Hochschule Karlsruhe
University of
Applied Sciences

Fakultät für
Informatik und
Wirtschaftsinformatik

Master-Thesis

Name: Katharina Post

Thema: Classification of arrhythmias based on electrocardiograms and photoplethysmography with deep learning

Arbeitsplatz: inovex GmbH, Karlsruhe

Referent: Prof. Dr.-Ing. Laubenheimer

Korreferent: Prof. Dr. Baier

Abgabetermin: 31.03.2023

Karlsruhe, 01.10.2022

Der Vorsitzende des Prüfungsausschusses

H. Körner

Prof. Dr. Heiko Körner

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig und ohne fremde Hilfe angefertigt habe. Textpassagen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht. Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Karlsruhe, 30.03.2023

Katharina Post

Zusammenfassung

Herzrhythmusstörungen sind weltweit eine der häufigsten Todesursachen. Diese Todesfälle können vermieden werden, wenn Herzrhythmusstörungen frühzeitig erkannt und überwacht werden. Die Elektrokardiographie ist ein wichtiges Instrument zur Erkennung von Herzrhythmusstörungen, erfordert jedoch spezielle Geräte und geschultes Personal. Um Ärzte bei der Elektrokardiogramm-Diagnose zu unterstützen, wurden bereits zahlreiche Machine- und Deep-Learning-Techniken entwickelt. Diese Techniken berücksichtigen verschiedene Aspekte wie die zu klassifizierenden Arrhythmien oder die Anzahl der verwendeten Elektrokardiogramm-Kanäle. Neuerdings wird Photoplethysmographie zunehmend als Alternative zur Überwachung der kardiovaskulären Gesundheit auch außerhalb des klinischen Umfelds eingesetzt. Während einige Smart Watches bereits in der Lage sind, einzelne Arrhythmien zu erkennen, sind die Ansätze zur Klassifizierung von multiplen Arrhythmien noch begrenzt. Das Ziel dieser Arbeit ist die Verbesserung der Klassifizierung von multiplen Herzrhythmusstörungen mit Photoplethysmographie. Dabei wird der Einsatz von Transfer Learning als eine Möglichkeit zur Verbesserung evaluiert. Dazu wird eine Modellstruktur entwickelt, die auf Elektrokardiogrammen trainiert wird. Diese vereint Vorteile und Aspekte aus früheren Arbeiten. Eine Herausforderung dabei ist die Unterrepräsentation einiger Klassen in der Elektrokardiogramm-Datenbank. Dies führt dazu, dass die entwickelte Modellarchitektur nicht die Benchmark-Ergebnisse aus vorherigen Publikationen erreicht. Daher wird zusätzlich ein Ensemble aus mehreren binären one-vs-all Klassifikatoren betrachtet. Das Ensemble-Modell erreicht bei der Klassifikation von fünf Herzrythmusklassen einen F1-Score von 91%. Bei der Klassifikation von Herzrhythmusstörungen auf der Basis von Photoplethysmogrammen führt die Anwendung von Transfer Learning zu einer Verbesserung gegenüber dem Modell, das nur auf Photoplethysmographie-Daten trainiert wurde. Die besten Ergebnisse erzielt jedoch wieder das Ensemble-Modell mit einem F1-Score von 89%. Die Arbeit zeigt, dass die Anwendung von Transfer Learning in der Arrhythmieklassifikation zu Verbesserungen führen kann. Das Ensemble-Modell ist jedoch ein Ansatz, der vielversprechendere Ergebnisse liefert und die Benchmark-Ergebnisse im Bereich der Klassifizierung von Herzrhythmusstörungen mittels Photoplethysmographie übertrifft.

Abstract

Cardiac arrhythmias are one of the leading causes of death worldwide. These deaths can be prevented if arrhythmias are detected and monitored early. Electrocardiography is a major tool for detecting arrhythmias. However, it requires specialized equipment and trained personnel. Many machine learning and deep learning techniques have been developed to help doctors diagnose electrocardiograms. These techniques incorporate different aspects such as the number of arrhythmias to be classified or the number of electrocardiogram channels used. More recently, photoplethysmography has been increasingly used as an alternative for monitoring cardiovascular health in nonclinical settings. While some smartwatches can already detect single arrhythmia classes, approaches to classify multiple arrhythmias are still limited. This work aims to improve the classification of multiple arrhythmias using photoplethysmography. Thereby, the use of transfer learning is evaluated to improve classification performance. For this purpose, a model structure is developed that is trained on electrocardiograms. This model design combines advantages and aspects from previous work. One challenge is the underrepresentation of some classes in the electrocardiogram database. As a result, the performance of the developed model architecture is behind the benchmarks published in previous work. Therefore, an ensemble of several binary one-vs-all classifiers is additionally considered. The ensemble model achieves an F1-score of 91% for the classification of five arrhythmia classes on electrocardiography data. When classifying arrhythmias based on photoplethysmography, the application of transfer learning leads to an increase in performance over the model trained on photoplethysmogram data solely. However, the ensemble model again achieves the best performance with an F1-score of 89%. The work shows that applying transfer learning to arrhythmia classification can lead to a performance gain. However, the ensemble model is an alternative method that yields more promising results and outperforms the benchmark results in multiclass arrhythmia classification on photoplethysmography.

Contents

1 Introduction

The introduction of this work begins with the motivation of why it is important to detect and classify cardiac arrhythmias based on electrocardiograms and photoplethysmography. Then, the research aims are presented. Afterwards, the company is introduced, with whose cooperation the thesis is written, and finally, an overview of the thesis structure is given.

1.1 Motivation

According to the World Health Organization (WHO), cardiovascular diseases are the most common non-communicable disorder worldwide, responsible for an estimated 17.8 million deaths in 2017. As the UN Sustainable Development Goals aim to reduce premature mortality from non-communicable diseases by one-third before 2030, accurate early detection systems for heart failure that can continuously monitor individuals are required [1].

Arrhythmia is a commonly diagnosed cardiac disease caused by an irregular heartbeat. It occurs when the electrical impulses that coordinate the heartbeat malfunction. This results in a heartbeat that is too fast or too slow, or a skipping of the heartbeat [2]. All humans experience irregularities in their heart rate. Thus, it is important to distinguish between different types of heartbeats. Initially, harmless irregular beats can develop into persistent cardiac arrhythmias over time, becoming life-threatening in the worst case. Therefore, it is important to classify the various arrhythmias [3].

Currently, cardiac arrhythmias are usually detected after complex examinations in the clinical environment with the help of electrocardiograms (ECGs). However, such examinations are normally carried out when the patient notices symptoms.

Some patients do not notice any changes even though they suffer from arrhythmias [4]. ECGs are used as a standard in everyday clinical practice. Thus, more than 300 million ECG strips are produced worldwide every year [5], which have to be evaluated manually by doctors. Nevertheless, not all individuals who would need such monitoring are regularly screened for arrhythmias.

The advancement of deep learning offers new possibilities in the automatic analysis of ECG strips. Several studies have been published in recent years on the application of machine and deep learning in the field of arrhythmia classification [6][7][8]. Despite a meanwhile very good automatic classification capability, ECGs are still too expensive and too impractical to obtain permanent monitoring of all individuals.

Photoplethysmography (PPG) is a technique that has become increasingly popular in recent years. It can be used to measure the volumetric change of blood in the microvascular bed of tissue, which correlates with the heart cycle. This produces waveforms that are related to those from the ECG. There are already initial efforts to use this technique for the detection of cardiac arrhythmias, as it is cheaper and easier to use than ECGs. This technique can be found, for example, in smart devices such as smartwatches, but can also be recorded using the mobile phone camera and flash.

However, previous methods in this area mostly use classic machine learning approaches in order to detect abnormalities in the heart function, which are very time-consuming due to manual feature generation [4][9]. In the last year, deep learning has been used in some research, but the performance does not reach that of ECG approaches yet.

To transfer the good results from ECGs to PPGs, initial efforts are being made to apply transfer learning. For example, models are trained with ECG data, and the last layers of the network are frozen and re-trained with PPGs afterwards. In 2021, two papers were published by Radha et al. [10] and Li et al. [11] that use transfer learning to classify the sleep stage of individuals. The models are each pre-trained on ECGs and then transferred to PPGs. Both find an improvement in accuracy compared to the models trained on PPGs only. In addition, an article published in 2021 by Ramesh et al. [12] applies transfer learning in the field of arrhythmia classification. The approach aims to classify one particular type of cardiac arrhythmia, based on the heart rate variability extracted from ECGs. Then the final layers are re-trained with the pulse rate variability from PPGs. Since initial successes have already been achieved with the transfer learning approach from ECGs to PPGs, this work will examine whether a transfer learning method can likewise achieve better results in the multiclass classification of arrhythmias.

1.2 Objective

Two primary objectives are investigated in this project.

1. How can a model for classifying cardiac arrhythmias be designed that combines the advantages/aspects of previous publications? This question is addressed by comparing different approaches in the literature and identifying the advantages of each publication. In addition, the challenges and deficits of the previous approaches are considered so that possible research gaps can be closed.
2. Does a model pre-trained with ECG data and applied to PPG data via transfer learning achieve a better F1-score in multiclass arrhythmia classification than a model trained on PPG data only? Therefore, the model resulting from the first research question is transferred to PPG data. If the model achieves an F1-score outperforming that of the model trained on PPG-only data, this can be considered a success.

In the following, the particular steps that need to be undertaken to answer the the first research question are specified:

- Conduction of a comprehensive literature review focusing on the classification of cardiac arrhythmias based on ECGs and PPGs.
- Identification of advantages and important aspects in the literature.
- Identification of challenges and shortcomings of previous publications to detect research gaps.
- Creation of a model structure and approach that combines the advantages of previous publications.

To answer the second research question following tasks are conducted:

- Implementation and training of the model structure developed in 1. with ECG sequences, taking into account further aspects that are found in the answering of the first research question.
- Implementation and training of the model with PPG data.
- Execution of transfer learning by re-training the final layers of the model trained on ECG data on the PPG sequences.
- Comparison of the performance of the PPG-only model with the transfer learning model.

Accordingly, this work results in a model that detects and classifies cardiac arrhythmias in ECGs. Various aspects regarding the advantages and disadvantages of previous publications are to be taken into account. Exemplary aspects are the classification possibility of several arrhythmias or that the model works as well in new patients as in subjects from the training data set. Furthermore, the transferability of the model to PPG data must be considered.

The proposed model is based on the paper *Topological Data Analysis for Arrhythmia Detection Through Modular Neural Networks* published in 2020 by Dindin et al. [13]. The authors developed a modular multichannel neural network consisting of an autoencoder, a deep convolutional neural network (DCNN), handcrafted features, and topological data analysis (TDA). The autoencoder, as well as the topological data analysis, are re-implemented. The structure of the DCNN is adopted to the data used in this work. An overview of the exact model structure is given in Chapter 4.

1.3 Environment

The thesis is written in cooperation with the company *inovex GmbH* in Karlsruhe. *inovex* is an innovation and quality-driven IT project house with a focus on digital transformation. Approximately 500 employees at seven locations in Germany support customers from different industries in the realization of digital use cases. The company's portfolio includes mobile apps, robotics, backend services, cloud services, data science, and deep learning [14].

The focus of *inovex* can be divided into three fields. The *Application Development* unit includes software development with backend and frontend, as well as UI/UX design. The employees from the *IT Engineering and Operations* division develop infrastructure for various projects. And the *Data Management and Analytics* department creates data-driven solutions using machine and deep learning methods, as well as big data and business intelligence [14].

Furthermore, *inovex* is involved in several research projects and science. New technologies are developed in the areas of Internet of Things (IoT), artificial intelligence, and e-health. Within this scope, many theses are offered in the company. Both academic questions and concrete use cases from the economy are investigated [15]. Such work belongs to the area of *inovex Lab*, the central unit for research and development [16]. This work was undertaken as part of the *inovex Lab* and aims to gain further insights into the topic of e-health.

1.4 Structure

Chapter 1 provides an overview of the motivation, objectives and the working environment of the project. Chapter 2 introduces relevant theoretical foundations in the field of medicine and topological data analysis. Chapter 3 discusses prior research that is relevant to this project, which is mostly research related to the automatic classification of arrhythmias based on ECGs and PPGs. Chapter 4 describes the methods used in the project, including the generation of the data set, the concept of the multichannel neural network and a description of each model component. Chapter 5 presents the implementation and results of the proposed network on ECG data and the transfer to PPG data. Finally, chapter 6 discusses and summarizes the findings of the thesis, describes the limitations of the methods, and gives suggestions for further research.

2 Theoretical Foundations

Relevant terms from the fields of medicine and informatics are presented below. The technology of electrocardiography and photoplethysmography is outlined to create an understanding of the differences between these two methods. In addition, cardiac arrhythmias are introduced and some classes that will be categorized in the later model are explained. Further, topological data analysis is delineated with a focus on persistent homology, which can be used to obtain features for the classification of arrhythmias.

2.1 Electrocardiography and Photoplethysmography

The most common diagnostic method for cardiac arrhythmias is the recording of an electrocardiogram (ECG). The advantage of this method is a quick and reliable assessment of heart function.

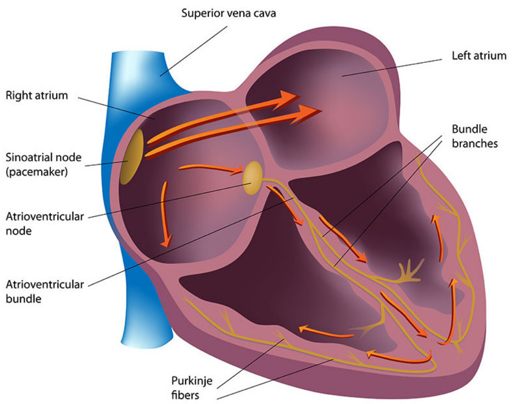

Figure 2.1: Cross-section of the heart – the sinus node gives the beat and thus determines the heart rhythm [17].

An ECG records the electrical activity of all heart muscle fibers. With every heartbeat, the heart muscles change their electrical properties. These voltage changes are measured on the patient's body with electrodes. The resulting heart electricity curve is called an electrocardiogram [18]. The ECG can be used to monitor the rhythm of the heart. Each heartbeat is represented in the ECG as a waveform. Three consecutive heartbeats with identical waveforms are defined as one rhythm. The similarity of the waveforms indicates that the beat has the same origin. Normally, the sinus node (cf. Figure 2.1) is the predominant pacemaker centre and thus the resulting rhythm is called sinus rhythm. Other pacemakers may be, for example, parts of the atrial myocardium or clusters of cells around the atrioventricular node. If a cell develops pathological automaticity, extrasystoles and arrhythmias can occur [19].

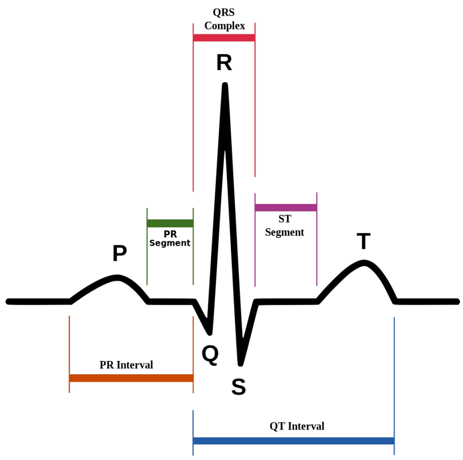

Figure 2.2: Waveform of one heart beat which consists of a P-wave, a QRS-complex and a T-wave [20].

