University of Sannio

Department of Engineering

Ph.D. in Information Technologies for Engineering

Artificial Intelligence for Continuous Health Monitoring and Predictive Diagnostics

Ph.D. Coordinator:
Professor Massimiliano Di Penta

Ph.D. Student:
Ilaria Mancino

Supervisor:
Lerina Aversano

Benevento, April 2026

Acknowledgements

This journey began somewhat by chance, like many things in my life, and for that I thank myself, along with the curiosity and initiative that define me.

This PhD accompanied me through a phase of discovery, new awareness, and significant personal and professional growth.

I am grateful to those who supported and stood by me in both my successes and my failures, always reminding me to recognize the value I deserve.

But more than anything, I want to raise a toast to myself and to what lies ahead.

Abstract

The healthcare sector is undergoing a profound transformation driven by the integration of Internet of Things (IoT) technologies and Artificial Intelligence (AI), leading to the emergence of the Smart Health paradigm. This new approach shifts healthcare from a traditional reactive and hospital centered model to a proactive, patient centered system that leverages continuous monitoring through wearable sensors and intelligent data analysis. Miniaturized devices collect real time physiological and behavioral data, which are transmitted securely via IoT infrastructures to enable continuous assessment outside clinical settings. AI plays a crucial role in transforming this massive, complex sensor data into actionable knowledge for accurate diagnosis, adaptive decision-making, and personalized care.

This thesis explores the synergy of IoT and AI, known as Artificial Intelligence of Things (AIoT), in Smart Health systems, which rely on edge, fog, and cloud computing architectures to enable local processing, data aggregation, and large-scale analysis. Despite these advances, challenges such as data privacy, security vulnerabilities, bias in AI models, and the opaque nature of DL hinder clinical adoption. Moreover, the literature highlights significant gaps: there is limited application of DL techniques to IoT health data in real-world conditions, few end-to-end AIoT healthcare applications operate in clinical workflows, and the diagnostic potential of multimodal data fusion remains underexplored.

To address these shortcomings, the thesis undertakes a systematic review of current research, an exploratory analysis of authentic IoT health datasets, and the design of an AI model tailored for Smart Health applications. The central objectives are to critically evaluate the state of the art, identify technological challenges and opportunities, and assess the effectiveness of DL methods in extracting robust, clinically relevant insights from multimodal physiological signals collected by wearable IoT devices.

This work aims to contribute to the advancement of predictive, preventive, and personalized medicine by highlighting how artificial intelligence can effectively support clinicians in the decision making process.

Keywords: Internet of Things (IoT), Artificial Intelligence, Hybrid Model, Smart Health

Contents

Acknowledgements **I**

Abstract **II**

Introduction **VI**

1 Preliminary on IoT and AI in Smart Health **1**

1.1 The Internet of Things 1

1.1.1 IoT Architecture and Core Components 2

1.1.2 Edge, Fog, and Cloud Computing Paradigms 3

1.2 IoT in Healthcare 4

1.2.1 Health IoT Devices and Sensor Ecosystem 5

1.2.2 Applications and Benefits of Health IoT Systems 7

1.2.3 Challenges and Emerging Trends in Health IoT 11

1.3 Artificial Intelligence 13

1.3.1 Machine Learning in Smart Health 14

1.3.2 Deep Learning in Smart Health 15

1.4 The Convergence of AI and IoT in Smart Health 15

1.4.1 Advantages of AIoT in Healthcare Systems 16

1.4.2 Applications of AIoT in Smart Health 17

1.4.3 Challenges and Future Directions 18

2 Study Definition and Planning **21**

2.1 Preliminary Study and Application of AI Techniques 27

2.2 Analysis of IoT Health Data Using Machine Learning Techniques 31
2.3 Model Design and Implementation 35
2.4 Workflow and Background 39
2.4.1 Data Collection and Understanding 40
2.4.2 Preprocessing and Segmentation 40
2.4.3 Model Design and Training 42
2.4.4 Hyperparameter Optimisation 53
2.4.5 Model Evaluation 54
2.4.6 Explainability 57

3 Approach and Methodologies 59
3.1 Preliminary Study and Application of AI Techniques 60
3.2 Analysis of IoT Health Data Using Machine Learning Techniques 62
3.2.1 Dataset 64
3.2.2 Preprocessing 67
3.2.3 Machine Learning Pipeline 70
3.3 Model Design and Implementation 72
3.3.1 Preprocessing 73
3.3.2 Data Splitting 74
3.3.3 Hyperparameter Configuration 74
3.3.4 DL Pipeline 75

4 Study Result 79
4.1 Preliminary Study and Application of AI Techniques 79
4.2 Analysis of IoT Health Data Using Machine Learning Techniques 90
4.3 Deep Learning Model Design and Experimental Evaluation for AF Detection 95

Conclusions 103

Bibliography 107

List of Publications 123

Introduction

The healthcare sector has undergone a rapid digital transformation, catalyzed by the convergence of information technologies and Artificial Intelligence (AI). The traditional healthcare model, predominantly reactive and hospital-centered, is progressively evolving toward a proactive, data-driven, and patient-centered approach. This transition, often referred to as **Smart Health**, represents the integration of Internet of Things (IoT) infrastructures, cloud and edge computing, and advanced data analytics for continuous health monitoring and intelligent clinical decision making [9].

The Smart Health paradigm refers to an ecosystem in which patients, healthcare providers, and intelligent devices interact seamlessly. Wearable technology has driven this transformation by enabling the continuous acquisition and transmission of physiological data in real time, thus promoting a more proactive approach to healthcare. Miniaturized sensors embedded in wearable devices can detect and collect high-frequency physiological and behavioral data including heart rate [85], blood pressure [26], sleep patterns [100], physical activity [113], and blood oxygen saturation (SpO2) [28].

These data streams, once confined within hospital environments, are now transmitted in real time across secure communication networks, enabling continuous assessment of patient status in everyday settings [123].

Internet of Things (IoT) technology provides the physical and digital infrastructure that interconnects devices and platforms, enabling bidirectional communication between patients and healthcare systems. The combination of miniaturized sensors, low-power wireless communication, and edge intelligence allows the development of wearable and portable devices capable of performing continuous and non-invasive monitoring. This technological progress supports early diagnosis, remote rehabilitation, chronic disease management, and emergency response, fostering the transition from episodic to continuous care.

However, managing the acquisition and conversion of massive amounts of raw physiological data into actionable knowledge remains a major challenge. This difficulty arises not only from the risks of cyberattacks and privacy breaches [15], but also from technical constraints such as data overload, bandwidth limitations, and latency in processing real-time sensor streams [36]. In addition, the vast quantity and inherent complexity of data generated by wearable and connected devices demand advanced analytical approaches capable of extracting clinically relevant insights.

