and computational cost. The methods of representation, model design, and fine-tuning process will be chosen sufficiently, regarding the field of interest.

C. Interpretability and Clinical Trust

Deep learning models are not interpretable, which is a significant obstacle to their implementation in clinical settings. Medical practitioners need clear, understandable mechanisms to facilitate decision-making, especially when making high-stakes decisions, such as cardiac diagnoses. Nevertheless, the majority of deep-learning models are black boxes, not explicitly explained in terms of their reasoning process. Such a lack of transparency may lower the levels of trust and acceptance among clinicians. Other methods, such as Grad-CAM, saliency maps, and attention visualizations, have been proposed to enhance interpretability by highlighting salient regions of the input data. Although these approaches provide some information, they are seldom sufficient to elucidate model behavior. The creation of inherently interpretable models and the application of explainability to the design process are areas of critical research that need to ensure clinical reliability and trustworthiness.

D. Computational Complexity and Deployment Constraints

The other major problem is the computational complexity of the deep learning models, particularly the large CNNs and transformer architectures. Their computational requirements for training and inference are also high, requiring high-performance GPUs and large memory. This can limit their applicability in real-time systems and resource-limited settings, such as wearable and mobile health applications and rural health centres. In addition, latency and power consumption can be important aspects of deployment, as timely and efficient processing is needed. The following model optimization techniques have been proposed to address these issues: pruning, quantization, and knowledge distillation, though they may introduce performance trade-offs between model efficiency and accuracy. That is why it is necessary to balance model performance and computational efficiency to enable the deployment of automated ECG classification systems that are practical and scalable.

VII. FUTURE RESEARCH DIRECTIONS

Automated ECG analysis using deep learning and transfer learning is a rapidly evolving field; however, some aspects of the methodology and the clinical viability of the solution warrant further development. The research of the future should not be focused on the use of pure supervised learning models, but on more data-saving systems that can be generalized and interpreted. The need to minimize the use of labelled data by leveraging unlabelled or weakly labelled data for self-supervised and semi-supervised learning paradigms is a key trend. Moreover, multimodal data sources, such as electronic health records (EHRs), imaging, and wearable sensor data, may be pooled to provide more information and contextual context, in turn resulting in more specific and tailored diagnostic systems. The other major area is the creation of light and energy-efficient designs that may be implemented in real-time settings, such as in mobile devices and wearable health systems. The creation of explainable artificial intelligence (XAI) should also be pursued to improve transparency and build trust between clinicians. Additionally, the federal learning models can be adjusted to accommodate joint model training across institutions without violating data privacy and safety. This will also be accompanied by practices such as standardizing datasets, assessment protocols, and benchmarking, which will greatly help ensure fair comparisons and the replicability of results. When combined, these recommendations are likely to eliminate the disconnect between research and the practical application of clinical solutions needed to achieve scalable, reliable, and patient-centric care.

A. Self-Supervised and Semi-Supervised Learning

The weaknesses of medical AI include the scarcity of labelled data, and self-supervised and semi-supervised learning appear to be an exceptionally productive research agenda. Self-supervised learning allows the model to learn meaningful representations from large quantities of unlabelled ECG data using pretext tasks such as signal reconstruction, contrastive learning, or temporal prediction. It is then possible to optimize these learned representations for downstream classification tasks, thereby substantially eliminating the need for annotated datasets. Semi-supervised learning then proceeds to a further

step: adding a small number of labelled data points to many unlabelled data points, thereby improving model generalization and robustness. These methods have been lurking in the medical circles: Pseudo-labeling, consistency regularization, and the teacher-student framework. Further research in this area is required to develop domain-specific pretext tasks that characterize physiological patterns specific to ECG signals. Besides, such techniques can also be improved with transfer learning to increase performance and scalability, which is why they are well-suited to real-world healthcare applications, where less data is labelled.

B. Multimodal Learning and Data Integration

The combination of multiple data modalities is another great opportunity to enhance the quality of diagnoses and clinical reasoning. ECG signals have been valuable for understanding cardiac activity; however, when combined with other sources of information, such as medical imaging, electronic health records (EHRs), lab analyses, and wearable sensors, a better understanding of the patient's health can be achieved. The concept behind multimodal learning models is to employ these divergent sources of data collaboratively to retrieve compensatory features and contextual data. This method can enhance the diagnosis of complex conditions that would be missed with a single modality. Nevertheless, issues such as data alignment, the absence of modalities, and the opposite data design should be addressed. The second step in research should focus on the means of using powerful fusion methods, such as attention-based and graph-based models, to successfully combine multimodal information. This type of system can enable personalized and precision medicine by providing an individual patient profile with a diagnosis and treatment.