A waveform (shown in Figure 2.2) consists of the P-wave, the QRS-complex and the T-wave. The P-wave represents depolarisation of the atrium, the QRS-complex indicates depolarisation and contraction of the ventricles and the T-wave denotes repolarization of the ventricles [2]. The distance between two R-spikes is used to calculate the heart rate. Normal sinus rhythm is characterised by regularity with a heart rate of 50 to 100 beats per minute. In addition, the P-wave has a constant morphology before each QRS-complex [19].

There are different types of ECGs, which differ in their number of electrodes and leads. There are a total of twelve leads (cf. Figure 2.3), which are measured with a sum of ten electrodes. Such an ECG is called a 12-lead ECG and is part

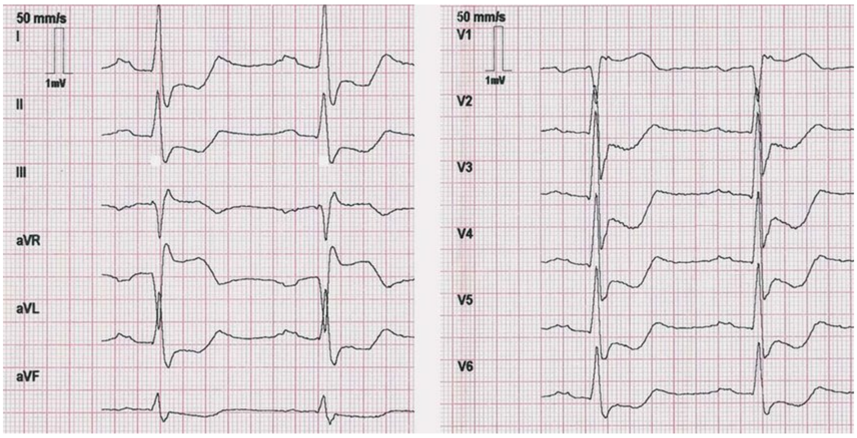

Figure 2.3: Example of a 12-lead-ECG. Lead II is used in this work [21].

of standard diagnostics. For a long-term ECG (24 hours), a total of three leads are recorded. In addition, there is also a one-lead ECG, which enables long-term monitoring of patients. However, only a single lead is recorded, so comprehensive diagnostics cannot be guaranteed. Channel II is usually used for such recordings [18]. In this work, single-lead ECGs are used because they are suitable for the constant monitoring of a patient. In addition, the same cardiac arrhythmias can be detected with a single-lead ECG as with photoplethysmography.

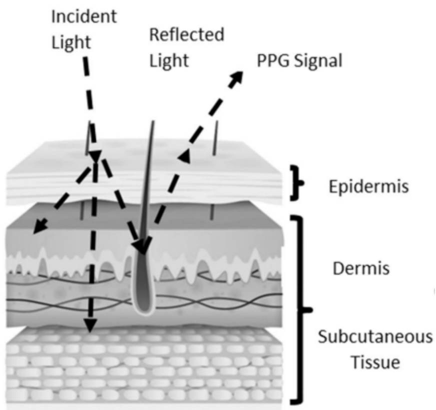

Figure 2.4: Working principle of PPG, where the dashed line indicates the light [7].

Photoplethysmography (PPG) is used to monitor the volumetric change of blood in the microvascular bed of the tissue [7]. For this purpose, the PPG-sensor emits a light that is reflected or transmitted in the tissue. The functioning of the PPG-sensor is shown in Figure 2.4.

Here, green light is used because the wavelengths of green light are strongly absorbed by the blood. During the systolic phase of the cardiac cycle, the arteries obtain more blood volume than in the diastolic phase and thus the change can be recorded by PPGs [22]. The resulting PPG-waveform is shown in Figure 2.5. By analyzing the PPG-waveforms, arrhythmias can be detected and classified.

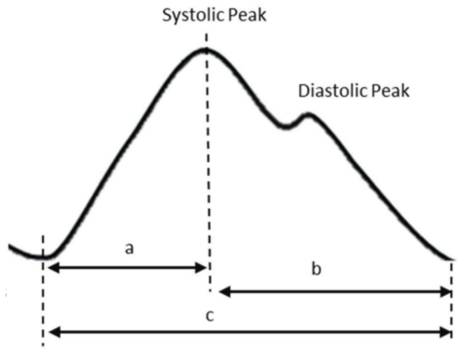

Figure 2.5: Representation of a PPG waveform, where a represents the depolarization of atria, b represents the repolarization of ventricles, and c represents the pulse width [7].

The ECG and PPG waveforms are related to each other. The QRS-complex of the ECG maps the contraction of the heart ventricles. Afterward, the blood is transported through the blood vessels to different body parts. Thus, the volume of blood increases after each heartbeat. The R-spikes of the ECG represent ventricular contraction. The systolic spikes follow the R-spikes with a certain delay (cf. Figure 2.6) because the blood needs some time to travel from the heart to the location of the PPG measurement (often the fingertip) [23].

2.2 Cardiac Arrhythmia

Cardiac arrhythmias occur among most people. They can be clinically relevant, but also harmless. Moreover, they are determined based on ECG or PPG recordings. The waveforms and distances between two R-spikes or systolic peaks are analyzed for this. Arrhythmias are divided into those in which the centre of excitation is in the sinus node and those which have their place of origin either in the supraventricular area of the sinus node or in the ventricular area [25]. Heart rates are also divided

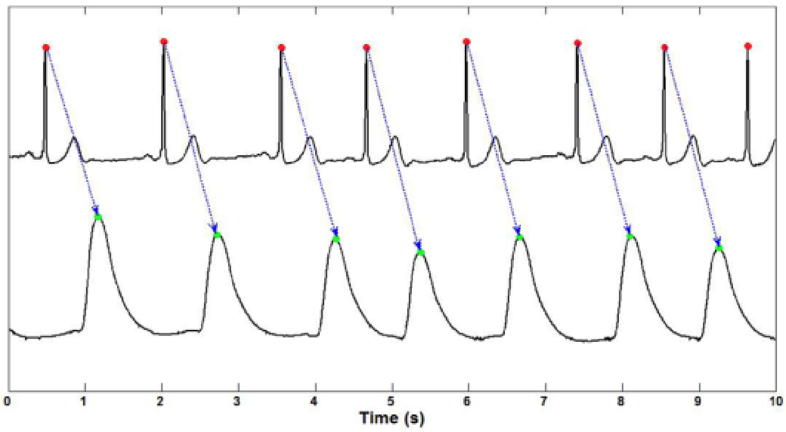

Figure 2.6: Relationship between ECG and PPG signals. The peak in the PPG waveform is slightly delayed compared to the peak of the ECG [24].

into those that are too fast (tachycardia) and those that are too slow (bradycardia). Furthermore, fibrillation can occur. Cardiac arrhythmia is defined as a pathological heart rhythm that is not physiologically adequate. Bradycardia during sleep is therefore not pathological, but during physical exertion it is [19].

In the following, some cardiac arrhythmias are presented that can be detected in both single-lead ECGs and PPGs.

Premature Ventricular Contractions: Premature Ventricular Contractions (PVC) are indicated by a wide and abnormal QRS-complex (see Graph **b)** in Figure 2.7). This complex occurs earlier than the expected systole in the cardiac cycle. Additionally, there is no P-wave preceding the QRS-complex. The PVC replaces a sinus beat and induces a delay in the next sinus beat. Thus, the RR-interval is prolonged. Because of this delay, the next sinus beat follows the basic rhythm, with the interval between the sinus beat that preceded the extrasystole and the next beat equaling two RR-intervals. The extrasystoles are harmless at first. However, they can cause persistent ventricular tachyarrhythmias.

PVCs can be classified further. For instance, there are PVCs in bigeminy (every second beat is a ventricular extrasystole) or couplets (two consecutive PVCs). If there are more than 30 consecutive PVCs and the heart rate exceeds 100 beats per minute, it is called sustained ventricular tachycardia [18].

Premature Atrial Contraction: Premature Atrial Contractions (PAC) leads to premature appearance of a QRS-complex, which has a normal shape. The P-wave has a different morphology compared to the Sinus-P-wave (Graph **c)** in Figure 2.7). Like PVCs, the extrasystole leads to a delay in the next sinus beat, which also prolongs the RR-interval. However, in PACs, the basic rhythm is not maintained. The interval between normal beats before and after the PAC is shorter than two RR-intervals. PACs are also usually harmless at first, but they can cause sustained supraventricular tachyarrhythmias such as atrial fibrillation [18].

Ventricular Tachycardia: Ventricular tachycardia (VT) (Graph **d)** in Figure 2.7) can be triggered by many causes, thus it cannot be determined by a specific characteristic. The heart rate of a VT is between 100 and 250 beats per minute. Other characteristics include more than three consecutive ventricular beats, which in turn are typified by a wide QRS-complex.

VT can be further subdivided into, eg., monomorphic and polymorphic VT. In this context, a monomorphic VT is characterized by a consistent morphology of the QRS-complex. On the other hand, a polymorphic VT has different patterned QRS-complexes and may also have a varying rhythm [18].

A VT can be confused with supraventricular tachycardias with wide QRS-complexes. Because VTs are potentially life-threatening, it is essential to distinguish VTs from supraventricular tachycardias. One way is to examine the onset of tachycardia. Irregular RR-intervals are indicative of a VT [19]. Furthermore, other features can be observed on the ECG for differentiation. However, these will not be discussed in detail because they are not relevant to this work.

Supraventricular Tachycardia: Supraventricular tachycardia (SVT) includes all tachycardias originating from a circuit or focus involving the atria or atrioventricular node. The term paroxysmal SVT refers to a subset of SVTs characterized by fast, regular tachycardias with abrupt onset and termination. The term SVT is synonymous with paroxysmal SVTs in this work. SVTs are characterized by narrow complexes. In most cases, the P-wave is hidden in the QRS-complex and thus not visible. In addition, an SVT is usually characterized by a short RP-interval. SVTs include atrial flutter, sinus tachycardia, and atrioventricular nodal re-entrant tachycardia (AVNRT) [26].

Atrial Fibrillation: Atrial fibrillation (AF) belongs to SVTs. However, AV is distinguished by an irregular rhythm, which is the reason why they are further differentiated from SVTs. In AFs, the P-wave is absent and the heart rate is irregular between 100 to 180 beats per minute. The line between the QRS-complexes is characterized by either fibrillation waves (f-waves) (Graph **e)** in Figure 2.7) or minor oscillations. The f-waves are small with varying morphology and amplitudes. AF is among the most common tachycardias [18].

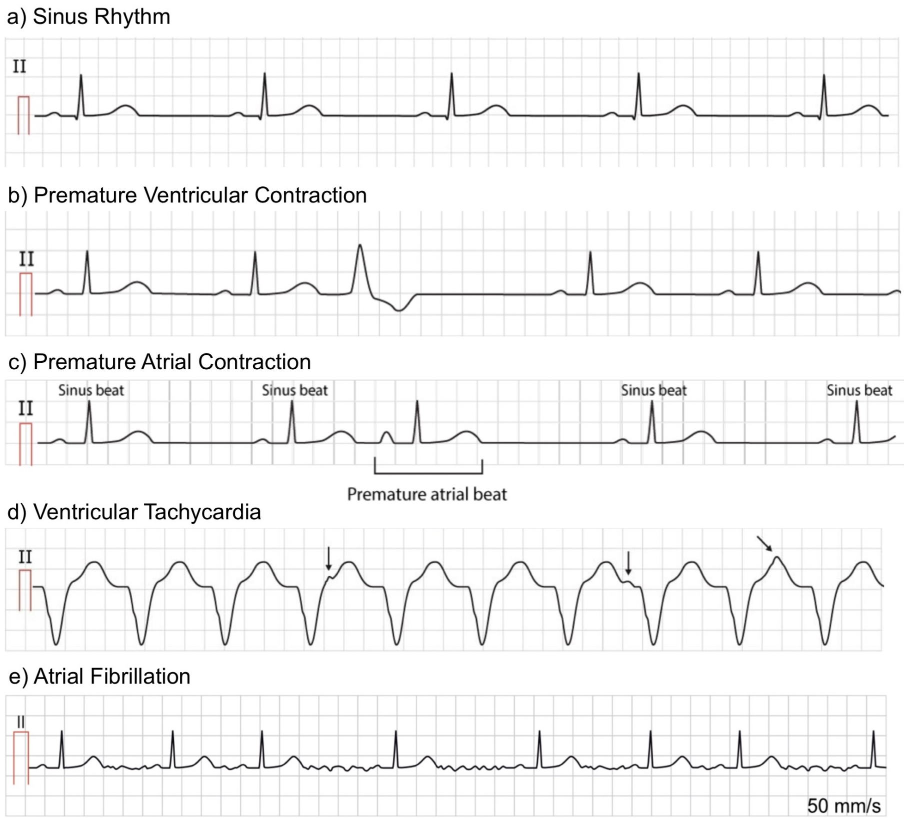

Figure 2.7: ECG signals with different arrhythmias. a) Normal sinus rhythm. b) Example of a premature ventricular contraction after two normal R-peaks. c) Typical premature atrial contraction. d) Ventricular tachycardia with P-wave depicted by the arrows. e) Atrial fibrillation with f-waves. [27].

2.3 Topological Data Analysis

Topological data analysis (TDA) has recently gained popularity through the work of Edelsbrunner et al. [28] and Zomorodian and Carlsson [29]. The underlying concept is that topology and geometry provide a powerful approach to obtain qualitative information about the structure of the data [30]. Topology represents a mathematical field that classifies topological spaces. Such spaces have the same topological properties within a class, so it is possible to discover the same structures in different spaces. Through high-dimensional data, patterns can be detected with the help of TDA.

One instrument of the TDA is persistent homology. It is a powerful tool for calculating, investigating, and encoding multi-scale topological features [30]. With the help of persistent homology, complicated structures can be described and are hence used in image processing and time series analysis [31]. In the context of arrhythmia classification, persistent homology is used to characterize the shape of ECG signals compactly. The TDA features are very robust to changes in the pattern of the ECG signal and are not affected by expansion and contraction in the direction of the time axis [13].

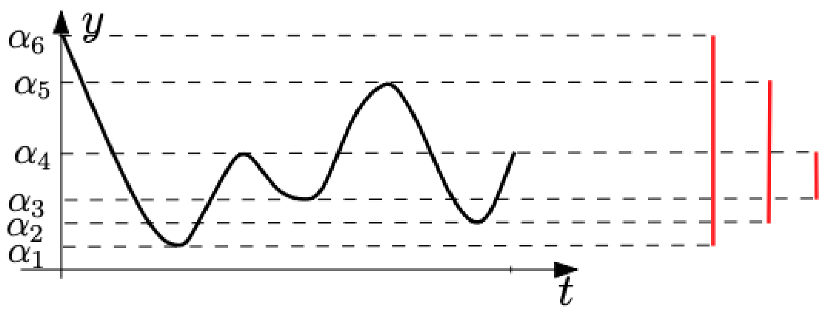

Figure 2.8: Persistence barcode of a 1D signal [13].

To characterize the heartbeats using persistent homology, the so-called sub-level (or upper-level) sets filtration can be applied to the time series. Here, the signal is considered as a function f defined in an interval I. For a given threshold α, related components

$$F_\alpha = \{t \in I : f(t) \leq \alpha\}$$

for the sub-level sets filtration and

$$F^{\alpha} = \{t \in I : f(t) \geq \alpha\}$$

for the upper-level sets filtration are obtained. As α increases (or decreases) some components appear and others are merged. The evolution of the components is encoded in a persistence barcode. In Figure 2.8 such a barcode is shown.

The start point of each interval corresponds to a value α at which a new component is created, while the endpoint corresponds to the value α' at which the created component is merged with another. Thus, the function f is the piecewise linear interpolation of the ECG time series [13].