In this context, **Artificial Intelligence (AI)** plays a key enabling role. Through intelligent algorithms, raw sensor data can be transformed into meaningful representations that support accurate diagnosis, continuous monitoring, and adaptive decision making [121]. By learning from historical and real-time data, AI systems can progressively improve their predictive capabilities, leading to more reliable and personalized healthcare solutions.

Among AI methodologies, **Machine Learning (ML)** techniques have been widely adopted to analyze biomedical data and detect patterns in physiological signals. ML models learn from historical datasets to perform classification, prediction, or anomaly detection tasks without being explicitly programmed for each scenario.

More recently, advances in **Deep Learning (DL)**, a specialized branch of ML based on multi-layer neural networks, have enabled the automatic extraction of complex temporal and morphological patterns directly from raw physiological signals. These architectures have demonstrated remarkable performance improvements in several healthcare applications, particularly in medical signal processing and diagnostic support.

The integration of AI and IoT, often referred to as the **Artificial Intelligence of Things (AIoT)**, amplifies the strengths of both domains. IoT provides a continuous flow of multimodal health data, while AI transforms this raw information into actionable insights. This synergy enables a new generation of context-aware healthcare systems capable of real-time reasoning and decision making. From an architectural perspective, AIoT-based Smart Health systems leverage the integration of *edge, fog, and cloud computing* [19, 94]. Edge intelligence allows local data processing directly on wearable or portable devices, minimizing latency and preserving patient privacy. Fog nodes serve as intermediate layers that aggregate and filter data before transferring them to the cloud for large-scale analysis and model retraining.

Despite these advancements, ensuring **privacy and data security** remains one of the most significant challenges [13]. The collection and analysis of large volumes of sensitive medical information inevitably raise concerns about the protection of personal data, which could be exposed or misused in the event of a cyberattack. Closely related to this is the issue of data bias. Since AI systems learn from the data on which they are trained, any imbalance, incompleteness, or representational bias within clinical datasets can translate into biased predictions or diagnostic inequalities. Such biases risk amplifying disparities in healthcare delivery and compromising the fairness of clinical decision making.

A further challenge concerns the transparency and explainability of AI-driven decisions. DL models, in particular, are often described as “black boxes,” since their internal reasoning processes are difficult to interpret. This opacity can reduce clinicians’ trust in AI tools and hinder their integration into clinical workflows.

Overall, the convergence of IoT and AI marks a fundamental shift in healthcare delivery from reactive treatment to predictive, preventive, and personalized medicine. This integration not only enhances the efficiency of healthcare systems but also empowers patients to play an active role in managing their own health. The resulting ecosystem of connected devices, intelligent analytics, and distributed computing constitutes the foundation upon which the subsequent chapters of this thesis are built.

Despite the rapid progress and the growing interest in the integration of IoT and AI, the scientific literature still reveals several limitations that hinder the full adoption of these technologies in healthcare.

A first limitation concerns the limited number of advanced ML approaches effectively applied to IoT-generated physiological data. Although ML techniques are widely explored, many existing works rely on relatively simple models or statistical analyses due to the complexity of real-world sensor signals.

More advanced approaches based on DL remain comparatively less explored, particularly in scenarios involving long-term physiological monitoring. This is often due to the lack of sufficiently large and well-annotated datasets required to train deep architectures, as well as the high variability and noise present in real-world IoT data streams.

A second challenge relates to the limited number of operational, end-to-end applications that integrate IoT and AI within real healthcare workflows. Although both technologies are mature and widely adopted in other domains, their deployment in clinical environments remains relatively rare. AI models are often developed in controlled environments using clean, curated datasets, making them difficult to generalize to real-world scenarios where IoT data are noisy, incomplete, and affected by environmental and patient-specific variability.

Furthermore, the current literature provides few contributions addressing multimodal approaches capable of jointly analyzing multiple IoT data modalities, including physiological, behavioral, and environmental information. Most works focus on a single signal type (e.g., ECG, accelerometry, or photoplethysmography), overlooking the significant diagnostic potential that arises from multimodal data fusion.

In this context, the present thesis aims to address these gaps through a structured research pipeline comprising a systematic review of existing literature, an exploratory analysis of real IoT health data, and the development of an AI-based model tailored to Smart Health applications. The overarching objective is twofold: first, to critically assess the current technological landscape, identifying limitations and opportunities in the integration of IoT and AI; and second, to experimentally investigate the capability of ML and DL techniques to process physiological signals acquired through wearable IoT systems, ultimately evaluating their potential to deliver more accurate, robust, and clinically meaningful diagnostic and predictive tools.

CHAPTER

1

Preliminary on IoT and Artificial Intelligence in Smart Health

Foundations and Context

1.1 The Internet of Things

The **Internet of Things (IoT)** has emerged as one of the most transformative technological paradigms of the 21st century, enabling pervasive interconnection between physical objects, digital infrastructures, and intelligent services. The concept was first introduced by Kevin Ashton in the late 1990s [7], envisioning everyday objects equipped with sensors, processors, and communication interfaces capable of autonomously exchanging data. What began as a conceptual framework has since evolved into a technological reality that now underpins modern industries, public infrastructures, and consumer applications.

The early 2000s marked the first wave of IoT adoption, supported by the diffusion of wireless technologies such as RFID, Bluetooth, and WiFi, which enabled automatic identification and tracking of objects. Subsequent advances in low cost sensors, embedded microcontrollers, and cloud computing accelerated the expansion of IoT systems, enabling billions of devices to interconnect across global networks [9] [41]. Over time, IoT has expanded far beyond industrial and logistics applications, converging with artificial intelligence, big data analytics, and edge computing to form the technological foundation of the **Fourth Industrial Revolution**

(Industry 4.0) [98].

Today, IoT technologies are ubiquitous across smart cities, environmental monitoring, transportation, agriculture, connected homes, and many other domains. These systems rely on distributed networks of sensors and actuators that capture real world information such as temperature, motion, humidity, or pressure and transmit it to computational platforms for analysis and control. This integration creates a cyber physical ecosystem capable of intelligent automation, resource optimization, and adaptive decision making.

1.1.1 IoT Architecture and Core Components

A modern IoT system is commonly described through a layered architecture [45, 135] designed to ensure scalability, interoperability, and robustness. The canonical model comprises three main layers:

- **Perception Layer:** Interfaces with the physical world through sensors, actuators, and embedded devices, converting physical phenomena into digital signals.
- **Network Layer:** Handles data transmission using communication standards such as BLE, ZigBee, WiFi, LoRaWAN, NB-IoT, or 5G. This layer may also include gateways that perform routing, protocol conversion, and local data preprocessing.
- **Application Layer:** Provides end user services such as monitoring dashboards, automation, and analytics, often deployed through cloud based or distributed computing platforms.