C. Lightweight Models and Edge Deployment

There is a need to design light, efficient models for real-world applications, specifically in resource-limited settings. There are also some advanced deep learning systems that are computationally expensive, which limits their use in mobile computing, wearable sensors, and remote healthcare systems. The model compression algorithms that could be used to compress the model dimension, pruning, quantization, and knowledge distillation are areas of future research to ensure that the cost of computation is also minimized, even though the execution of the model is not wholly minimized. In addition, real-time ECG monitoring and diagnosis can be implemented using an edge-specifically optimized architecture. The other benefits of edge computing include achieving low latency, greater privacy, and reduced reliance on cloud systems. Internet of Things (IoT) devices might also be integrated with deep learning models to facilitate continuous health monitoring and the timely identification of cardiac abnormalities. The right balance between efficiency and accuracy will be essential to ensuring that healthcare solutions are scalable and accessible.

D. Explain ability, Trust, and Clinical Integration

One of the major concerns regarding the use of AI systems in clinical practice is transparency. Medical workers should witness the rationale and interpretable models to understand the mechanism behind the predictions and to demonstrate that they are correct. The next round of research should focus on developing explainable AI approaches that provide clinically actionable information, e.g., indicating which part of the waveform to worry about or identifying condition-specific features. The process of interpretability may be improved by methods such as attention visualization, saliency maps, and concept-based explanations, but further research is needed to ensure their consistency and reliability. In addition, the introduction of AI systems into the clinical processes should be supported by a particular emphasis on usability, adherence to the rules, and ethics. Clinicians, human-in-the-loop systems, and clinicians collaborating with AI models can facilitate improved decision-making, and there should be a balance between accountability. Strict validation will be the only way to win the trust of AI-based healthcare systems, to undergo their stringent evaluation, and to collaborate closely with researchers, clinicians, and policymakers.

VIII. CONCLUSION

The paper introduces automated ECG classification, considering both deep learning and transfer learning, with specific reference to the prospects of applying vision-based models to interpret medical signals. The paper has highlighted the evolution of classical signal processing and handcrafted feature extraction to modern end-to-end learning models capable of learning complex temporal and spatial features. One-dimensional ECG signal mapping to two-dimensional representations has made the concepts of convolutional neural networks and Vision Transformers much more useful, as they can better perform in diagnosis and be scaled. The review also verified various ECG representation strategies and transfer learning techniques and demonstrated their great relevance for overcoming the limitations of limited labelled data and high annotation costs. It was discovered after a comparative analysis of the existing literature that the hybrid and transformer-based models are merged with transfer learning to achieve state-of-the-art performance because they can identify both local and global dependencies. However, the paper has also identified key limitations, including data variability, domain gaps, interpretive challenges, and computational constraints, which continue to limit its clinical application. These problems, such as the need to pay attention to self-supervised learning, the integration of diverse data types, the development of lightweight

models, and the explanation of AI, defined the directions of future research. These needs should drive some level of improvement to develop solid, efficient, and trustworthy systems that can be readily assimilated into the actual context of healthcare environments. In recap, biomedical signal processing with vision-based deep learning, aided by transfer learning, has enormous potential to enhance automated ECG analysis, ultimately resulting in more precise, scalable, and accessible cardiac diagnostics.

Interest Conflicts: The author(s) declare(s) that there is no conflict of interest concerning the publishing of this paper
Funding Statement: Not Applicable

IX. REFERENCES

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
[2] Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The met Press, 2016, 800 pp., sib: 0262035618. Genetic programming and evolvable machines, 19(1), 305-307.
[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.
[4] Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. IEEE engineering in medicine and biology magazine, 20(3), 45-50.
[5] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv preprint arXiv:1707.01836.
[6] Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual review of biomedical engineering, 19, 221-248.
[7] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.
[8] Chen, Y. (2025). Deep Semi-Supervised Point Cloud-Based Models with Uncertainty Awareness for Efficient Abnormality Detection in 3D Medical Imaging (Doctoral dissertation, Queensland University of Technology).
[9] Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., ... & Dosovitskiy, A. (2021). Mlp-mixer: An all-mlp architecture for vision. Advances in Neural Information Processing Systems, 34, 24261-24272.
[10] Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. Advances in Neural Information Processing Systems, 32.
[11] Webster, J. G. (Ed.). (2009). Medical instrumentation: application and design. John Wiley & Sons.
[12] L. Sörnmo and P. Laguna, Bioelectrical Signal Processing in Cardiac and Neurological Applications, Elsevier, 2005.
[13] Rangayyan, R. M. (2015). Analysis of Concurrent, Coupled, and Correlated Processes.
[14] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21-27.
[15] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. nature, 542(7639), 115-118.
[16] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large-scale visual recognition challenge. International journal of computer vision, 115(3), 211-252.
[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[18] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
[19] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In the 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255). Ieee.
[20] Wang, Z., & Oates, T. (2015, January). Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In Workshops at the twenty-ninth AAAI conference on artificial intelligence (Vol. 1, No. 1, pp. 20-954).