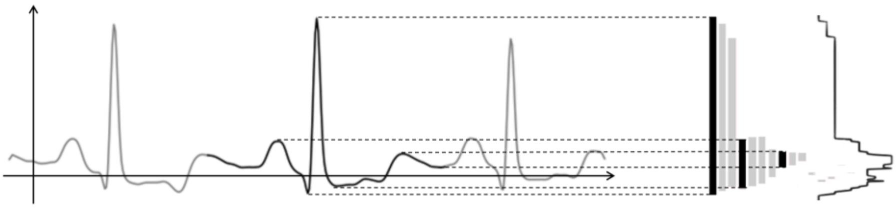

Figure 2.9: Generation of a persistent barcode and Betti curve from heartbeats [13].

The persistent barcodes provide important information. However, they cannot be used as input data for machine learning techniques because the barcode is a sparse image and the number of components of the barcode is not constant. Consequently, Umeda [32] presented an approach in which they derived the so-called Betti curves from the barcodes. The Betti curve represents the number of barcodes at time m. Formally, the Betti curve with homological dimension k is defined as

$$\beta_k(t) = \sum_{(b,d) \in dgm} w(b,d) 1_{t \in [b,d]}$$

where w is a weight function defined on Δ [30]. The Betti curves are calculated and discretized on the interval bounded by the minimum and maximum of the birth and death values of each persistent diagram. A fundamental property of the Betti curves of one-dimensional signals is the stability concerning the temporal reparameterization and the scaling of the signal value. This makes it possible to generate a uniform input for classical one-dimensional convolution models and to solve individual differences. Because persistence intervals measure the relative height of the peaks of the signal and not the width, the invariance to time scaling is guaranteed [13]. Figure 2.9 shows the process of barcode generation from three consecutive heartbeats and the subsequent retrieval of the Betti curve.

3 State of the Art

In the following, the current State of the Art in arrhythmia detection and classification based on ECGs and PPGs is reviewed. Different prerequisites of the various approaches are explained and thus, a sorting of the methods is undertaken. In addition, an overview of frequently used data sets is given.

Data-driven identification and classification of arrhythmias are primarily based on the analysis of ECGs and PPGs (cf. Section 2.1). In 2021, Neha et al. [7] gathered the results from various papers published in the past on the classification of arrhythmias. In the process, they also examined the number of articles that use ECG and PPG sensors to classify arrhythmias. In 2021, 1209 publications use ECG data and 109 papers are based on PPGs. This can be explained by the fact that PPGs have gained popularity within the last few years. Especially the development of semiconductor technology, i.e. light-emitting diodes (LED) and photodiodes, has significantly improved the reliability of the technology [23].

The individual approaches can be subdivided based on the technological prerequisites. For ECGs, a distinction is made between single-lead sensors and multiple-lead sensors. Most publications use single-lead sensors (812). Furthermore, the publications can be grouped by the number of classified labels. Some approaches examine the recordings for a specific cardiac arrhythmia. Other approaches perform a multi-class classification. In addition, there are also differences in the methodology used. Older approaches particularly use descriptive statistics or classic machine learning models. Newer approaches, on the other hand, often use deep learning, which enables them to achieve automatic feature generation. In addition, there are performance differences in some approaches, as the models created are partly tuned to patients and do not work for new individuals. The suggested classification of the different concepts from the literature can be seen in Figure 3.1.

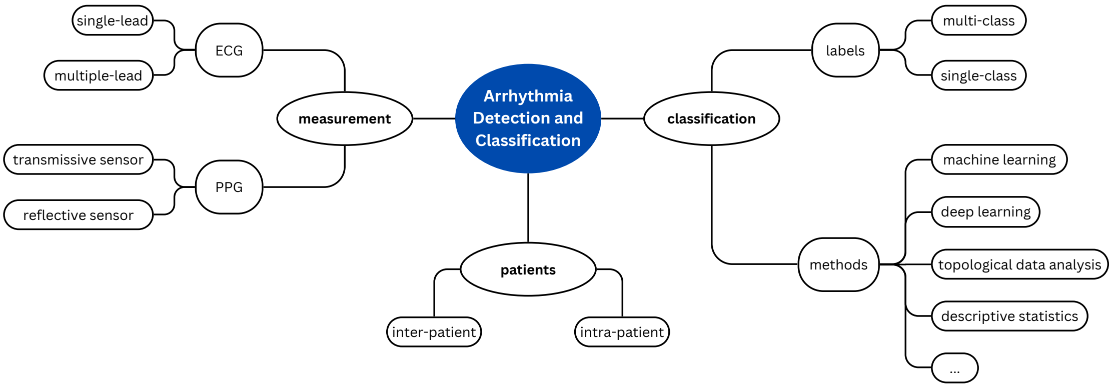

Figure 3.1: Different aspects and factors of the methods for arrhythmia classification presented in previous work. Publications are divided into measurement technique, patient selection, number of labels and classification methods.

In the following, first several existing data sets are introduced and afterwards the individual classification differences are discussed in more detail and existing work is presented.

3.1 Data Sets

There are several data sets that contain cardiac records. In the following, some of these data sets are presented and the advantages and disadvantages are highlighted.

The most commonly used database is the *PhysioNet* database. It was initiated by the Massachusetts Institute of Technology and contains data from Beth Israel Hospital (MIT-BIH). Since 1999, it has provided free access to ECG and PPG arrhythmia data, among others. The ECG arrhythmia data contained in the MIT-BIH Arrhythmia Database includes 48 data sets from 47 subjects. The frequency of the signals is 360 Hz. The majority of the signals are acquired from lead II. Others from leads V2, V4, or V5. In addition, the database contains PPG signals found in the MIMIC-II and Challenge 2015 and 2017 data sets. Furthermore, there is one ECG signal per PPG signal. The signals are sampled at a frequency of 125 Hz. The PPG recordings are not labeled but can be assigned a label using the ECG data. The labels included are listed in Table 3.1. Heartbeats are divided into five supercategories and then specified in more detail in subcategories [7]. In addition, there is a normal sinus rhythm data set (MIT-BIH Normal Sinus Rhythm Database), which contains 18 long-term ECGs. To complement the supraventricular arrhythmia data in the arrhythmia database, there is the MIT-BIH Supraventricular Arrhythmia Database with 78 half-hour ECG recordings and the MIT-BIH malignant ventricular ectopy database.

Group	Symbol	Class
N	N ou.	Normal beat
any heartbeat not	L	Left bundle branch block beat
in other group	R	Right bundle branch block beat
	e	Atrial escape beat
	j	Nodal (junctional) escape beat
SVEB	A	Atrial premature beat
supraventricular	a	Aberrated atrial premature beat
ectopic beat	J	Nodal (junctional) premature beat
	S	Supraventricular premature beat
VEB	PVC	Premature ventricular contraction
ventricular	E	Ventricular escape beat
ectopic beat		
F	F	Fusion of ventricular and normal beat
fusion beat		
Q	P ou/	Paced beat
Unknown beat	f	Fusion of paced and normal beat
	U	Unclassified beat

Table 3.1: Principle types of heartbeats presented in the MIT-BIH database [8].

The American Heart Association (AHA) generated a database of eight arrhythmia classes and normal sinus rhythm from more than 6300 hospitals and 400 health systems. The signals are 12-lead ECGs sampled to 250 Hz. However, the database is not publicly available and does not contain PPG data [7].

A database for multichannel ECGs is the UCI Arrhythmia Database. The data is from 1998 and contains ECG signals from 452 instances. There are 16 heart rhythm classes and additional patient information such as age, gender, height, and weight. Associated PPG records are not available.

A European community created a database in 1985 known as the European ST-T Database. The database contains 90 annotated ECG recordings from 79 patients. Two channels are documented at the time of recording. The database only records ST and T-wave changes (cf. Figure 2.2) and also does not contain PPG data [7].

A PPG data set was developed by Liu et al. [4] in the context of their publication *Multiclass Arrhythmia Detection and Classification From Photoplethysmography Signals Using a Deep Convolutional Neural Network*. They recorded 118,217 PPGs of ten seconds each from a total of 228 patients. There are six classes of arrhythmia: sinus rhythm, PVC, PAC, VT, SVT, and AF. In total, there are 38,081 labelled sinus rhythms, 11,372 PVC, 11,248 PAC, 5783 VT, 12,539 SVT and 39,194 AF. PPG data are downsampled to 100 Hz. 60% of patients are randomly assigned to the training set. This corresponds to 137 patients and 71,390 segments. For validation and testing, 20% of the data are used respectively. This equals 46 and 45 patients with 23,443 and 23,384 PPG segments. The training data are not publicly available. However, validation and test data sets are made available, giving a data set of 46,827 signals.

3.2 Classification with Electrocardiography

As explained in the introduction of this chapter, single-lead ECGs or multi-lead ECGs can be used to classify arrhythmias. Single-lead ECGs allow constant monitoring of a patient because, in contrast to multi-lead ECGs, one device is sufficient to generate an ECG. The disadvantage, however, is that not all cardiac arrhythmias can be detected with a single-lead ECG.

Rahhal et al. [33] introduced a publication based on multi-lead ECGs and use stack denoising autoencoders (SDAEs) to learn feature representations from ECG data. After the feature learning phase, the authors add a Softmax regression layer on top of the resulting hidden representation layer, creating a deep neural network (DNN). During the interaction phase, an expert labels each iteration's most important and uncertain ECG beats. These are then used to update the DNN weights. The MIT-BIH Arrhythmia and SVBD data sets, which contain two leads, and the INCART data set, which has 12 leads, are used as the data set (cf. Section 3.1).

Rahhal et al. present the results in the form of (class PVC to [normal beat, S, and F]) and (class S to [normal beat, PVC, and F]). The authors achieve an overall accuracy of 98% (MIT-BIH and SVDB) and 99% (INCART). A major disadvantage of the approach is the time-consuming manual labeling of an expert.

Irfan et al. [34] solved the problem and developed a deep learning framework that integrates different networks by stacking similar layers in each network. The network are tested on two data sets: The UCI Arrhythmia data set and the MIT-BIH Arrhythmia data set (cf. Section 3.1). Several ECG-leads of the data are used and it is a multiclass approach. The presented framework reaches an accuracy of 99.35% in the classification of five arrhythmia subclasses (N, L, R, A, and V). Despite the partial inaccuracy of single-lead ECGs, they are better suited for monitoring many patients in an ambulatory setting. Moreover, the classification possibilities of single-lead ECGs are the same as those of PPGs. For this reason, this work focuses on single-lead ECG approaches.

Different types, but also different numbers of arrhythmias are classified in several publications. For example, while Rahhal et al. [33] performed two binary classifications, Irfan et al. [34] categorize five sub-classes of arrhythmias in a multiclass environment. In some applications, it is convenient to perform binary/one-class classification. For example, Zhang et al. [35] classify atrial fibrillation. This is useful because some individuals do not notice irregular and rapid heartbeats. However, atrial fibrillation can lead to stroke, making detection necessary. Nonetheless, there are high risks associated with other arrhythmias as well, making a multiclass classification preferable. Therefore, the focus of the following work is on multiclass classification.

Previously, mainly machine-learning-based methods are used for arrhythmia classification. A major disadvantage of such approaches is the need for manual feature recognition, complex models, and long training times [36]. In addition, heartbeats can have different morphologies even within the same patient, which are difficult to capture with hand-crafted features [37]. Automatic feature engineering built into deep learning can capture these different morphologies invisible to the human eye [38]. Therefore, deep learning has become increasingly attractive in recent years, as it obviates the need for manual feature recognition. A study on the use of deep learning was published by Wu et al. [36] in 2021. The authors present a 12-layer deep one-dimensional convolutional neural network (CNN) to categorize five micro-classes. They use the MIT-BIH Arrhythmia database. All heartbeats from one channel are extracted individually. The ECG signals are denoised using the wavelet transform method. The wavelet transform is an algorithm that decomposes non-stationary signals into scale signals of different frequency bands. Finally, the pre-processed ECG signals are fed into the CNN and a class is determined. The network achieves an accuracy of 97.41%.

Rahman et al. [3] presented an approach in 2022 whereby they use transfer learning to categorize five arrhythmia classes. They compare three models: ResNet, SqueezeNet, and AlexNet. The data set published on *Kaggle*, which in turn is based on the MIT-BIH Arrhythmia data set, serves as the data basis. The data are additionally augmented to increase the number of images in the data set. In addition, the data has been preprocessed before being passed on to the three models. Several layers are reused and the last three layers are re-trained. The ResNet achieves an accuracy of 97%, the SqueezeNet 75%, and the AlexNet 96%.

The two previously presented approaches (Wu et al. [36] and Rahman et al. [3]) carry out extensive preprocessing steps. These are time-consuming and costly. Hannun et al. [5] provided a deep learning model that does not require any substantial preprocessing such as wavelet transform or equivalent. With their model, they achieve a performance similar to cardiologists. The single-lead ECGs can be classified into twelve categories. Thereby Hannun et al. [5] stand out in terms of the number of labels compared to other publications. The authors also created their own data set consisting of 91,232 ECG records from 53,549 patients. The model achieves an average area under the receiver operating characteristic curve (ROC) of 0.97. The CNN consists of 34 layers and accepts raw ECG data as input. The model is tested on the 2017 PhysioNet Challenge data. It achieves an F1-score of 0.83 and is thus among the best performers in the PhysioNet Challenge. Despite the generalizability of the model, differences between individuals are not explicitly addressed. The model could perform worse on data from new subjects.

Therefore, in 2018 Mousavi and Afghah [39] developed a concept for the inter- and intra-patient heartbeat classification. They also make use of the MIT-BIH arrhythmia data set. Channel II of the ECG is selected. The intra-patient approach combines training and evaluation data from the same patients. In contrast, the evaluation data set is retrieved from new patients in the inter-patient approach. This makes the model more suited to reality, where data from new patients is classified using the existing model. In addition, the authors address the challenge of imbalanced data. There are underrepresented classes in the data sets. Thus, Mousavi and Afghah developed a sequence-to-sequence deep learning model and an oversampling method called Synthetic Minority Over-sampling Technique (SMOTE). First, preprocessing is done to extract individual heartbeats. Then the preprocessed heartbeats are fed into a CNN. Afterwards, the results of the CNN are used as the input to a RNN sequence to sequence model. The sequence to sequence model is designed based on the encoder-decoder idea. The encoder encodes the input sequence, while the decoder computes the category of each input. When testing the intra-patient concept, training and test data are randomly selected. In the inter-patient approach, however, the data is split based on the patients, whereby only ECGs from patients who do not appear in the training data are used for the test data. The model achieves an accuracy of 99.92% in the intra-patient setting and 99.53% in the inter-patient scenario. The complexity in Mousavi and Afghah's work is to generate synthetic data using SMOTE to compensate for minority classes. In reality, this can lead to difficulties due to the above-average representation of sinus rhythm. An approach that deals with imbalanced data differently and also take the variations between individuals into account was presented by Dindin et al. [13] in 2020.

Dindin et al. [13] present an approach that classifies different arrhythmias and performs well in new patients. They use a multichannel neural network and combine topological data analysis, handcrafted features, and deep learning. At the same time, they address the challenge of unbalanced data. The available training data for sinus rhythm is higher than for arrhythmias. The authors use several data sets from the *Physionet* platform. Among them are MIT-BIH Normal Sinus Rhythm Database, MIT-BIH Arrhythmia Database, and MIT-BIH Supraventricular Arrhythmia Database. Among the data, one lead of the ECG recordings is used.