In addition to these layers, many contemporary architectures incorporate a **middleware layer** that manages interoperability, resource abstraction, access control, and integration with external platforms. Middleware components typically rely on standardized APIs, messaging protocols, and data models to ensure seamless communication across heterogeneous devices and services.

Thanks to this modular structure, IoT systems can be adapted to a wide range of environments from low power personal devices to large scale industrial networks while maintaining flexibility, scalability, and resilience. This layered perspective provides the technological foundation upon which specialized IoT domains, including Smart Health, are built.

1.1.2 Edge, Fog, and Cloud Computing Paradigms

The massive amount of data generated by modern IoT ecosystems requires efficient and distributed computing frameworks capable of supporting low latency, high scalability, and robust privacy. Traditional cloud centric architectures often struggle with bandwidth constraints, communication delays, and the overhead associated with transmitting large volumes of data to remote data centers. To overcome these limitations, contemporary IoT infrastructures increasingly adopt a hierarchical computing model that integrates **edge**, **fog**, and **cloud** paradigms [5, 14, 101].

- **Edge Computing:** Edge computing refers to a distributed computing paradigm in which data processing and analysis are performed directly at or near the data source, such as sensors, wearable devices, or local gateways. By bringing computational capabilities closer to the physical environment where data are generated, edge computing reduces communication latency and bandwidth consumption while enabling real-time or near real-time analytics. This paradigm is particularly advantageous in applications requiring rapid response times, continuous monitoring, or localized decision making, such as industrial automation, autonomous systems, and healthcare monitoring platforms [46, 101].

- **Fog Computing:** Fog computing extends the edge computing paradigm by introducing an intermediate layer of distributed computing resources positioned between edge devices and centralized cloud infrastructures. Fog nodes provide additional capabilities for data aggregation, temporary storage, and context-aware processing, allowing part of the computational workload to be handled closer to the network edge while still benefiting from coordinated system-level management. This hierarchical architecture improves scalability, reduces unnecessary data transmission to remote data centers, and supports latency-sensitive applications that require both local responsiveness and broader system coordination [14].

- **Cloud Computing:** Offers centralized, large scale resources for long term data storage, batch processing, and advanced analytics [5, 72]. Cloud platforms provide high computational power and global accessibility, supporting complex workloads such as DL model training, large scale simulations, and big data analysis.

The integration of these three paradigms forms a coherent distributed intelligence architecture, in which computational tasks are dynamically allocated based on factors such as latency requirements [46], available resources, and data sensitivity. For example, edge nodes may execute preliminary filtering or anomaly detection, fog layers may handle context aware processing or data fusion, and the cloud may perform intensive analytics or coordination across geographically distributed systems.

Emerging technologies such as **federated learning** [71] extend this model by enabling collaborative AI training across multiple edge or fog nodes without requiring the transfer of raw data. This approach supports data locality and reduces cloud dependency while fostering privacy preserving and scalable intelligence across distributed IoT environments.

1.2 IoT in Healthcare

The rapid expansion of the Internet of Things has naturally extended into the medical and clinical domains, giving rise to what is commonly referred to as **Smart Health**.

While traditional *eHealth* and *mHealth* solutions mainly focused on the digitization of health records and the use of mobile applications to deliver health services, Smart Health goes one step further by integrating pervasive sensing, continuous connectivity, and intelligent analytics into the entire healthcare continuum [9]. In this paradigm, the patient is no longer observed only within the hospital or clinic but becomes the center of an extended cyber physical ecosystem, where physiological, behavioral, and environmental data are collected and analyzed in real time across different contexts (home, workplace, community).

By leveraging the same principles of pervasive sensing, ubiquitous connectivity, and distributed computation that characterize general purpose IoT systems, healthcare oriented

IoT infrastructures enable continuous monitoring, seamless device integration, and efficient management of biomedical data flows. This specialized ecosystem, commonly known as the *Health Internet of Things (H-IoT)*, interconnects wearable sensors, implantable devices, hospital instrumentation, and mobile health platforms within a unified digital environment. Unlike conventional monitoring systems, which are often siloed and limited to specific settings, H-IoT aims to provide an integrated view of the patient's condition over time, across locations, and along the entire care pathway.

From an architectural perspective, H-IoT systems typically combine in home devices and gateways, hospital networks, cloud platforms, and edge nodes. Data flows may originate from the patient's body, their immediate surroundings, or clinical equipment, and are progressively aggregated, filtered, and analyzed as they traverse different layers of the infrastructure. This multi layered organization allows Smart Health platforms to support use cases ranging from simple home telemonitoring to advanced clinical decision support systems [131], enabling a progressive digital transformation of healthcare delivery.

1.2.1 Health IoT Devices and Sensor Ecosystem

The deployment of IoT in healthcare relies on a diverse ecosystem of devices designed to measure physiological, behavioral, and environmental parameters with varying levels of invasiveness and autonomy. These devices differ not only in form factor and sensing capabilities but also in terms of computational resources, energy constraints, communication protocols, and integration requirements with clinical workflows [79]. They can be broadly grouped into the following categories:

1. **Wearable Sensors**: smartwatches, chest straps, wristbands, patches, and textile-based sensors capable of continuously capturing vital signs such as ECG, heart rate [85], respiration, temperature [23], blood oxygen saturation (SpO_2), blood pressure [26], or physical activity [79]. These devices are designed to be worn over long periods with minimal discomfort, making them particularly suitable for long-term monitoring of chronic patients, rehabilitation programs, or lifestyle assessment.

2. **Implantable Medical Devices**: miniaturized systems placed inside the body to measure internal physiological variables (e.g. intracardiac pressure, intracranial pressure, glucose levels, neural activity) or deliver therapeutic interventions such as pacemakers, defibrillators, neurostimulators, or insulin pumps. These devices must meet strict requirements in terms of biocompatibility, reliability, fault tolerance, and energy autonomy, as maintenance operations (e.g. battery replacement) are often invasive and risky.

3. **Ambient and Environmental Sensors**: non invasive sensors deployed in homes, nursing facilities, or hospitals to detect falls, monitor air quality, assess room occupancy, track movement patterns, or detect anomalies in daily activities. Typical examples include motion sensors, pressure mats, door and window sensors, and environmental probes for temperature, humidity, or CO_2. These devices are crucial in supporting assisted living, elderly care, and context aware monitoring, particularly for frail or cognitively impaired individuals.

4. **Ingestible and Disposable Sensors**: temporary diagnostic devices, including swallowable capsules and disposable skin patches or bandages, used for gastrointestinal monitoring, wound assessment, medication adherence tracking, or short term physiological data collection. Such devices extend the reach of Smart Health into areas that are otherwise difficult to assess through non invasive technologies, providing novel means for early diagnosis and targeted therapy.

Each category presents specific trade offs involving energy consumption, measurement accuracy, patient comfort, and data security. Wearable devices must balance battery life with sampling frequency and processing capabilities, while implantable devices require ultra low power designs and robust fault management strategies. Ambient sensors must ensure unobtrusiveness and low maintenance, whereas ingestible sensors must guarantee safety, biocompatibility, and complete elimination from the body after use.

Furthermore, achieving interoperable communication across diverse device manufacturers and platforms remains a major challenge. Standards such as IEEE 11073 for personal health devices, Open mHealth data schemas, and Continua Design Guidelines have been proposed to harmonize data formats and communication protocols, facilitating integration with electronic health record (EHR) systems and external platforms. Despite these efforts, fragmentation still persists, often requiring custom middleware and data harmonization strategies in real world deployments.

A growing trend is the adoption of **multimodal sensing** within single devices or coordinated device ensembles. By combining optical, motion, electrical, and biochemical sensing modalities, these systems can provide a richer characterization of the patient's status. For example, the joint analysis of ECG, PPG, and accelerometer data enables more robust estimation of cardiorespiratory parameters and physical exertion, while the fusion of environmental and behavioral data [123] can improve the detection of context in which physiological changes occur. This multimodal perspective is particularly beneficial when combined with advanced AI models capable of exploiting complex temporal and cross sensor relationships.

1.2.2 Applications and Benefits of Health IoT Systems

Health IoT systems support a broad spectrum of clinical and operational applications, covering all stages of the care continuum from prevention and early detection to treatment, rehabilitation, and long term follow-up. The following application areas illustrate how H-IoT technologies are reshaping healthcare delivery.

Remote Monitoring and Home Based Care

Remote Patient Monitoring (RPM) is one of the most widespread and impactful applications of Health IoT [105]. Its primary objective is to monitor patients outside traditional clinical settings by collecting continuous or frequent measurements of physiological and behavioral parameters [2]. This paradigm is particularly relevant for chronic diseases such as heart failure, diabetes [131], COPD, and hypertension, which require ongoing assessment and early detection of clinical deterioration.

An RPM system typically integrates wearable sensors (e.g. ECG patches, PPG wristbands, blood pressure monitors), smart home devices, and mobile health applications. Data are acquired in real time and transmitted through secure wireless protocols to cloud or edge platforms, where they undergo preprocessing and AI based analysis. Alerts are generated when anomalies or critical trends are detected and delivered to clinicians or caregivers.

RPM has been shown to reduce hospital readmissions, improve patient adherence, and enable earlier intervention. Continuous monitoring provides a richer and more representative picture of a patient's condition compared to sporadic in hospital measurements.

Key challenges include signal noise, patient compliance, data overload for clinicians, communication interruptions, and privacy concerns [13]. Standardized protocols are needed to integrate RPM outputs into clinical workflows.

Examples include wearable ECG patches for arrhythmia detection, smart blood pressure cuffs for hypertension control, and smart scales that characterize fluid retention in heart failure patients [19, 95].

Assisted Diagnosis and Decision Support

Assisted diagnosis is one of the most clinically validated applications of AIoT systems. By combining continuous sensing from IoT devices with advanced AI models, Smart Health platforms can support clinicians in identifying pathological conditions earlier, more accurately, and with reduced diagnostic ambiguity.

IoT devices such as ECG patches, PPG wearables, accelerometers, and environmental sensors collect granular physiological and behavioral data at frequencies far exceeding traditional hospital based measurements. These data streams capture transient episodes such as paroxysmal arrhythmias, respiratory instability, epileptic pre ictal states, or gait alterations that often go undetected during sporadic clinical evaluations.

ML and DL models transform raw signals into actionable diagnostic information by performing segmentation, feature extraction, classification, and anomaly detection. CNNs, LSTMs, and Transformer based architectures excel in analyzing morphological and temporal patterns in ECG, EEG, and multimodal sensor signals. Explainable AI (XAI) enhances transparency by providing saliency maps, feature attribution scores, and uncertainty estimates.

The output of AIoT driven diagnostic pipelines feeds into Clinical Decision Support Systems (CDSS), which alert clinicians when anomalies exceed predefined thresholds or when risk patterns emerge. Examples include arrhythmia detection from wearable ECG patches, respiratory anomaly monitoring, and early identification of neurological impairments through gait and tremor analysis.

These systems reduce diagnostic delays, improve triage efficiency, and decrease human error by highlighting clinically relevant events and filtering irrelevant data. CDSS platforms supported by AIoT enhance diagnostic objectivity and reproducibility.

Challenges include noise in real world data, device heterogeneity, false-positive management, regulatory requirements, and cybersecurity risks.

Predictive and Preventive Medicine

Predictive and preventive medicine seeks to anticipate clinical deterioration before symptoms arise and to prevent adverse outcomes through early intervention. AIoT systems are uniquely suited for this paradigm thanks to their ability to collect continuous, longitudinal physiological data and extract predictive biomarkers.

Physiological signals such as ECG, PPG, accelerometry, respiration, and glucose levels enable models to forecast events including:

- heart failure decompensation,
- arrhythmia onset,
- hypoglycemic episodes,
- epileptic seizures,
- respiratory distress.

Temporal AI models (LSTMs, Transformers, Temporal CNNs) detect subtle deviations from patient specific baselines.

AIoT systems build individualized baselines and detect deviations indicating early stage deterioration, enabling prioritization and targeted follow-up.

When deterioration risk increases, AIoT systems can:

- alert clinicians,

- recommend lifestyle or treatment adjustments,
- trigger closed loop therapies (e.g. insulin pumps),
- suggest preventive diagnostic tests.

Large scale IoT infrastructures allow detection of environmental and behavioral risk factors, supporting digital epidemiology.

These include the need for large datasets, cross population generalization, alarm calibration, and strict privacy management.

Rehabilitation, Assisted Living, and Smart Environments

AIoT technologies provide essential support in rehabilitation, functional assessment, and independent living particularly for elderly individuals and patients with neurological or motor impairments.

Wearable inertial sensors, smart garments, and sensorized home environments continuously track gait cycles, joint angles, muscle activation, and postural transitions. AI models evaluate motor performance, detect compensatory behaviors, and quantify adherence to rehabilitation protocols.

AI algorithms adapt rehabilitation exercises to patient performance, providing real-time feedback through haptic, auditory, or visual cues. These capabilities support tele-rehabilitation and remote physiotherapy.