The first step is to create an autoencoder that learns sinus rhythm using unsupervised learning. The autoencoder can then be used to classify the data in binary terms. Afterwards, data points labeled as abnormal are classified in a second step. Betti curves (cf. Section 2.3) are generated using topological data analysis, which compensates for individual differences in the data and thus makes the model usable for new patients. In addition, the ECG signal is processed in a CNN. Together with handcrafted features and discrete Fast Fourier Transform, the data is labeled. The model structure is shown in Figure 3.2. The authors achieve an accuracy of 90% on the test data set for the detection of arrhythmias (autoencoder). For the classification of 13 classes, an accuracy of 80.5% is achieved. In addition, the approach is tested with eight classes used in other papers. The model has an accuracy of 99% when classifying eight arrhythmias.

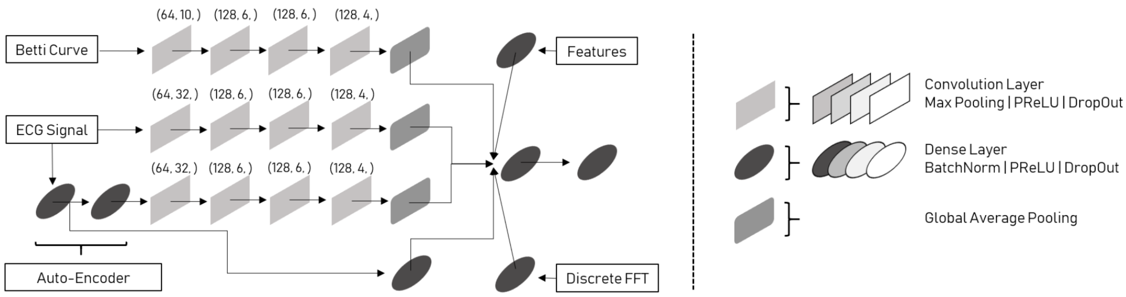

Figure 3.2: Model structure presented by Dindin et al. [13]. They created a multi-channel neural network which consists out of an autoencoder, the generation of Betti curves that are further processed in convolutional layers, handcrafted features, discrete fast fourier transform and convolutional layers that are processing the ECG signal itself.

In their paper, the authors present the autoencoder and the Betti curves in more detail. However, it is not possible to retrieve more information about the handcrafted features that influence the accuracy. Besides, the creation of handcrafted features is time-consuming, so an approach without feature generation should be considered.

3.3 Classification with Photoplethysmography

The trend of wearable devices has also increased the demand for wearable sensors that can permanently monitor heart rhythm. PPG is a suitable solution because it is inexpensive, non-invasive, and does not require a large number of electrodes to be reattached after a while. In 2021, Neha et al. [7] concluded that previous classification methods focus on binary class classification. An example of such a method was presented by Polania et al. [9]. The authors classify PVC and VT from normal sinus rhythm and SVT. They first developed a preprocessing stage. Then discriminative features are extracted and used as input to a Support Vector Machine (SVM). The extracted features originate from the time-domain and frequency-domain context. In addition, non-linear dynamics are explored. For all cases, high accuracies above 90% are achieved. Nevertheless, there is a need to develop methods for multiclass classification. The same conclusion was drawn by Neha et al. [7].

In the last year, work on the multiclass classification of arrhythmias has been made. For example, Neha et al. [40] use dynamic time warping (DTW) to detect four arrhythmia classes in PPGs. Automatically generated warping features are extracted through DTW. Subsequently, the classification is carried out with a feed-forward neural network. The PhysioNet MIMIC-II data set serves as the data basis. The class distribution of the data set is unbalanced. There are 400 normal PPG signals and 90 signals for PVC, VT, and AF respectively. The authors first implemented the preprocessing. Besides removing the outliers, a second-order Butterworth low pass filter is applied. Moreover, the signals are normalized. Then DTW is used for feature extraction. In this process, warping features are determined by mapping two PPG signals and calculating the optimal warping path. For arrhythmia detection, a normal PPG signal is warped with an arrhythmic signal. These changes in the optimal warping paths observed when mapping normal-normal PPG signals and normal-abnormal PPG signals are used to identify the arrhythmia type. The resulting features are passed to a neural network. The network consists of three hidden layers, whereby the first two layers consist of 100 neurons and the last layer contains 50 neurons. Three arrhythmia types and sinus rhythm are classified. The method achieves an accuracy of 95.97%. Manual feature extraction increases the workload and makes the model more complex. Therefore, it is reasonable to use deep learning.

An approach using deep learning to classify multiple arrhythmias was presented by Liu et al. [4] in 2022. They developed a deep convolutional neural network (DCNN) to classify six types of rhythms. First, they prepare the data by performing down-sampling, denoising, segmentation, and normalization. The DCNN is based on the VGGNet-16 architecture. It is adapted to be suitable for one-dimensional input signals. A self-made data set serves as the database. The PPG recordings of 228 patients are divided into ten-second segments. This produced a data set with

118,217 recordings. The following rhythms are recorded: sinus rhythm, PVC, PAC, VT, SVT, and AF. The data set is randomly divided into independent training, validation, and test data. Only signals from patients not present in the training data are used in the test data. The proposed method is therefore an inter-patient approach. The model achieves an accuracy of 85%. By contrast, Neha et al. [40] reach an accuracy of 95.97% for four classes on PPG data. The model offers potential for improvement in terms of performance.

In this work, the advantages of the presented publications are combined. A model is developed that can classify several types of arrhythmia by using deep learning on single-channel ECGs. An inter-patient approach is followed so that the model can also be transferred to new patients. At this point, TDA is suitable because individual differences are already taken into account when feeding the model. The finished model is then transferred to PPGs. The aim here is to develop a multiclass deep learning PPG approach that takes into account differences in patients and requires minimal manual feature engineering effort. The exact procedure and structure of the model are described in the following chapter.

4 Methodology

Based on the fundamentals and findings described in the *State of the Art* chapter, this chapter outlines the requirements for the model to be developed, the data selection and the system design. Finally, two further approaches are presented, which are evaluated in this work: The use of transfer learning and an ensemble of multiple binary classifiers.

The system is designed with the research questions under consideration. Firstly, an approach that includes and combines all aspects of previous systems is developed. In addition, the system is designed to be transferable to PPG sequences. The following aspects are addressed:

- The model can receive and process ECG and PPG signals.
- The system is designed to be trained on ECG sequences and fine-tuned on PPG sequences.
- Transferability to new patients is taken into account. This is achieved by using topological data analysis and ensuring good performance on new patients by dividing patients into training and testing.
- The input sequences are second-long signals and are not elaborately preprocessed. This is because the used PPG sequences do not have peak markers and labels are for the entire series. Thus, preprocessing steps to center the peaks cannot be implemented. In addition, detection in second-long segments is more practical for future applications in smart devices.
- Deep learning is used. This is because heartbeats can have different morphologies even within the same patient and the automatic feature engineering is able to capture these differences.
- The system design should incorporate the findings of the paper by Dindin et

al. [13]. The modular multichannel approach should be emphasized.

- The system is designed to accept single-channel ECG data as well as data from reflective PPG sensors.
- Multiple heartbeats and rhythms should be classifiable. The focus is on the classes present in the PPG data.

In the following, the data selection and the preprocessing are discussed.

4.1 Data Selection and Preparation

This work aims to develop a classification system for multiple arrhythmias in ECG and PPG data. The focus of the work is on single-lead ECGs, as they are more extensively available than multi-lead ECG databases. In Section 3.1, various single-lead ECG databases are reviewed, including the MIT-BIH Normal Sinus Rhythm Database, the MIT-BIH Arrhythmia Database, and the MIT-BIH Malignant Ventricular Ectopy Database. These databases are used to obtain sequences of normal sinus rhythm, different types of arrhythmias, and underrepresented arrhythmias such as PVC and SVT.

Reflective PPG data is a commonly used acquisition technique, and therefore this work focuses on the classification of multiple arrhythmias in reflective PPG data. Due to the limited availability of labeled PPG data, the publicly available validation and test data from the Liu et al. [4] publication are selected for our analysis. Whilst research on single-class classification with PPGs is well established, the objective of this work is the multiclass classification. However, the relatively few different labeled classes in the PPG data limit this study. Therefore, the ECG model is trained with the classes that occur in the PPG data, namely N, PVC, PAC, VT and SVT. Some of these classes are underrepresented in the otherwise large database of ECGs.

The ECG sequences are prepared by filtering out pacemaker-triggered sections. In addition only lead II is used, so corresponding sequences from patients in whom other leads are recorded are removed. This results in data from 86 patients, with sequences of different frequencies due to acquisition from different databases. All data are thus resampled to 360 Hz.

The ECG database contains labels for each heartbeat, with the label positioned at the peak of the ECG waveform. Labels are extended to all associated data points. The ECG data consists of a signal resampled to 360 Hz, a heartbeat label, and a rhythm label for each data point, with each signal having a duration of 30 minutes.

For this study, the signals are segmented into non-overlapping two-second sequences to generate input samples for the model. After applying the detrend function from the *SciPy* package, the signals are scaled between zero and one using the *scikit-learn* min-max preprocessing package. Scaling between minus one and one is also tested, but yield worse results. The label of a sequence is determined by examining all unique values from the list of all labels occurring in the two-second signal. Only sequences corresponding to the classes N, VT, SVT, PVC, and PAC are retained for this study. The decision to use two-second sequences is based on an assessment of peak occurrence and the number of data per class to ensure a sufficient number of samples per class. At least three consecutive heartbeats of the same class are required to define a rhythm. This definition of heart rhythm is considered when choosing the number of peaks that should occur in a sequence. Due to the different heart rates of the patients, the data does not always contain three peaks. Nevertheless, by using two-second segments, there are an average of three peaks per sequence.

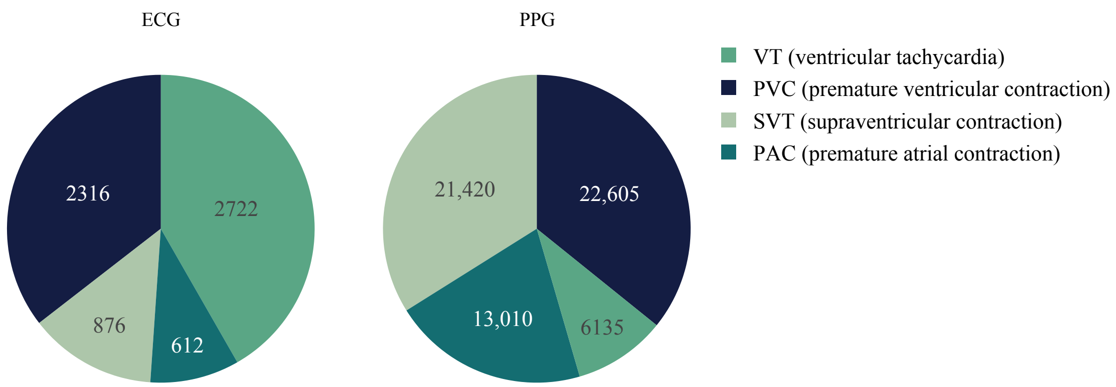

Figure 4.1: Number of total ECG (left) and PPG (right) two-second samples. The total number of ECG samples is lower than the number of PPG samples.

In this work PPG data from 91 patients that have already been preprocessed and normalized by Liu et al. [4] is used. The authors segmented the PPG signals into ten-second sequences, which were labeled based on ECG data recorded in parallel and classified by two physicians. In this work the sequences are resampled from 100 Hz to 360 Hz and divided into two-second sequences.

This work takes an inter-patient approach and separates the data by the patients. Training data consists of 80% of the patients, while the remaining 20% serves as test data. The approach is implemented to ensure that the patient-specific ECG morphology [41] is captured and the model works correctly with new patients. The data set includes ECG data from 86 patients and PPG data from 91 patients, resulting in a training data set of 69 and 73 patients, respectively, and a test data set of 17 and 18 patients. Figure 4.1 shows the number of samples per class, excluding class N due to its overrepresentation. The sinus rhythm is represented by 672,141 ECG samples and 77,755 PPG samples.

4.2 System Design

This work addresses the first research question by extending previous approaches to detect arrhythmias from ECG and PPG signals using a two-stage deep learning model (cf. Figure 4.2). The model is designed to identify normal and abnormal rhythms and subsequently classify the abnormal signals into various arrhythmia classes. The model components are chosen based on the publication by Dindin et al. [13]. However, the division into two separate networks and the architectures of the models differ from their approach, allowing to classify normal and abnormal signals before processing further the abnormal signals. This separation has the advantage that sequences labeled as normal are not processed further, thus saving computing power.

The present work uses an autoencoder to distinguish between normal and abnormal sequences. The anomalous signals are further handled by a two-channel network consisting of a Betti-CNN and a CNN. The Betti-CNN generates two Betti curves from the input signal, which are then processed through convolutional layers. The CNN operates directly on the signal without any prior processing.

This work combines the latter two models into a multichannel framework. Unlike the ensemble model, where multiple models are trained to predict a single class,

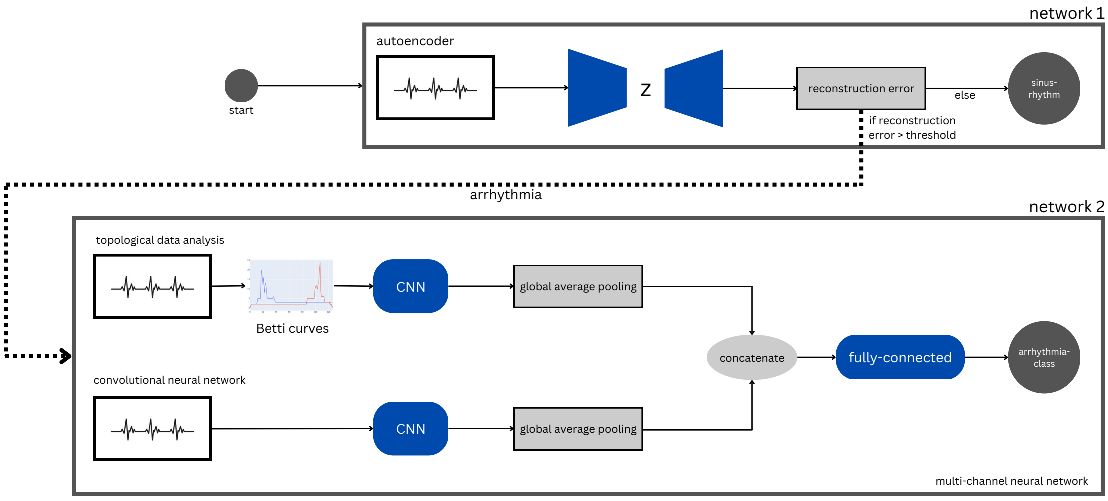

Figure 4.2: Architecture of the proposed model. Network 1 is pre-trained on normal sinus rhythms. Afterward, the network weights are frozen. Each input then passes through the first network. If the reconstruction error is greater than a threshold, this indicates an arrhythmia and the input is processed further in the second network. Otherwise it is classified as sinus rhythm.

the multichannel model creates different input channels that are part of a single model. Each input channel corresponds to a different aspect or type of input data that is processed by different parts of the model. The different input channels are combined at varying stages of the model to generate a common prediction. Specifically, this work combines information from Betti curves with information from raw ECG or PPG signals. In contrast, the ensemble model involves training multiple models with different parameter settings while the input to the network remains constant. Each model generates its prediction before a final prediction is selected using various techniques. This type of model is also evaluated in this work, along with the multichannel approach integrated into the overall model architecture. The details of this ensemble model are presented in Section 4.4. The individual components of the model structure are described in more detail in the following.

The two-step model architecture starts with an autoencoder network that distinguishes between normal and abnormal heartbeats and -rhythms. The separation of the model from the rest of the model architecture is a response to the challenge of an imbalanced data set. Due to the prevalence of sinusoidal rhythms, there are more normal training samples, which are used to train the autoencoder in an unsupervised manner [13]. The model is trained on normal sinus rhythm to learn its structure. During training, the input signal is converted into a low-dimensional representation. The original signal is then reconstructed from this representation. The differences between input and output are minimized using the mean squared error as the loss function.