Ambient IoT and wearables support:

- fall detection and fall prediction,
- monitoring Activities of Daily Living (ADLs),
- sleep analysis,
- detection of behavioral anomalies related to cognitive decline.

Long term data analysis identifies deviations from routines that may indicate dementia progression, depression, frailty, or loss of autonomy.

Challenges include device intrusiveness, battery limitations, elderly acceptance, calibration complexity, and privacy concerns.

Operational Efficiency and Smart Hospitals

Smart hospitals integrate AIoT technologies to enhance operational workflows, patient safety, and resource allocation.

IoT tracking (RFID, BLE, UWB, LoRaWAN) enables real-time localization of equipment such as infusion pumps, ventilators, and portable monitors, reducing delays and losses.

Smart beds and sensorized wards monitor posture, mobility, vital signs, and environmental factors to detect deterioration early and prevent complications.

AIoT systems assist with patient flow optimization, bed assignment, emergency triage, and medication logistics. Predictive models identify bottlenecks and anticipate resource needs.

Aggregated data enable administrators to identify inefficiencies, adapt staffing, and optimize resource planning.

Secure communication, device authentication, and adherence to standards (HL7 FHIR, DICOM, IEEE 11073) are essential for safe integration.

AIoT enabled hospitals demonstrate:

- reduced operational costs,
- improved staff efficiency,
- enhanced patient safety,
- improved decision making.

1.2.3 Challenges and Emerging Trends in Health IoT

Despite their potential, H-IoT systems face a range of challenges that hinder large scale, routine adoption.

From a technical standpoint, **data quality** is a critical issue. Signals acquired in daily life are often affected by noise, motion artifacts, missing segments, and variable adherence to wearing protocols. Robust preprocessing pipelines and artifact detection algorithms are therefore necessary to avoid misleading interpretations. Moreover, the heterogeneity of devices, firmware versions, and communication protocols complicates integration and maintenance, especially in multi vendor environments.

Security and privacy are particularly sensitive in healthcare due to the critical nature of medical data. The distributed and often wireless nature of H-IoT networks increases the attack surface, making them vulnerable to unauthorized access, data tampering, and denial of service attacks. To mitigate these risks, strong encryption, authentication mechanisms, secure key management, and continuous monitoring of network behavior are required. Regulatory frameworks such as GDPR [35] in Europe and HIPAA [116] in the United States impose strict requirements on data handling, storage, and consent management, which must be carefully taken into account when designing and deploying H-IoT solutions.

Interoperability remains an open challenge. While standards like HL7 FHIR facilitate data exchange between clinical systems, many consumer grade IoT devices still use proprietary formats and closed platforms, limiting seamless integration. Middleware solutions and semantic interoperability layers are often needed to map heterogeneous data into common representations and terminologies (e.g. SNOMED CT, LOINC). Without such harmonization, it is difficult to aggregate data across devices, institutions, and studies, thereby limiting the generalizability of AI models and the scalability of Smart Health services.

Another important dimension concerns **clinical validation and regulatory approval**. For H-IoT systems that support medical decision making, robust validation in real world conditions and compliance with medical device regulations are mandatory. This requires prospective studies, rigorous evaluation of performance metrics, and clear demonstration of clinical benefit. The dynamic nature of AI enabled H-IoT systems, which may adapt or be retrained over time, poses additional challenges for traditional regulatory frameworks that assume static devices.

Recent technological trends are helping to address some of these limitations. **Edge and federated learning** enable decentralized AI computation while preserving data locality, reducing the need to transmit sensitive information to central servers. **Blockchain based**

frameworks are being explored for tamper evident logging and fine grained access control to medical data. In parallel, the integration of multimodal sensing and advanced AI algorithms is enabling more accurate, robust, and context aware health assessments, even in the presence of noisy or incomplete data.

In summary, the Health IoT ecosystem is gradually evolving from isolated telemonitoring solutions to comprehensive, intelligent, and interoperable Smart Health platforms. Understanding its technological foundations, benefits, and challenges is essential for contextualizing the research contributions of this thesis, which build precisely on the convergence of IoT infrastructures, AI methods, and physiological data analysis.

1.3 Artificial Intelligence

Artificial Intelligence (AI) is one of the most influential fields in modern computer science and its origins trace back to the mid 20th century, when researchers first began exploring whether machines could emulate human reasoning. The term "Artificial Intelligence" was formalized by John McCarthy during the 1956 Dartmouth Conference, marking the official birth of AI as an academic discipline [121]. Early AI research focused on symbolic reasoning through rule based systems, where knowledge was encoded using logical rules and expert crafted representations.

A major paradigm shift occurred with the rise of ML, which reframed intelligence as the ability of algorithms to identify patterns and make decisions directly from data rather than predefined rules. ML encompasses three fundamental learning paradigms:

- **Supervised Learning:** Algorithms learn from labeled datasets to perform prediction or classification. This paradigm underpins many medical applications, such as diagnosing diseases from ECG signals or predicting clinical outcomes.

- **Unsupervised Learning:** Models discover hidden structures in unlabeled data, enabling clustering, anomaly detection, or dimensionality reduction. In healthcare, unsupervised models support patient stratification, phenotype discovery, and identification of abnormal physiological patterns.

- **Reinforcement Learning (RL):** An agent learns optimal behaviors through trial and error interactions with an environment, receiving rewards or penalties as feedback. RL is increasingly used in healthcare for optimizing treatment policies and adaptive monitoring strategies.

In the 2010s, the emergence of DL, a subset of ML based on artificial neural networks, further accelerated the progress of AI. DL models automatically learn hierarchical representations from raw data, eliminating the need for manual feature extraction. This capability has driven breakthroughs in computer vision, speech recognition, and natural language processing, and has had a profound impact on biomedical data analysis.

1.3.1 Machine Learning in Smart Health

Within the Smart Health domain, ML plays a central role in transforming clinical, behavioral, and physiological data into actionable insights. ML models can process heterogeneous modalities ranging from electronic health records and laboratory measurements to real time wearable sensor data and supporting tasks such as disease prediction, risk stratification, and early anomaly detection.

A key component of ML based health analytics is the extraction of meaningful features from raw physiological signals. Time-domain metrics (e.g. variability indices, entropy), frequency-domain descriptors (e.g. Fourier or Wavelet coefficients), and morphology-based features are commonly used to characterize signals such as ECG, EEG, or accelerometry. Traditional ML algorithms (e.g. Random Forests, Gradient Boosting, SVMs) are particularly effective when transparency and clinical interpretability are required properties that are highly valued in regulated medical contexts.

In resource constrained environments, such as wearable or IoT based systems, lightweight ML models enable on device analytics, reducing latency [46] and preserving patient privacy by limiting data transmission. This makes ML an integral component of real-time monitoring platforms for chronic disease management, physical activity tracking, and personalized healthcare interventions.