After training, the weights of the model are frozen. If an abnormal rhythm is introduced into the network during inference, the model is not able to reconstruct it well and the reconstruction error is larger than for normal sequences. The mean squared error is chosen to take greater account of larger errors. This penalizes erroneous reconstructions of peaks in the ECG data. To distinguish between normal and abnormal sequences, a dynamic threshold based on the loss distributions of normal and abnormal samples is used. This threshold is determined by identifying the point of intersection of the loss distribution of normal and abnormal sequences. Any sample with a reconstruction error below this threshold is classified as normal, while others above are considered abnormal. Samples classified as abnormal are processed further in the second network.

The second step of the approach involves a two-channel network. The first channel, the Betti-CNN, generates two Betti curves, which are processed in convolutional layers. The two generated curves each of size 128 represent the upper/lower set filtration (cf. Section 2.3). Each of these curves are further processed in parallel in an one-dimensional CNN. The second channel digests the signal directly in an one-dimensional CNN. After applying global average pooling to both channels respectively, they are concatenated and further processed in fully connected layers to determine the arrhythmia class. It should be noted that due to possible misclassifications of regular sinus rhythms in the first stage, they must also be considered as a label in the second network. The purpose of the combination of the two channels is to extract features from both models which will be critical for classification. In addition, models that tend to overfit can be compensated. Each model of the second step is implemented and tested individually before being combined to the multichannel network.

4.3 Transfer Learning

The proposed model structure is designed to accept both ECG and PPG signals as input data. In addition to developing separate models for each data type, transfer learning (TL) techniques are also explored. The aim is to learn features from ECG samples to subsequently transfer the generated knowledge to PPGs. As the waveforms of the ECGs and PPGs are related to each other, the PPG models take advantage of the previously learned features. The relationship between the two waveforms has already been discussed in Section 2.1. Before applying TL to the multichannel model, TL is applied separately to each model component for better evaluation. In addition, each model is also trained on PPG data only to get reference results.

There are several TL approaches involving freezing or fine-tuning the weights of individual steps based on the similarity of the data and the amount of data available. Freezing refers to maintaining the weights of the model during training, while fine-tuning allows the weights to be adjusted. Depending on the similarity and amount of data, individual levels can be frozen or the entire model can be trained. The latter approach is used when sufficient training data is available for the new task, allowing the information already learned from the first task to be used and the weights to be further adjusted with new data [42].

In this work, the database of PPGs is found to be better than the ECG database, making the fine-tuning approach suitable. In general, however, the TL approach is intended to compensate the generally poorer database of PPGs by pre-training on ECGs. However, some of the arrhythmia classes used in this study are underrepresented in the otherwise large ECG database. Therefore TL approaches requiring many samples of both databases can be applied. The present work investigates different TL strategies for all models. Specifically, in the autoencoder model, parts of the encoder or the decoder can be frozen while other layers are fine-tuned. In this work, the weights of the encoder are fixed and only the decoder is tuned using PPG data. In the second model, two different approaches are used and compared. In the first approach, only the fully connected layers of each model are tuned, while the weights of all other layers remain fixed. The second approach involves further training all layers of the model on the PPG data.

4.4 Ensemble of Multiple Binary Classifiers

The network structure of the second model is originally designed to implement multiclass classifiers. However, an ensemble of multiple binary classifiers is more suitable due to the underrepresentation of individual classes in the ECG data. The ensemble allows each classifier to focus on detecting patterns in a specific class, without being affected by the imbalance of other classes. The exact difference and architecture are highlighted below.

In the ensemble of multiple binary classifiers, the original multiclass problem is divided into subtasks whose combined output performs the classification [43]. The idea behind this is that binary classification tasks are easier to solve than multiclass tasks. However, the challenge is to combine the binary classifiers to ensure correct classification afterward [44].

There are two strategies for binarizing multiclass problems. One approach is one-vs-all classification. Here, for m classes, m binary models are created that separate each class from all others. The other strategy is one-vs-one classification. For this, m classes are divided into $m(m-1)/2$ binary problems. Each binary problem learns to distinguish two classes from each other [44]. In this work, the one-vs-all approach is followed because fewer models need to be generated, and thus the computational effort is lower.

There are also different approaches to merging the individual binary classifiers. Two methods for one-vs-all classification are presented in the following. The simplest strategy is the voting strategy. The class of the classifier with the highest positive response is taken. This approach is also useful for a multilabel problem where multiple classes can be output. For this, the list of all positive predictions is returned [44]. Another strategy was presented in 2008 by Hong et al. [45]. The authors trained a Naive Bayes classifier along with all other binary models. The classifier determines the order in which the binary classifiers are processed. The class whose model returns a positive value first is considered. This is done dynamically for each sample. This means that the confidence of the individual models is not relied on. Since a multilabel problem is interesting for the use case of this work, the voting strategy is used.

The combination of several binary models is of interest for network two. This work compares classification by a multiclass model with an ensemble of several one-versus-all binary classifiers for the Betti-CNN as well as for the CNN and the multichannel network. The network architectures are the same. The only difference is the generation of five binary models per channel and their subsequent combination. The resulting architecture for the CNN is shown as an example in Figure 4.3.

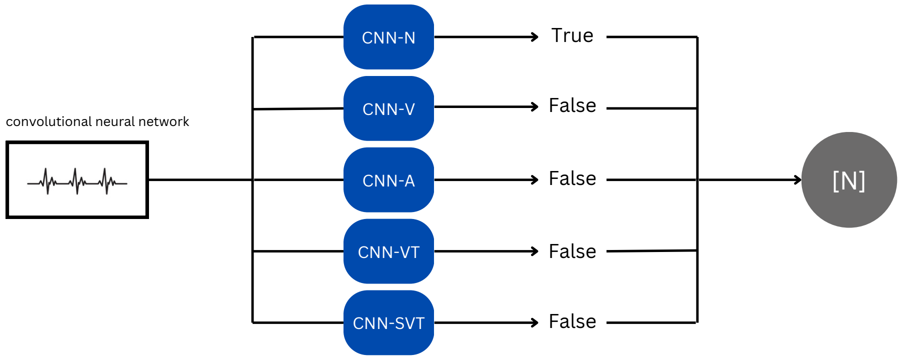

Figure 4.3: Architecture of the CNN with an ensemble of binary one-vs-all classifiers to predict the heartbeat or -rhythm class. All labels which models predicted True are returned.

5 Implementation and Results

This chapter deals with the implementation of the models and different approaches described in the previous chapter. First, each of the components is considered individually, before the implementation of the overall model structure is discussed. The frameworks used are also presented. The implementation and results are introduced in the following order. First, the models are implemented for ECGs. Then, PPG data are applied to the model structure. Next, the transfer learning approach is applied and the results are compared with the PPG-only approach. Finally, the ensembles of binary classifiers for both ECG and PPG data are described.

5.1 Frameworks and Programming Language

The models and approaches are all implemented in *Python* and organized using *Kedro*. *Kedro* is an open-source *Python* framework for writing maintainable and modular data science code. The framework allows the generation of pipelines that can be used to structure different experiments. *Kedro* provides a folder structure for each pipeline so that the code structure is also standardized and easy to maintain. An integrated data catalog allows data to be loaded and saved in various file formats. Through the data catalog, developed models can also be versioned [46].

A total of 20 pipelines have been created in this work. Three of them contain steps of data preprocessing. The components of the designed overall architecture are first considered individually in different pipelines each for ECG and PPG data. Then another pipeline is created for the multichannel network. Separate pipelines for each model are built for transfer learning as well as for implementing the ensemble of binary classifiers. For the latter, the pipelines are designed individually for ECGs and PPGs. Despite the partially reusable code, the division into different pipelines has the advantage that the experiments can be run individually and the associated models can be versioned without having to adapt the code. This allows the best-performing model and approach to be extracted from the code as an independent pipeline for further work.

The models are all developed using *PyTorch*, an open-source machine learning framework that can be used for both research prototyping and production deployment [47]. The training of the models take place on the CPU of a *MacBook* from 2020 with an M1 chip and 16 GB RAM.

5.2 Multiclass Model

When implementing the system described in the previous chapter, the various models are first implemented individually. The implementation of the autoencoder, the CNN, and the Betti-CNN are described in the following.

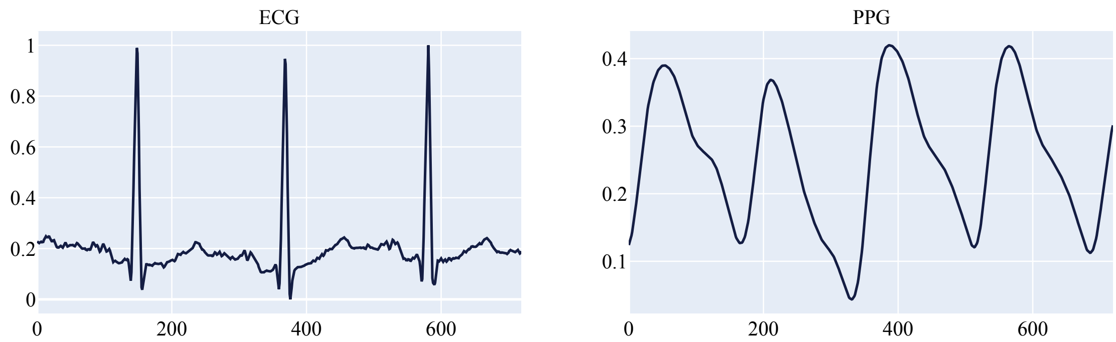

Figure 5.1: Exemplary samples of normal sinus rhythm recorded with an ECG (left) and a PPG (right).

The autoencoder is trained on samples of normal sinusoidal rhythms. Each sample has an input size of 720 data points. Examples of ECG- and PPG-samples are shown in Figure 5.1. The encoder and decoder each consist of five layers. Each layer includes a linear layer and a rectified linear unit (ReLU) activation function. The first four layers of the encoder also contain batch normalization. In the last decoder activation function, ReLU is replaced by a Sigmoid function. This is because the input signals are normalized between zero and one. Since the Sigmoid function also returns values between zero and one, the reconstruction of the signal is reduced to this range of values. The exact network architecture is shown in figure 5.2.

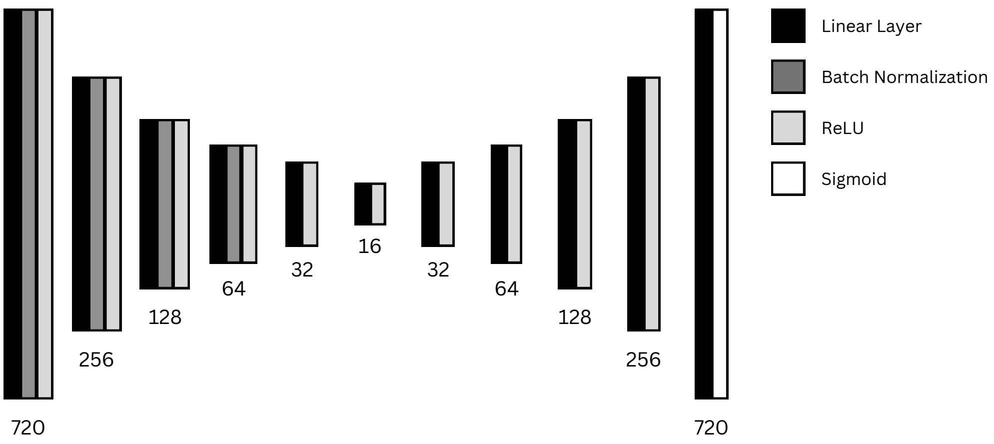

Figure 5.2: Architecture of the autoencoder identical for ECG and PPG data.

The Xavier initialization is used to initialize the weights, and the Adam optimizer is selected to update the network weights. The learning rate for training is 1e-3 and the weight decay is 1e-8. The autoencoder is trained with a batch size of 128 for 1500 epochs. The Mean Squared Error is chosen as the loss function. Using this loss function, larger deviations in the reconstruction are weighted more heavily, thus penalizing incorrect reconstruction of peaks.

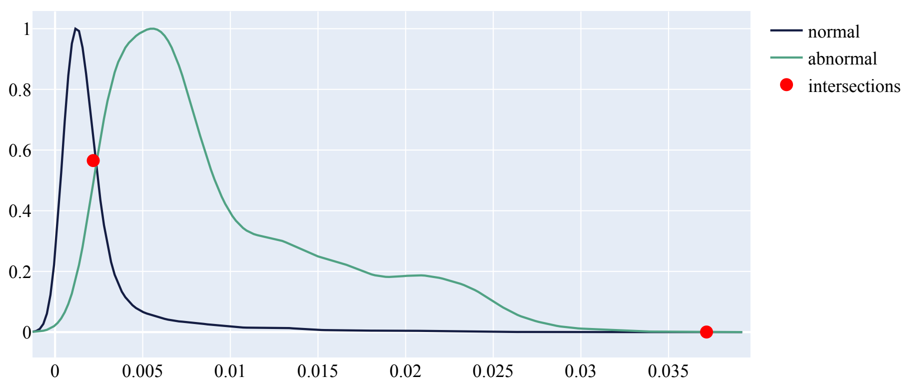

Figure 5.3: Distribution of normal sequence and abnormal sequence losses of ECG data with intersection points to determine the threshold for classifying the sequence in normal or abnormal.

After training the model, weights are frozen and the threshold is determined using the test data. This is done by calculating the loss of all normal and all abnormal samples separately. The distributions of the losses are then considered and the intersection between the normal and abnormal distributions is calculated. Thereby the frequencies are scaled between zero and one. The presentation of the loss distribution curves for ECG data with the intersections can be seen in Figure 5.3. The intersection point furthest to the left is set as the threshold, resulting in a threshold value of 0.002.

Two kinds of errors can occur when classifying using the threshold. Either a normal rhythm is classified as abnormal or an abnormal rhythm is classified as normal. The latter requires more attention in this context, as there are greater risks associated with such a misclassification. Figure 5.4 shows the confusion matrices of classifications with the errors described above.

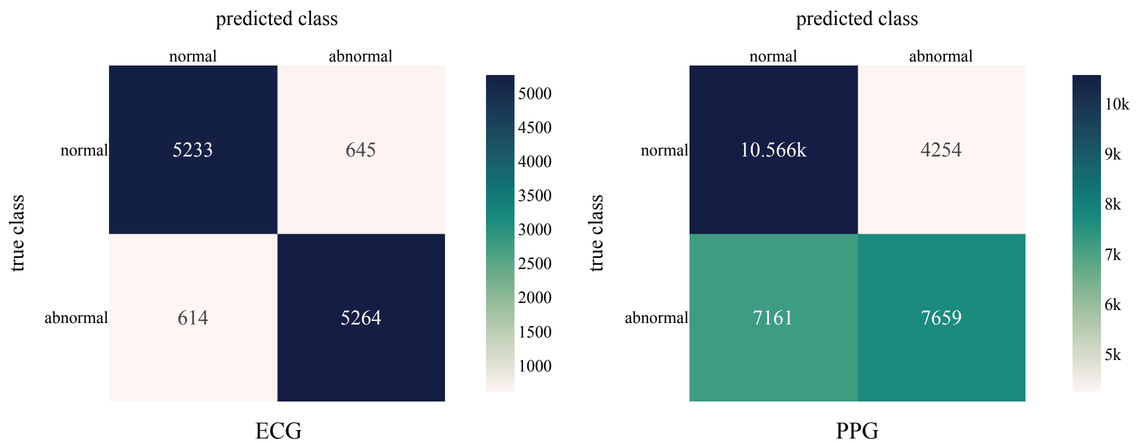

Figure 5.4: Confusion matrices of the autoencoder's performance on ECG data (left) and PPG data (right).