1.3.2 Deep Learning in Smart Health

DL has expanded the analytical capabilities of Smart Health systems by enabling end-to-end processing of high dimensional biomedical data. Convolutional Neural Networks (CNNs) are widely used for medical imaging tasks (tumor detection, anatomical segmentation, radiological triage etc) often achieving performance comparable to expert clinicians.

For sequential physiological data, architectures such as Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM) networks, and more recently Transformers, capture long range temporal dependencies and complex nonlinear dynamics. These models have been successfully applied to ECG based arrhythmia detection, EEG based seizure prediction, respiratory monitoring, and multimodal fusion of wearable sensor streams. Despite their accuracy, DL models often face challenges related to interpretability, computational cost, and regulatory acceptance.

Overall, the integration of ML and DL within Smart Health ecosystems represents a major step toward personalized, predictive, and preventative medicine, enabling continuous and data driven insights that can significantly enhance clinical care and patient well being.

1.4 The Convergence of AI and IoT in Smart Health

The integration of **Artificial Intelligence (AI)** and the **Internet of Things (IoT)** has given rise to a new technological paradigm known as the *Artificial Intelligence of Things (AIoT)*. This convergence brings together the sensing and connectivity capabilities of IoT with the cognitive and analytical power of AI, enabling systems that are both context aware and autonomous [9]. Within the healthcare domain, AIoT represents a decisive step toward the realization of truly intelligent and distributed health systems that can perceive, reason, and act in real time, enabling advanced diagnostic, monitoring, and decision-support capabilities in modern healthcare systems [34]. In a traditional IoT environment, devices primarily act as *data collectors*, capturing physiological or environmental information and transmitting it to a centralized server for processing. However, as the number of connected devices increases exponentially, this model becomes unsustainable due to limitations in network bandwidth, latency [46] and privacy [13]. AI integration fundamentally reshapes this workflow by embedding intelligence at multiple levels of the IoT architecture. A typical AIoT based healthcare pipeline can be conceptualized as follows:

1. **Data Acquisition:** IoT sensors continuously capture physiological signals such as electrocardiograms (ECG), photoplethysmograms (PPG), temperature, or motion data.

2. **Data Transmission and Preprocessing:** Raw data are filtered, normalized, and transmitted via secure wireless protocols to local gateways or edge nodes.

3. **Feature Extraction and Analysis:** At the edge or fog layer, AI models perform lightweight analytics detecting anomalies, segmenting relevant time windows, or extracting statistical and spectral features.

4. **Model Inference and Decision Support:** DL or ensemble ML algorithms analyze aggregated data to provide diagnostic outcomes or risk assessments.

5. **Feedback and Actuation:** Based on model predictions, alerts or recommendations are sent to clinicians or patients, and, in some cases, automatic actions (e.g. drug release, emergency notifications) can be triggered.

AI enhances IoT systems by transforming raw sensor data into **actionable knowledge**, while IoT empowers AI with access to diverse, high-frequency, and real world data. As a result, the combined system exhibits properties such as autonomy, self learning, and contextual awareness, which are essential for next generation Smart Health services.

1.4.1 Advantages of AIoT in Healthcare Systems

The convergence of AI and IoT brings several advantages that make healthcare systems more efficient, resilient, and adaptive.

- **Real-Time Intelligence:** On device inference allows the immediate analysis of critical signals (e.g. heart rhythm irregularities [19, 95]) without relying on remote servers. This is crucial in emergency care and intensive monitoring scenarios.

- **Personalization and Adaptivity:** AI models can continuously learn from individual patient data, adjusting predictions to personal baselines and clinical histories. This supports the paradigm of *personalized medicine*.

- **Scalability:** Edge and cloud integration enables scalable deployment across multiple users and healthcare facilities, allowing collective learning from distributed data sources.

- **Operational Efficiency:** AI driven automation reduces the workload of healthcare professionals by filtering irrelevant information and highlighting clinically significant events.

- **Privacy Preservation:** By performing analytics locally at the edge, AIoT minimizes the transmission of sensitive data, reducing exposure to potential security risks.

These benefits contribute to a significant improvement in both patient outcomes and system level performance. For instance, predictive models can detect anomalies hours or even days before symptom onset, enabling early interventions and reducing hospitalization rates. Moreover, the integration of explainable AI (XAI) techniques ensures that model outputs remain interpretable by clinicians, reinforcing trust and accountability in automated decision processes.

1.4.2 Applications of AIoT in Smart Health

AIoT has been successfully applied in several healthcare domains, demonstrating its versatility and clinical relevance.

1. **Cardiovascular Monitoring:** Edge based ML and DL algorithms can identify arrhythmias [21, 131], ischemia, or paroxysmal atrial fibrillation (PAF) from continuous ECG signals. Lightweight models such as Random Forest or XGBoost provide interpretable and energy efficient solutions for wearable devices, while CNN and LSTM architectures achieve higher accuracy in cloud assisted analysis.

2. **Neurological and Cognitive Health:** IoT wearables combined with AI can monitor tremor intensity, sleep patterns [100], and gait abnormalities, supporting early diagnosis

of neurodegenerative diseases such as Parkinson's [114] or Alzheimer's. Edge inference allows patients to be monitored unobtrusively in their daily environments.

3. **Chronic Disease Management:** AIoT systems can continuously track glucose levels, medication adherence, or blood pressure to support patients with diabetes [131] and hypertension [25]. Predictive analytics identify risk trends and suggest personalized adjustments to treatment plans.

4. **Rehabilitation and Assisted Living:** Motion sensors, combined with AI driven analysis of movement quality, can assist in post stroke rehabilitation, balance monitoring, or fall prevention in elderly populations.

5. **Public Health and Epidemiology:** Large scale IoT networks integrated with AI enable real-time epidemiological monitoring, contact tracing, and resource optimization in pandemic response scenarios.

These examples illustrate how AIoT systems bridge the gap between raw physiological data and intelligent, adaptive healthcare services. They enable both individual level monitoring and population level insights, reinforcing the central role of digital technology in the future of medicine.

1.4.3 Challenges and Future Directions

Despite their potential, AIoT systems in healthcare still face several challenges. These include the need for robust interoperability among heterogeneous devices, standardized data representation, and compliance with privacy [13] and ethical regulations. Moreover, many AI models are computationally intensive and require optimization for deployment on resource constrained devices.

Another key issue is the **trustworthiness** of AIoT decisions. Ensuring model transparency, fairness, and explainability is crucial for clinical acceptance. Techniques such as model distillation, uncertainty quantification, and explainable DL are being actively explored to bridge the gap between performance and interpretability.