An F1-score of 89.1% is obtained for ECG data and 64.88% for PPG data. A sample for each classification error is shown in Figure 5.5. In the poor reconstructions of sinus rhythm, it is noticeable that the curves show greater variance than other signals with normal sinus rhythm. This can be explained by the differences between patients, as in some patients the sinus curve has a greater variance than in others. Additionally, there are differences in abnormal heartbeats depending on the arrhythmia. In class VT there is no misattribution to sinus rhythm. In class

PAC, however, 11.4% are misclassified as sinus rhythm. Class PVC and SVT are misclassified 0.01% of the time.

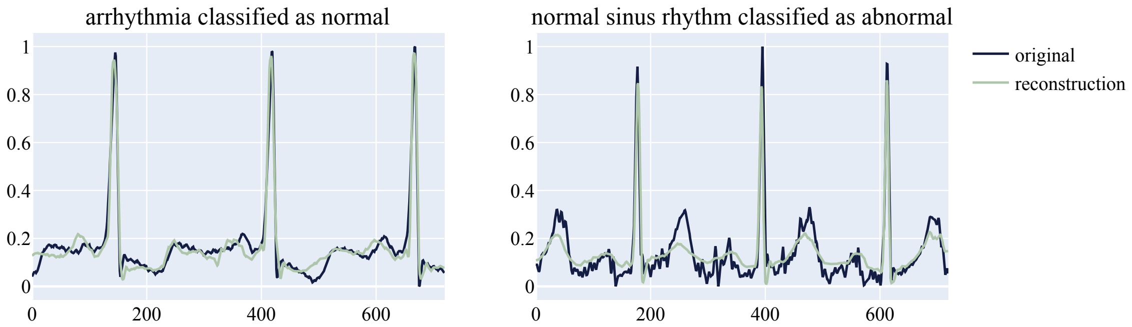

Figure 5.5: One example of each misclassification of the autoencoder, where a sinus rhythm is classified as abnormal (on the left), and an arrhythmia is classified as normal (on the right).

In this work, there is a potential for incorrect label assignment when dividing the PPG sequences into two-second sequences, which may affect the classification into normal/abnormal, as well as in the multiclass categorization. Splitting the signals into shorter segments increases the risk for mislabeling of PVC and PAC samples, as these abnormal segments are transient and the rhythm is otherwise normal. As a result, normal sequences may be labeled as abnormal.

To investigate the effect of potential mislabeling, multiple tests are performed to compare the performance of the model on ten-second and two-second sequences using the F1-score. The analyses are performed on the network architecture presented by Liu et al [4]. The results show no significant difference in the scores for the two variants (37.32% for ten seconds and 37.92% for two seconds). However, the confusion matrices (cf. Figure 5.6) show that the model is more likely to classify abnormal sequences as normal for the ten-second sequences. As the proportion of abnormal parts in ten-second sequences is small, it is difficult for the model to detect them. In contrast, normal sequences are more often misclassified as abnormal in the two-second sequences. However, this behavior is preferred because the misclassification of abnormal sequences has a greater impact, as it imposes a potential severe health risk. Overall, using two-second sequences provides a better trade-off between the risk of misclassification and data availability.

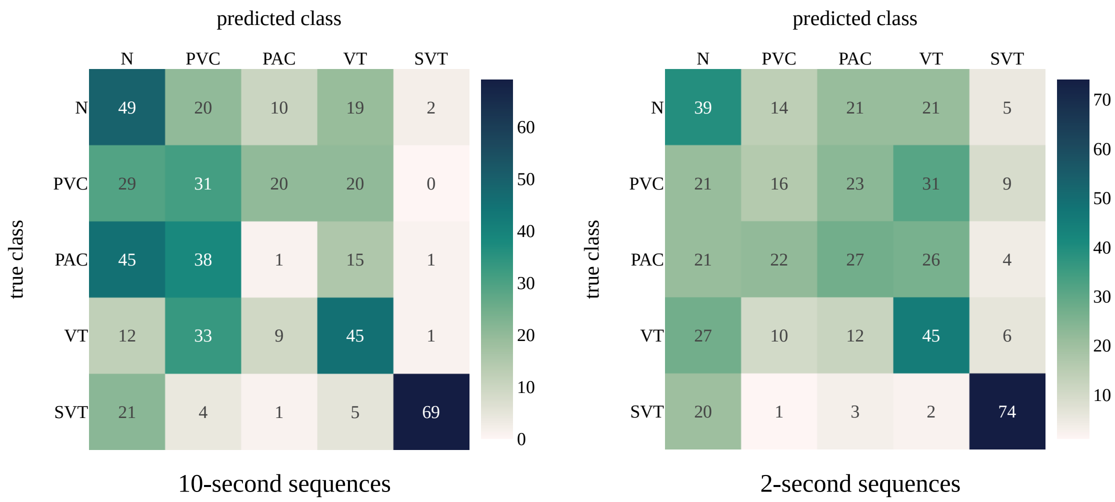

Figure 5.6: Confusion matrices showing the classification performance for ten-second (left) and two-second sequences (right).

In the second network, additional data preprocessing is carried out by reducing the number of samples to a maximum of 2000 for each class. The purpose of this step is to produce a balanced data set. However, in the case of the ECG data, the PAC and SVT classes suffer from a scarcity of samples (cf. Figure 4.1), resulting in an unbalanced data set. A reduction in the number of samples to match the class with the fewest samples results in a severe loss of performance, so this variant is excluded. Conversely, the PPG data set is balanced by limiting the number of samples to 2000 for each class. The performance of the models is tested on 100 samples per class. This selection is imposed by the limited number of samples available for certain classes, and thus the maximum use of the data for training purposes. As with the training data, the underrepresented classes in the training set also showed a lack of samples in the test set. In order to address this issue, a comparable ratio as in the training data of samples for each class is maintained in the test set. The labels of the samples are additionally one-hot encoded before the models are implemented.

To implement the first channel of the second network, the first step is to generate Betti curves from the samples. The curves can be generated using the *GUDHI* package. The package is only supported for Anaconda in combination with the M1 chip. For this reason, the generation of the Betti curves is outsourced from the project and realized via a Jupyter Notebook. In addition, the TDA toolbox from Dindin [48] is used to generate the Betti curves. First, the persistence barcode is generated and then the Betti curves are computed. The number of points to be used for the curves is specified as a parameter in the curve generation function. If the number of points is too high, the curve is too fine-grained and the topology behind the curves is not as clear. For this reason, 128 points are chosen.

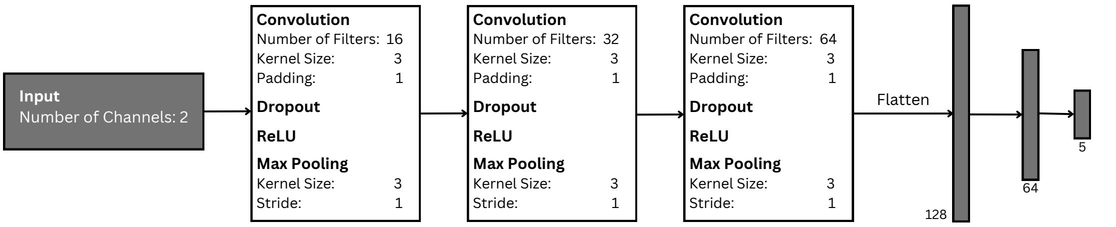

Figure 5.7: Model structure of the Betti-CNN identical for ECG and PPG data.

The generated Betti curves are used as input to a one-dimensional CNN. The CNN consists of three convolutional layers, each with a dropout of 0.5 and a maximum pooling operation. ReLU is used as the activation function. After the three convolutional layers, there is a flattening operation, followed by three linear layers. The last layer uses the Softmax function to obtain a probability for the five classes. The architecture is shown in Figure 5.7. The two Betti curves are processed in parallel as two channels in the CNN. Training is done for 150 epochs with a batch size of 128, a learning rate of 0.001, and a weight decay of 0.005. The cross-entropy loss is used as the loss function and optimization is done with Adam. An F1-score of 62% is achieved on ECG data and 34.94% on PPG data. The corresponding confusion matrices are shown in Figure 5.8.

When trained on ECG data, the total misclassification of all class PAC and SVT samples is striking. These classes are underrepresented in the data. For comparison, the model is trained without the underrepresented classes and reached an F1-score of 90%. Additionally, analysis of the confusion matrices of the models trained on PPG data shows that a great number of arrhythmias are classified as normal. This behaviour is also observed with the autoencoder.

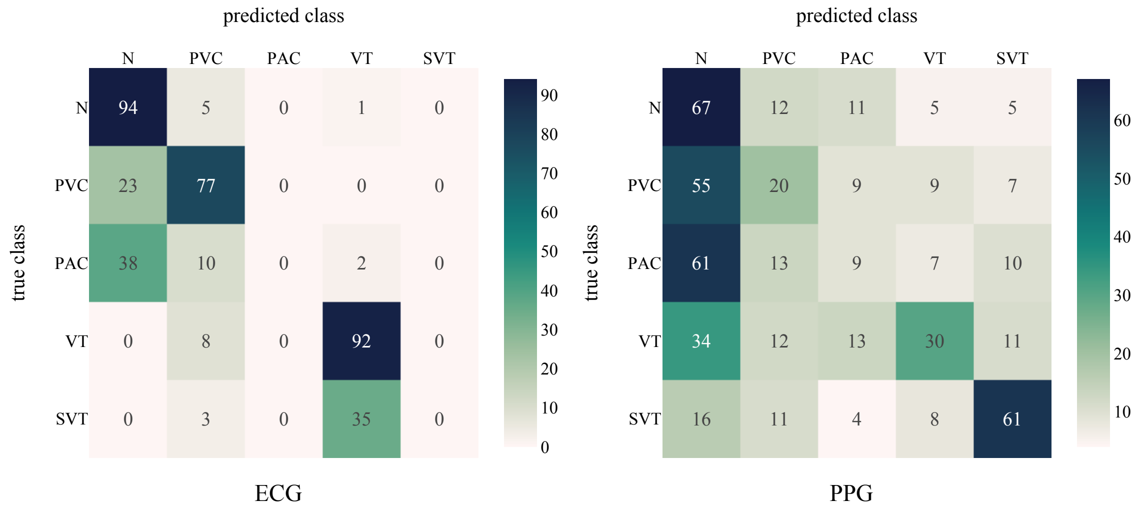

Figure 5.8: Confusion matrices of the Betti-CNN's performance on ECG data (left) and PPG data (right).

In the second channel, the signals are processed using a CNN, which is closely modeled on the Betti-CNN structure. However, unlike the Betti curve, the signal itself is used as the input to the CNN as a single channel. The CNN has the same number and structure of layers as the Betti-CNN but differs in kernel size. While the kernel size remains the same in the Betti-CNN, it increases in the CNN. A visual representation of this architecture is shown in Figure 5.9.

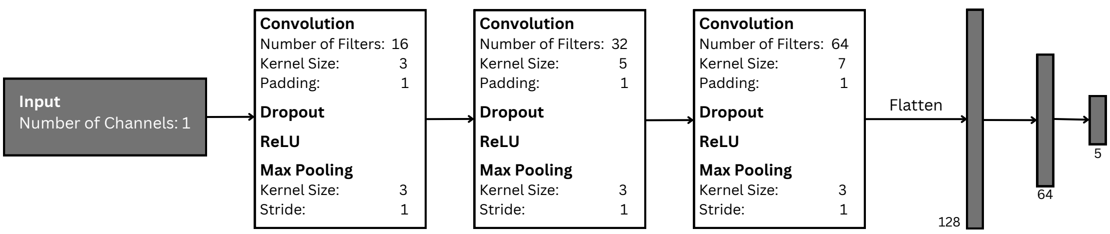

Figure 5.9: Model structure of the CNN identical for ECG and PPG data.

The performance of the CNN in classifying five classes on ECG data is evaluated with an F1-score of 56%. As also observed with the Betti-CNN, none of the classes PAC and SVT receive any sample assignments during the inference process. Again, for comparison purposes, the model is compared with the classification of the three classes. A significant performance improvement is also observed, with an F1-score of 92%. In contrast, the CNN performs poorly when applied to PPG data, achieving an F1-score of 6.67%. In this case, samples are randomly assigned to a single class during inference.

In the process of merging the two models into a multichannel model, the convolutional layers are adopted, followed by a global average pooling. The outputs of the pooling operation are then concatenated, and the resulting output is further processed in three fully connected layers. As before, a Softmax operation, which converts the output into probabilities per class, is used to generate the final output. The structure of the layers remains unchanged, except for the addition of global average pooling and concatenation operations. The resulting architecture is shown in Figure 5.10. In the case of ECG data, the multichannel model achieves a score of 61%, while for PPG data the score is 18%.

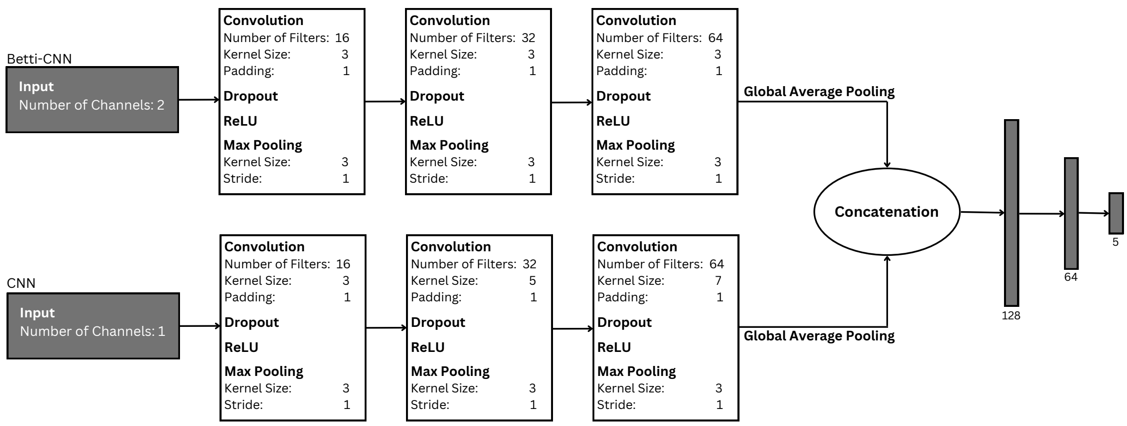

Figure 5.10: Model structure of the multichannel network identical for ECG and PPG data. The multichannel network is a combination of the Betti-CNN and the CNN.

5.3 Ensemble of Multiple Binary Classifiers

Given the uneven distribution of arrhythmia classes in the ECG data, an ensemble of several binary classifiers is evaluated to improve the performance of the second network. For each class, a model is created to distinguish it from all other classes. Consequently, a change in the model structure is required for the second network. Specifically, five models are created for each channel, maintaining the model architecture of the Betti-CNN and CNN models. The resulting structure of the CNN channel has already been explained in Section 4.4.

The data preprocessing method for the ensemble of binary classifiers differs from the multiclass approach in one aspect. Instead of one data set with different class labels, five data sets with binary differences are required. To achieve this, a data set is created for each class, where a True value corresponds to that class and a False value corresponds to a sample from all other classes. The number of samples from other classes is limited to the number of available samples from the True class. Apart from this modification, all other preprocessing steps remain identical between the two approaches.

Class	Betti-CNN	CNN	Number of Samples
sinus rhythm	91%	96%	> 5,000,000
premature ventricular contraction	93%	92%	2,316
ventricular tachycardia	92%	89%	2,731
premature atrial contraction	66%	50%	612
supraventricular tachycardia	91%	88%	867

Table 5.1: F1-scores of each model in the binary ensemble trained on ECG data for each class and the number of samples per class. The score for the class with the fewest samples performs the worst.