Finally, the emergence of federated and distributed learning paradigms is reshaping the development of AIoT in healthcare. These approaches allow collaborative model training across multiple hospitals or devices without centralizing patient data, thereby addressing privacy [13] and data governance concerns.

In the coming years, the synergy between AI and IoT is expected to evolve toward **self-organizing, adaptive, and privacy preserving healthcare ecosystems**, where each component contributes intelligently to patient well being while respecting ethical and societal constraints.

CHAPTER

2 Study Definition and Planning

The convergence between artificial intelligence (AI) and wearable technologies is profoundly reshaping the Smart Health landscape. This interaction has attracted growing attention from both academia and industry, as it offers unprecedented opportunities for improving healthcare delivery and enhancing individuals' well-being.

Wearable devices, equipped with increasingly miniaturized and sophisticated sensors, represent the foundation of modern digital health ecosystems. They continuously acquire a wide range of physiological, behavioral, and movement-related data—such as heart rate [85], blood pressure [26], sleep quality [100], physical activity levels [113], and blood oxygen saturation (SpO_2) [28]. When combined with contextual, environmental, and personal health information, these data streams provide a rich and multidimensional portrait of an individual's health status and lifestyle patterns.

Nevertheless, the large volume, heterogeneity, and temporal granularity of wearable-generated data introduce considerable analytical challenges. Extracting clinically meaningful information requires advanced computational tools capable of identifying subtle patterns, modelling temporal dependencies, and supporting personalized health assessment. In this regard, artificial intelligence plays a pivotal role.

ML and DL methods have shown impressive capabilities in converting raw sensor data into actionable insights, forecasting health trajectories, and enabling tailored interventions. Their ability to learn complex relationships and continuously improve through data makes them particularly suitable for real-time and adaptive health monitoring scenarios.

Motivated by this context, the present research aims to conduct a systematic examination of the current landscape of AI applied to wearable-based Smart Health systems. The objective is to provide a comprehensive overview of existing methodologies, emerging trends, and the main challenges characterizing this field. In particular, the study pursues the following goals:

- analyse the most recent and relevant advancements in AI-driven Smart Health applications;
- identify whether ML or DL approaches are predominantly employed in current research contributions;
- assess the performance, robustness, and reliability of AI models designed for wearable-based health monitoring and prediction;
- evaluate the feasibility of translating these AI solutions into real-world clinical and everyday health scenarios.

The rapid evolution of digital technologies has profoundly reshaped modern healthcare systems, giving rise to the paradigm of *Smart Health*. Combining real-time data acquisition and artificial intelligence, Smart Health enables continuous, personalized, and proactive care, moving beyond the traditional episodic and hospital-centered model. The combination of Internet of Things (IoT) devices and advanced Artificial Intelligence (AI) techniques represents one of the most promising drivers of this transformation. IoT systems allow the acquisition of physiological, behavioral, and environmental data with unprecedented frequency and granularity, while AI algorithms, especially ML and DL, enable the extraction of complex, clinically relevant patterns from these heterogeneous data streams.

However, despite the vast potential demonstrated in the literature, the integration of IoT and AI in real health applications introduces several scientific and technological challenges. IoT data are often heterogeneous, noisy, and susceptible to missing values; they may exhibit temporal irregularities, sensor drift, or variability due to patient-specific factors. In parallel, AI models require large, representative, and well-curated datasets, and must be validated not only statistically but also clinically, ensuring robustness, interpretability, and fairness.

The broader objective of this thesis is to systematically explore the opportunities and limitations of applying AI to IoT-generated health data, combining a comprehensive analysis of the literature with empirical experimentation and model development. To achieve this, the research pipeline is structured into three interconnected phases. First, a systematic review [9] consolidates the state of the art, highlighting current technological approaches, methodological trends, and persistent challenges in merging IoT and AI for Smart Health. Second, an exploratory analysis [10] is conducted on real IoT-derived health data to evaluate the feasibility and potential of ML techniques in extracting meaningful information and supporting preliminary diagnostic tasks. Finally, the third phase [8] focuses on the design, development, and evaluation of a DL model tailored to Smart Health scenarios, with the goal of enhancing diagnostic accuracy, predictive capabilities, and overall decision support.

Phase 1 - Systematic Review of IoT and AI in Smart Health

The first phase involved a rigorous and structured systematic review of the literature, aimed at mapping existing research on the integration of IoT, ML, and DL in healthcare. Given the fragmented nature of the domain, where studies often focus on individual diseases, specific sensor modalities, or isolated AI techniques, this review served multiple purposes:

- assess how IoT technologies are currently used to collect health-related data;
- analyze the ML and DL approaches applied to these data and their performance;
- identify recurring methodological strengths and weaknesses;
- examine challenges in scalability, interoperability, data quality, privacy, and interpretability;
- highlight research gaps and open challenges that motivate new methodological contributions.

ML algorithms, particularly when applied to traditional classification tasks, are often preferred because they offer greater interpretability and require fewer computational resources. In contrast, training reliable DL models typically requires large and well-annotated datasets.

Although the availability of healthcare data has increased significantly in recent years, not all medical applications—especially those involving rare diseases or events—are well suited to DL approaches [90].

As the volume of available data continues to grow, DL technologies are increasingly capable of supporting disease diagnosis and predictive analytics by identifying complex patterns that may be difficult for traditional analytical methods to detect. This capability has the potential to reduce ambiguity in clinical decision-making and support the development of more intelligent diagnostic systems. Furthermore, one of the long-term challenges for DL lies in its ability to integrate heterogeneous data sources originating from different health informatics domains, paving the way for more comprehensive precision medicine frameworks.

ML models usually require fewer computational resources compared to DL models, particularly those composed of multiple neural layers. DL architectures often demand substantial computing power and longer processing times, both during training and inference. A critical avenue for future research also concerns the enhancement of privacy and security mechanisms for health data, as these aspects remain insufficiently explored in the current literature.

Traditional ML techniques are generally based on handcrafted feature extraction. These approaches often require lower computational resources, can be trained on smaller datasets, and usually provide a higher degree of interpretability, which is particularly valuable in clinical contexts where transparency and traceability of decisions are essential.

By contrast, DL approaches enable end-to-end learning directly from raw or minimally processed signals, making them more suitable for capturing complex temporal dependencies, nonlinear relationships, and subtle morphological variations in physiological data. This often results in improved predictive performance, particularly in the analysis of long-term and high-dimensional signals such as ECG recordings. However, these advantages are counterbalanced by higher computational demands, reduced interpretability, and a stronger dependence on large, well-annotated datasets.