The models for the ensemble of binary classifiers are trained sequentially, with the training parameters remaining the same. The overall model architecture also remains unchanged, except for the number of neurons in the final output layer during the Softmax operation. This number is reduced from the previous five to two, as binary classification is now performed. During inference, each test sample is passed through each model and each model returns a True or False value. A True value indicates the presence of the class assigned to that model. A list with all True and False values of the individual models is returned. This list is then used to determine the labels. This approach may result in the assignment of multiple labels per sequence, but this scenario is not encountered during the experiments, and thus the handling of such cases is not addressed in this work.

For the models trained on ECG data, the performance of each class is evaluated for both the Betti-CNN and CNN models. The results are presented in Table 5.1. It is observed that the models in the class with the fewest samples (class PAC) have the lowest performance. Furthermore, the multichannel model achieves an F1-score of 91% on the ECG data, which is a significant improvement of 30% over the previous result. This indicates the effectiveness of the proposed approach in improving the classification performance of ECG data, especially in the presence of unbalanced classes.

Given the significant improvement in performance observed with the ensemble approach on ECG data, the same technique is applied to PPG data. The results show that the Betti-CNN model achieves an F1-score of 58.98%, an improvement of 24% over the previous result. The CNN model also benefits from the ensemble technique and achieves an F1-score of 71.15% on the PPG data. Notably, the multichannel network shows a strong improvement in performance, achieving an F1-score of 89%. This result highlights the effectiveness of the proposed ensemble approach in improving the classification performance of PPG data. The results of the ensemble model in direct comparison with the multiclass model are shown in Table 5.2.

Model	Multiclass Approach	Binary Ensemble
Betti-CNN	35%	59%
CNN	7%	71%
Multichannel Model	18%	89%

Table 5.2: Comparison of F1-score performance between a multiclass approach and an ensemble of binary classifiers for the classification of PPG signals. The ensemble model outperforms the multiclass model.

It is worth noting that utilizing an ensemble of several binary classifiers can lead to a significant improvement in classification accuracy. The relevance of these results are discussed in section 6.1.

5.4 Transfer Learning

This work investigates the use of transfer learning (TL) to improve the classification of PPG data. In particular, it explores the potential of using basic structures learned from ECG data to improve the classification of PPG data. In Section 4.3 the concept of transfer learning and several variants is described.

To implement the TL approach, a new function that freezes the weights of different layers is introduced. The rest of the implementation remains unchanged. First, the transfer learning with the autoencoder is evaluated. The weights of all encoder layers are frozen and the decoder layers are tuned with PPG data. Afterwards, the autoencoder achieves an F1-score of 66.6%, an improvement on the previous 64.88%.

This work also explores two TL approaches on each of the models in the second network, which are compared with models trained on PPG data only. The first approach freezes the weights of the convolutional layers and only fine-tunes the linear layers. In the second approach, the model is pre-trained on ECG data and all layers are fine-tuned using PPG data.

Model	PPG-only	Fine-tuning Fully Connected Layers	Fine-tuning all Layers
Betti-CNN	34.94%	16.04%	35.4%
CNN	6.67%	36.42%	38.82%
Multichannel Model	18.33%	20.26%	29.98%

Table 5.3: Comparison of F1-score performance between a multiclass model trained solely on PPG signals and the use of transfer learning, where the model is first pre-trained on ECG data and then fine-tuned with PPG signals.

In the first approach, only the final layers are fine-tuned and the Betti-CNN model achieves an F1-score of 16.04%. This result is a decline from the previous 34.94%.

However, the performance of the CNN model improves significantly from a non-learning model with a score of 6.67% to a classifier with an F1-score of 36.42%. Comparing the multichannel model trained on PPG data only with the first TL approach, an improvement from 18.33% to 20.26% is noticeable. The second approach involves fine-tuning the entire model, which improves the performance of the Betti-CNN model to 35.4%. The performance of the CNN model also improves with this approach from the previous 6.67% to an F1-score of 38.82%. The same applies to the performance of the multichannel model, which achieves an F1-score of 29.98% using the second TL approach. When comparing the two TL approaches, it should be noted that fine-tuning the entire network architecture produces better results than fine-tuning individual layers. The results are also shown in Table 5.3.

Although some of the improvements achieved through transfer learning may be small, there is a noticeable performance increase within the limits of imprecision.

6 Discussion and Conclusion

The results obtained and the design choices made are discussed in the following. Limitations are also highlighted, and possible further work is described. Finally, a summary of all the relevant results is given in the conclusion.

6.1 Discussion

Several approaches have been tested to improve the classification of arrhythmias using PPG data. First, a reference model was developed using ECG data, combining various aspects of previous publications regarding data selection, number of classes, and methods used. The reference model differentiated between normal and abnormal sequences using an autoencoder with an F1-score of 89%. Thereby, the PAC class has the highest percentage of misclassification compared to the other classes. This result may be related to the small sample size of the PAC class compared to other classes. Therefore, it is reasonable to conclude that the effectiveness of the autoencoder model on ECG data could be enhanced by increasing the sample size for underrepresented classes.

When classifying the sequences into five classes, an F1-score of 61% is obtained, which do not meet the benchmark performance. This result could be attributed to the underrepresentation of some classes, as the classification accuracy for the three most common classes is 90-92%. This underrepresentation is due to the limited choice of classes available in the PPG data. During the TL evaluation, the same classes have to be used in both data sets, limiting the otherwise large ECG database.

To overcome the challenge of an imbalanced data set, this work investigates the potential of using multiple one-vs-all binary classifiers instead of a single multiclass classification model. This type of classification improves the original approach by

30%, resulting in a score of 91%. This approach is especially useful in situations with limited data sets and reduced model structures. The superior performance of an ensemble of multiple binary classifiers can be attributed to two factors. First, discrepancies in the number of samples per class within the training data are equalized. In cases of uneven distribution, the model learns to consider the likelihood of a class occurring. As a result, classes that are underrepresented in the training data are selected less frequently during inference. This is demonstrated by the results on ECG data for the multiclass model, where eliminating underrepresented classes from the data significantly improve the classification. In the single binary models, in turn, each sample from each model is considered and the probability is identical for each class in binary relevance. Secondly, performance is improved because binary discrimination is less complicated and easier to learn. In particular, smaller model structures perform better in binary classification than in multiclass classification. Furthermore, binary classifiers have previously been shown to outperform multiclass approaches in published research (cf. Chapter 3). Our approach takes this concept of binary classification and applies it to the multiclass context via an ensemble model.

The developed model was afterward applied to PPG data. First, the performance of the model structure was assessed by training it on PPG data only. The autoencoder achieve an F1-score of 65%, while the multichannel network achieve 18%. Notably the CNN is unable to learn anything and the Betti-CNN achieve a score of 35%. Due to the limited sample size of individual classes in the ECG data small model architectures are used in the second network to avoid overfitting. However, the small model architectures are not sufficient to extract useful information from the PPG data. Consequently, the CNN cannot learn anything. In contrast, the Betti-CNN is able to classify some of the samples correctly because the features have already been extracted by preprocessing the signals into Betti curves. The curves, which represent the topology of the sequences, are easier for the CNN to interpret and the features are more obvious. The superior performance of the Betti-CNN is also evident in the classification on ECG data. Another advantage of using Betti curves is their smaller size, which allows them to be processed more quickly. This smaller size also enables processing with smaller model structures.

The study aims to evaluate the effectiveness of transfer learning, which involves fine-tuning models trained on ECG data with PPG data. The results show slight improvements of up to two percent for the autoencoder and the Betti-CNN. The most significant improvement is seen in the CNN, which transforms from a non-learning model to a classifier with an F1-score of 39%. Transfer learning is used to extract coarse structures from the ECG sequences, which are then recognized in the PPG data, resulting in a performance gain. The increase in training samples also explains the improvement with TL. The ECG data is used to learn simple structures, which can then be consolidated and developed on the PPG data. The similar structure of the two curves is advantageous in this context. Even though the improvement with TL on the autoencoder and Betti-CNN is small, an improvement could be observed. Therefore, transfer learning can be said to improve the results within the limits of imprecision. Other approaches, such as the one presented by Li et al. [11], also achieved only a few percentage points of improvement when applying TL to ECG and PPG data.

Given the significant improvement achieved by the ensemble method in classifying ECG data, this approach is also applied to PPG data. As a result, a significant increase in performance is observed, with results approaching those obtained with ECG data, resulting in an F1-score of 89%. Again, the non-learning CNN evolves into a classifier that can determine the correct class with an F1-score of 70% when using the ensemble approach. This result reinforces the assumption that smaller model structures can perform binary classification better than multiclass classification. In addition, the smaller size of binary models makes them more suitable for running on smart devices. Another advantage of the ensemble is its versatility for multilabel classification. In this study, two-second sequences without centering the peaks are used. This approach is very practical as preprocessing sequences can be time-consuming and require high computational power, making it difficult to apply on smart devices. Multiple labels can appear within a single sequence by selecting sections of seconds. Using the ensemble, multiple labels can be assigned to a sequence simultaneously. One drawback of the segmentation into second-long sequences is the difference in the number of QRS-complexes, thus total number of heartbeats in the sequence. If the number of QRS-complexes remain constant, performance might improve.

In terms of the overall model architecture, there are also some aspects that need to be considered. An important aspect of the model architecture is the separation into two networks. The autoencoder is responsible for the distinction between normal and abnormal sequences so that the latter is only processed by the subsequent network. This approach optimizes computational resources by focusing only on abnormal heart rates and rhythms. However, for this approach to be successful, it is necessary to accurately classify all abnormal heartbeats. Furthermore, in the ensemble approach the binary classifiers include a class that separates sinus rhythm from all other classes, which is consistent with the autoencoder's separation of normal and abnormal sequences. Therefore, it would be logical to first run sequences through the binary model that recognizes sinus rhythm and use the other classifiers based on the output when the performance of the binary model outperforms the autoencoder. This is for example the case for the CNN trained on ECG data reaching an F1-score of 96%.

Furthermore, the necessity of the multichannel approach is questionable. In general, the performance of the multichannel network is inferior to that of the best single model, except when an ensemble of binary classifiers is used on PPG data. A potential advantage of this approach is the reduction of overfitting that can occur within a single model. In theory, combining different models should lead to the generation of new features and thus improve performance. However, the current underperformance of the multichannel approach may be due to the poor performance of a single channel. Therefore, further research is needed to evaluate the potential benefits of the multichannel strategy using larger data sets and more extensive model structures.

The comparability of the approach presented in this work with other existing work is limited by several factors. Firstly, comparing the performance of this approach with other ECG-based approaches, such as the one by Dindin et al. [13], is complicated due to differences in the classes and data used. This work focuses on classifying both heartbeats and -rhythms, whereas Dindin et al. [13] focused solely on heartbeats. Additionally, due to the selection of different classes, the database used in this work is relatively limited, which result in a smaller number of available training data. Nevertheless, comparable results are achieved in the area of multiclass classification based on ECGs using the ensemble of binary classifiers.

When it comes to classification using PPGs, a comparison with the work of Liu et al. [4] is relevant, as a part of the database from this publication is taken. However, the authors used an imbalanced data set and evaluated the performance of the classifier in terms of accuracy, which can be misleading. A high accuracy score on an imbalanced data set does not necessarily indicate the true performance of the classifier, as it tends to predict the most frequent class. For this reason, in this work, the data set is balanced and the F1-score is used as the evaluation metric. In addition, Liu et al. [4] only published the validation and test data, which accounts for 40% of the data. To provide a baseline for an approximate comparison, the approach of Liu et al. [4] was re-implemented using the previously balanced smaller data set. The F1-score achieved is significantly lower than the accuracy stated in the publication. However, the ensemble of binary classifiers developed in this work outperforms the previous work with an F1-score of 89%, even assuming that the accuracy stated in the publication reflects the true performance.

Another relevant approach to the multiclass classification of PPG is that of Polania et al. [9]. In their paper, the authors distinguished four classes (N, PVC, VT, and premature VT). However, PVC and VT are distinguished from the other two classes, so there is no direct multiclass classification. The authors reported an accuracy of over 90%, but it is not clear whether the data used in their study are balanced. In addition, they used complex handcrafted features, making the approach complex, probably domain-knowledge driven and time consuming.

In summary, due to the limitations of the aforementioned works and different class choices, it is not possible to make a concrete comparison between the classification approach presented in this work and the current State of the Art in ECG and PPG classification.

6.2 Limitations

The results of this work are significantly limited by the restricted accessibility of ECG and PPG data. In particular, the available PPG data on labeled arrhythmias is very poor, and through the use of TL the ECG classes must contain the same categories as the PPG data. However, the categories present in the PPG data are poorly represented in the otherwise extensive ECG database. Consequently, this study requires the use of small model structures to avoid overfitting. The results need to be analyzed in light of small model structures and limited training data and must be evaluated further on larger model structures with more comprehensive data to ensure generalizability.

Furthermore, no information on the quality of the data can be provided. One of the problems that can arise during data preprocessing is the possibility of incorrect label assignments, particularly when segmenting PPG sequences into shorter segments. These incorrect label assignments can be learned by the model, potentially leading to poor performance. In addition, certain signals had to be upsampled due to the use of different databases with different frequencies. Upsampling introduces new data points that can distort the original signal and subsequently affect the model training. In addition, the PPG data was collected in a clinical context, which may lead to performance differences when the model is applied to PPG data collected from a smartwatch. Retraining is therefore required.

It should be noted that the results of this study have not been cross-validated, so there may be slight variations in the F1-score when applied to other data splits. In addition, the medical context in which the study was conducted requires further consideration. Misclassification of an arrhythmia can have serious consequences, so errors in this regard should be weighted more heavily, which has not been done in this study yet. Another critical issue in the medical context is the interpretability of the models. The decision-making process of deep learning models is not easily understood, which poses a challenge to their use. However, in the context of medical applications, the interpretability of results is of significant importance. Therefore, the use of explainable AI should be considered.

It is important to emphasize that models trained on single-lead ECGs or the similar PPGs cannot replace a physician's examination with multichannel ECGs. This is because not all arrhythmias can be captured by single-lead ECGs. Therefore, caution should be taken when interpreting the results of such models and their limitations must be acknowledged.

6.3 Further Work

The results of this work have been evaluated by experimental means. However, due to the extensive nature of the investigation, certain aspects have been omitted and may be further explored in future work.

Initially, the components of the model were implemented and evaluated separately. The model's architecture aims to combine these individual components into a cohesive concept focusing on future practical applications. Due to time constraints, this step was not feasible within the current work. In addition, the individual model components could be optimized through hyperparameter tuning to improve the F1-score, which is not the primary focus of this study.

A limitation of this study is the insufficient amount of training data available. To overcome this problem, a sliding window approach could be applied to increase the number of data samples. Specifically, the window could be shifted by a few data points during sequence generation to produce more diverse sequences that partially represent the same arrhythmias. Increasing the training data allows larger model architectures to be used and the results to be thoroughly evaluated.

The data preprocessing stage provides another opportunity to generate new samples of underrepresented classes by using upsampling techniques. Different upsampling methods should be evaluated for suitability. In addition, signal preprocessing can be optimized by dynamic time warping and other techniques, although it should be considered that each step increases the complexity of the approach. In particular, the study identified several sequences that appeared suspicious to the untrained eye. To combat this, an expert review of label assignments is recommended. Furthermore, the testing of the approach on new patients with different measurement devices will be of interest. An important issue in data collection and selection is the consideration of patient metadata such as BMI, gender, and age. Such factors can affect the model performance and should be considered in further work. Moreover, the approach can be extended to other classes by collecting newly labeled data. The collection of personalized PPG data using smartwatches is an interesting avenue for further exploration. This requires refining the model for production and identifying a solution for the edge device application. This is where Tiny-ML comes into its own.

The ensemble approach used in this study could benefit from further refinement using alternative methods. For example, the effectiveness of using the one-vs-one strategy and other techniques as the one presented by Hong et al. [45] (cf. Section 4.4) to merge individual models should be evaluated. Another possibility to consider is the adoption of a multilabel approach, allowing input sequences to have multiple labels, which is particularly relevant for later application on smart devices. In addition, knowledge distillation is a viable option for merging individual models. Moreover, the TL approach can be combined with the ensemble method. Since TL was found to slightly improve the results, the same effect could occur applying TL on the ensemble model.

Another compelling idea is the potential application of the approach to multiple ECG leads. This allows for the capture of additional information that is not discernible in a single-channel ECG. Consequently, the CNN would receive multiple channels as input, analogous to the Betti-CNN. It should be noted that this idea is not intended for use on edge devices as these cannot record multi-channel ECGs.

The autoencoder can also be optimized. A semi-supervised training approach can be implemented that includes both the normal and abnormal classes during training. This adjustment would also allow the threshold to be modified. In the inference of the autoencoder, a score can be introduced to reflect the reliability of the classification. This score can be calculated by measuring the distance from the threshold. If a sample is assigned to the normal rhythm but is close to the threshold, it can be processed in the second network. This approach reduces the number of incorrectly as normal labeled arrhythmias, which is relevant in a medical context.

6.4 Conclusion

This work examined whether an arrhythmia classification model pre-trained on ECG data could achieve a higher F1-score than a model trained on PPG data only. Therefore a reference model trained on ECG signals was developed that incorporates the advantages and aspects of previous studies. Afterward, the model was applied to PPG data which included an evaluation of transfer learning. During the study, the challenge of the underrepresentation of certain classes in the ECG data was encountered. To address this issue, the performance of an ensemble of binary classifiers was compared to that of the original multiclass model.

The model structure developed in this work is characterized by the division into a binary classification of normal and abnormal heart rhythms, and the multiclass classification of all abnormal heart rhythms. This design helps to save computational resources, as all rhythms classified as normal do not require further processing. Deep learning was chosen in the classification of abnormal sequences as it can better account for variations in individual heartbeat morphologies. Other methods, such as machine learning, require pre-designed features that may not be able to capture such differences. In addition, the use of TDA, specifically Betti curves, has been incorporated into the overall framework to compensate for individual patient differences. The developed model is designed to be transferable to new patients. To achieve this, the training and test data sets were partitioned based on patient identity. Patients included in the test data set were not previously included in the training data set. This approach ensures that the model can classify the heart rhythms of new patients.

The models were first implemented and evaluated individually. The autoencoder's initial classification of normal and abnormal heart rhythms achieved an F1-score of 89% on ECG data. The Betti-CNN model achieved a score of 62%, while the CNN achieved a score of 56%. The multichannel approach, which combined the last two models, resulted in an F1-score of 61%. Notably, the two underrepresented classes, PAC and SVT, were not selected during inference of the second network, which contributed to the poor performance. An ensemble of several one-vs-all binary classifiers was evaluated to address this underrepresentation. The ensemble significantly improved the performance of the multichannel model, resulting in a score of 91%.

The performance of the autoencoder on PPG data was weaker than on ECG data, resulting in a score of 65%. The Betti-CNN model achieved an F1-score of 35% on the PPG data. However, similar to the autoencoder model, many abnormal samples were incorrectly classified as normal. Moreover, the CNN model failed to learn from the PPG data, with all samples randomly assigned to a class during inference, resulting in a low F1-score of 7%. This failure to learn may be due to the small architecture of the model. Due to the limited size of the ECG database for the classes used, smaller model structures were chosen to avoid overfitting. However, previous deep learning approaches to arrhythmia classification on PPGs have utilized large model architectures. The smaller structures used in this study may not be able to effectively capture the features in the PPG data. In contrast, the Betti-CNN may have learned from the PPG data because the generation of the Betti curves extracted features that the CNN could process more efficiently. When the two models are combined into a multichannel model, a score of 18% was achieved.

One way to improve classification with PPGs is through the use of TL. Several approaches were tested to evaluate TL on the models. The most successful approach is to pre-train the models on ECG data and then fine-tune the whole model on PPG data. An improvement of 0.46% for the Betti-CNN was obtained with this approach. The greatest improvement was observed when transfer learning was applied to the previously non-learning CNN, enhancing the score to 38.82%. By fine-tuning the decoder layers, the autoencoder performance was improved by 1.72%. One of the advantages of transfer learning is the increased amount of training data, which allows simple structures to be learned from the ECG sequences that are similar to the PPG data, allowing the model to learn fine-grained features in subsequent iterations. The multichannel model also showed improvement, achieving a score of 30%. Although the increase in performance observed is small in some cases, there is an improvement within the limits of imprecision.

Due to the strong improvement in results when using an ensemble on ECG data, this approach was also tested on PPG data. It was observed that the binary classification provided a simplification from which the approach could benefit with PPG data. The results showed that the ensemble approach outperformed all other PPG models with an F1-score of 59% for the Betti-CNN, 71% for the CNN, and 89% for the multichannel model. The superior performance of the ensemble method can be attributed to the simplified binary classification, which is easier to learn than multiclass classification. Additionally, the ability to assign multiple classes to a sequence using a multilabel approach makes the method more practical and easier to implement on smart devices.

However, the results obtained should be viewed in the context of the underrepresentation of certain classes in the ECG data and the relatively small sample size within the otherwise extensive ECG database. It is possible that different results could be obtained using larger model structures and more extensive data. In addition, to confirm the score of the results obtained, it is recommended that this approach be cross-validated and tested on other data.

When it comes to place the proposed model in existing work, several factors make this difficult. Nevertheless, this work introduces an ensemble approach which performance is comparable to benchmark performance on ECG data and even surpasses it on PPG data. Additionally, this work represents an advance in the simultaneous classification of heartbeats and -rhythms on ECG data and proposes a technique for training models with good performance on limited data. In conclusion, although further development is needed, the potential benefits of implementing this approach for widespread cardiovascular health monitoring make it a promising area for future research.

A Abbrevations

a	aberrated atrial premature beat
A	atrial premature beat
AF	atrial fibrillation
AHA	American Heart Association
AVNRT	atrioventricular nodal re-entrant tachycardia
Betti-CNN	convolutional neural network that processes Betti curves
CNN	convolutional neural network
DCNN	deep convolutional neural network
DNN	deep neural network
DTW	dynamic time warping
e	atrial escape beat
E	ventricular escape beat
ECG	electrocardiogram
f	fusion of paced and normal beat
F	fusion of ventricular and normal beat
j	nodal (junctional) escape beat
J	nodal (junctional) premature beat
L	left bundle branch block beat
LED	light-emitting diodes
MIT-BIH	Massachusetts Institute of Technology - Beth Israel Hospital
N	normal beat
P	paced beat
PAC	premature atrial contraction
PPG	photoplethysmography
PVC	premature ventricular contraction

Q	unknown beat
R	right bundle branch block beat
ReLU	rectified linear unit
ROC	average area under the receiver operating characteristic curve
S	supraventricular premature beat
SDAE	stack denoising autoencoder
SMOTE	synthetic minority over-sampling technique
SVEB	supraventricular ectopic beat
SVM	support vector machine
SVT	supraventricular tachycardia
TDA	topological data analysis
TL	transfer learning
U	unclassified beat
VEB	ventricular ectopic beat
VT	ventricular tachycardia
WHO	World Health Organization

B List of Figures

C List of Tables

D Bibliography

[1] WHO CVD Risk Chart Working Group, "World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions.," *Lancet Global Health*, vol. 7, 2019.

[2] K. A. Salam and G. Srilakshmi, "An algorithm for ecg analysis of arrhythmia detection," in *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1–6, 2015.

[3] A.-U. Rahman, R. N. Asif, K. Sultan, S. A. Alsaif, S. Abbas, M. A. Khan, and A. Mosavi, "Ecg classification for detecting ecg arrhythmia empowered with deep learning approaches," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[4] Z. Liu, B. Zhou, X. Chen, Y. Li, M. Tang, and F. Miao, "Multiclass arrhythmia detection and classification from photoplethysmography signals using a deep convolutional neural network," *Journal of the American Heart Association*, vol. 11, 2022.

[5] A. Hannun, P. Rajpurkar, M. Haghpanahi, G. Tison, C. Bourn, M. Turakhia, and A. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, 2019.

[6] V. Nagarajan, S.-L. Lee, J.-L. Robertus, C. Nienaber, N. Trayanova, and S. Ernst, "Artificial intelligence in the diagnosis and management of arrhythmias," *European Heart Journal*, vol. 42, 2021.

[7] Neha, H. K. Sardana, R. Kanwade, and S. Tewary, "Arrhythmia detection and classification using ecg and ppg techniques: a review," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 4, pp. 1027–1048, 2021.

[8] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144–164, 2016.

[9] L. Polania, L. Mestha, D. Huang, and J.-P. Couderc, "Method for classifying cardiac arrhythmias using photoplethysmography," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2015, pp. 6574–6577, 2015.

[10] M. Radha, P. Fonseca, A. Moreau, M. Ross, A. Cerny, P. Anderer, X. Long, and R. M. Aarts, "A deep transfer learning approach for wearable sleep stage classification with photoplethysmography," *npj Digital Medicine*, vol. 4, 2021.

[11] Q. Li, Q. Li, A. Cakmak, G. Da Poian, D. Bliwise, V. Vaccarino, A. Shah, and G. Clifford, "Transfer learning from ecg to ppg for improved sleep staging from wrist-worn wearables," *Physiological Measurement*, vol. 42, 2021.

[12] J. Ramesh, Z. Solatidehkordi, R. Aburukba, and A. Sagahyroon, "Atrial fibrillation classification with smart wearables using short-term heart rate variability and deep convolutional neural networks," *Sensors*, vol. 21, no. 21, 2021.

[13] M. Dindin, Y. Umeda, and F. Chazal, "Topological data analysis for arrhythmia detection through modular neural networks," in *Advances in Artificial Intelligence* (C. Goutte and X. Zhu, eds.), pp. 177–188, Springer International Publishing, 2020.

[14] inovex GmbH, "Über uns." `https://www.inovex.de/de/ueber-uns/`. (accessed: 25.10.2022).

[15] inovex GmbH, "Forschung & entwicklung." `https://www.inovex.de/de/ueber-uns/forschung-entwicklung/`. (accessed: 25.10.2022).

[16] inovex GmbH, "inovex lab." `https://www.inovex.de/de/ueber-uns/inovex-lab/`. (accessed: 25.10.2022).

[17] MedLexi.de, "Atrioventrikularknoten." `https://medlexi.de/Atrioventrikularknoten`. (accessed: 04.11.2022).

[18] P. A. Iaizzo, *Handbook of Cardiac Anatomy, Physiology, and Devices*. Springer Cham, 2015.

[19] A. Rawshani, "Pocket guide to ecg interpretation." `https://ecgwaves.com/pocket-guide-to-ecg-interpretation-pdf`. (accessed: 11.10.2022).

[20] A. Rawshani, "Ruhe-ekg." `http://kardiologie-sporn.at/ruhe-ekg`. (accessed: 19.10.2022).

[21] Z. Taghiyev, A. Nia, N. Gassanov, S. Baldus, and F. Er, "[56-year-old patient with angina pectoris and progressive shortness of breath].," *Deutsche medizinische Wochenschrift (1946)*, vol. 138, pp. 1513–4, 2013.

[22] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors—past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, 2014.

[23] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, pp. R1–39, 2007.

[24] V. Kalidas and L. S. Tamil, "Cardiac arrhythmia classification using multimodal signal analysis," *Physiological Measurement*, vol. 37, no. 8, pp. 1253–1272, 2016.

[25] R. Klinge, *Das Elektrokardiogramm: Leitfaden für Ausbildung und Praxis.* Thieme, 2011.

[26] L. Bibas, M. Levi, and V. Essebag, "Diagnosis and management of supraventricular tachycardias," *Canadian Medical Association journal*, vol. 188, pp. 17–18, 2016.

[27] A. Rawshani, "Ekg & echo training." `https://ekgecho.de`. (accessed: 12.10.2022).

[28] Edelsbrunner, Letscher, and Zomorodian, "Topological persistence and simplification," *Discrete Comput. Geom.*, vol. 28, no. 4, p. 511–533, 2002.

[29] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete and Computational Geometry*, vol. 33, pp. 249–274, 2005.

[30] F. Chazal and B. Michel, "An introduction to topological data analysis: Fundamental and practical aspects for data scientists," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

[31] E. Munch, "A user's guide to topological data analysis," *Journal of Learning*

Analytics, vol. 4, pp. 47–61, 2017.

[32] Y. Umeda, "Time series classification via topological data analysis," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 32, no. 3, 2017.

[33] M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[34] S. Irfan, N. Anjum, T. Althobaiti, A. A. Alotaibi, A. B. Siddiqui, and N. Ramzan, "Heartbeat classification and arrhythmia detection using a multi-model deep-learning technique," *Sensors*, vol. 22, no. 15, 2022.

[35] X. Zhang, J. Li, Z. Cai, L. Zhang, Z. Chen, and C. Liu, "Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection," *Medical & Biological Engineering & Computing*, vol. 59, 2021.

[36] M. Wu, W. Yang, and S. Wong, "A study on arrhythmia via ecg signal classification using the convolutional neural network," *Frontiers in computational neuroscience*, vol. 14, 2021.

[37] R. Hoekema, G. Uijen, and A. van Oosterom, "Geometrical aspects of the interindividual variability of multilead ecg recordings," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 551–559, 2001.

[38] S. Somani, A. J. Russak, F. Richter, S. Zhao, A. Vaid, F. Chaudhry, J. K. De Freitas, N. Naik, R. Miotto, G. N. Nadkarni, J. Narula, E. Argulian, and B. S. Glicksberg, "Deep learning and the electrocardiogram: review of the current state-of-the-art," *EP Europace*, vol. 23, no. 8, pp. 1179–1191, 2021.

[39] S. Mousavi and F. Afghah, "Inter- and intra- patient ecg heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1308–1312, 2019.

[40] Neha, H. Sardana, N. Dogra, and R. Kanawade, "Dynamic time warping based arrhythmia detection using photoplethysmography signals," *Signal, Image and Video Processing*, vol. 16, pp. 1–9, 2022.

[41] M. Merone, P. Soda, M. Sansone, and C. Sansone, "Ecg databases for biomet-

ric systems: A systematic review," *Expert Systems with Applications*, vol. 67, pp. 189–202, 2017.

[42] "Cs231n: Deep learning for computer vision." https://cs231n.github.io/transfer-learning/, 2022. (accessed: 13.02.2023).

[43] A. Lorena, A. Carvalho, and J. Gama, "A review on the combination of binary classifiers in multiclass problems," *Artificial Intelligence Review*, vol. 30, pp. 19–37, 2008.

[44] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognition*, vol. 44, no. 8, pp. 1761–1776, 2011.

[45] J.-H. Hong, J.-K. Min, U.-K. Cho, and S.-B. Cho, "Fingerprint classification using one-vs-all support vector machines dynamically ordered with naive bayes classifiers," *Pattern Recognit.*, vol. 41, pp. 662–671, 2008.

[46] "Kedro." https://kedro.org/. (accessed: 24.02.2023).

[47] "Pytorch." https://pytorch.org/. (accessed: 24.02.2023).

[48] M. Dindin, "TDA Toolbox," 12 2019.