Therefore, the comparison emerging from the literature suggests that the choice between ML and DL cannot be based solely on predictive accuracy. Despite the superior performance of DL models, traditional ML approaches may remain preferable in scenarios characterized by limited data availability, strict computational constraints, or the need for high interpretability, such as edge-based or safety-critical clinical applications. In Smart Health and clinical monitoring scenarios, it must also account for robustness, explainability, computational feasibility, and real-world deployment constraints. This methodological tension between interpretability and predictive performance provides an important rationale for the subsequent phases of this thesis, in which both paradigms are explored and experimentally compared in the context of long-term ECG analysis for atrial fibrillation detection.

Phase 2 - Exploratory Data Study and Preliminary ML Experiments

Building on the conceptual understanding obtained from the literature, the second phase focused on the empirical exploration of a real IoT-derived health dataset. Before designing complex architectures or multimodal learning systems, it was essential to evaluate the feasibility of extracting predictive information from the available data and to examine how classical ML approaches perform when applied to physiological measurements collected in real-world settings. This phase consisted of several steps:

- Data inspection and cleaning: assessment of signal quality, missing values, sensor inconsistencies, and temporal alignment.
- Exploratory analysis: study of the distribution of physiological features, correlations among variables, and potential biomarkers.
- Feature engineering: extraction of time-domain, frequency-domain, and statistical features connected to the physiological nature of the signals.
- Baseline ML experiments: evaluation of classical models such as Random Forests, SVMs, Gradient Boosting, and KNN to determine whether IoT-derived features provided discriminative power.

The objective of this stage was to present a lightweight and interpretable ML pipeline. The resulting framework is computationally efficient and explainable, making it particularly suitable for deployment on resource-constrained systems such as wearables or edge devices.

Despite the promising results, this study has several limitations. First, the evaluation was performed on a single dataset, which, although large and detailed, may not fully reflect the variability encountered in different patient populations or acquisition settings. Second, the current segmentation and classification process treats each ECG segment independently, without modeling temporal dependencies or patient-level context. Third, the approach relies exclusively on handcrafted features. While effective, this strategy may miss complex patterns that could be captured by DL architectures.

Phase 3 - Model Design, Implementation and Evaluation

The third phase represents the core methodological contribution of the thesis. Leveraging the insights from both the systematic review and the exploratory analysis, this phase aimed to design, implement, and evaluate an AI model capable of processing IoT-related health data in a clinically meaningful way. Key objectives of this phase included:

- designing a model architecture capable of capturing temporal dependencies and physiological dynamics in IoT data;
- ensuring robustness to noise and missing values, which are inherent in IoT-based monitoring;
- exploring the contribution of ML and DL techniques within a unified methodological framework;
- optimizing the computational efficiency of the model for potential edge or mobile deployment;
- ensuring interpretability and transparency to facilitate clinical adoption.

The analysis confirmed that the DL model not only minimized false positives but also improved sensitivity to AF episodes, highlighting its robustness and reliability in clinical scenarios. Overall, the comparative analysis demonstrates the advantage of end-to-end learning strategies over handcrafted feature extraction. While ML classifiers remain useful for rapid prototyping, DL architectures are better suited to capture subtle and long-range dependencies in ECG signals.

The next paragraphs will describe the objectives and research questions that will guide the research work.

Section 2.1 presents the research questions concerning the first phase of the project development, namely ”Preliminary study and application of artificial intelligence techniques” will be discussed; in section 2.2, we find those relating to phase 2, namely ”Collection and analysis of heterogeneous clinical data” and finally in section 2.3 we find those relating to the third and final phase ”Model design and implementation”.

2.1 Preliminary Study and Application of AI Techniques

The first phase of this research was devoted to a systematic investigation of the existing literature according to the guidelines of the Kitchenham method [56], with the aim of exploring the research landscape regarding the application of Artificial Intelligence (AI) techniques in conjunction with Internet of Things (IoT) technologies within Smart Health applications.

In recent years, a large number of studies have explored how wearable devices, body sensors, and IoT platforms can be combined with ML and DL methods to support diagnosis, monitoring, prediction, and decision support in healthcare. However, these contributions are often fragmented: they focus on specific diseases, narrow use cases, single sensor types, or isolated algorithmic solutions. A structured and critical analysis of this fragmented landscape was therefore required to:

- understand which technological approaches are currently adopted to integrate IoT and AI in Smart Health;
- identify recurring challenges and limitations from both a technical and clinical perspective;

- assess the effective contribution of ML and DL models in extracting clinically meaningful information from IoT-generated data;
- clarify to what extent these models can empirically improve diagnostic accuracy, predictive capabilities, and decision support.

Research Questions

With the objective of identifying the main technological approaches, existing challenges, and open research gaps that would inform the subsequent phases of this work, a systematic review was conducted in accordance with established methodological guidelines from evidence-based software engineering and medical informatics.

The search strategy targeted articles combining the terms *Artificial Intelligence*, *Smart Health*, and *Wearable / IoT devices*, and was applied to major scientific databases (including, for example, IEEE Xplore and ScienceDirect).

After applying inclusion and exclusion criteria related to language, time frame, relevance to IoT-based Smart Health, and explicit use of ML or DL methods, a final set of studies was retained for in-depth analysis. The selected articles were analyzed along multiple dimensions, including:

- type of IoT devices and sensing modalities employed (e.g. wearable sensors, environmental sensors, mobile devices);
- data types and pre-processing strategies (physiological signals, activity traces, clinical records, multimodal data);
- AI techniques used ML algorithms, DL architectures, hybrid approaches);
- targeted application domains (diagnosis, monitoring, prediction, rehabilitation, risk assessment, etc.);
- performance metrics and reported results;
- considerations on privacy, security, and deployment constraints.

Starting from the evidence collected in the systematic review, the Research Questions (RQs) were formulated to capture the technological, methodological, and clinical dimensions of IoT–AI integration in Smart Health.

RQ_1: What are the current technological approaches, challenges, and limitations in integrating IoT and AI within Smart Health systems?

RQ_1 focuses on the *integration layer* between IoT and AI in Smart Health. Rather than examining algorithms in isolation, this question considers the entire technological pipeline, including:

- the types of IoT and wearable devices used for health monitoring (e.g. ECG patches, smartwatches, inertial sensors, smart textiles);
- the architectures adopted to connect devices, gateways, and cloud or edge platforms (e.g. edge/fog computing, cloud-centric solutions);
- the strategies used for data acquisition, synchronization, pre-processing, and feature extraction;
- the main application scenarios (monitoring, diagnosis, prediction, rehabilitation, assisted living, smart hospitals);
- the challenges encountered in practice, such as data quality issues, sensor heterogeneity, battery constraints, communication latency, and limited standardization;
- non-technical constraints, including privacy regulations, security vulnerabilities, clinical acceptance, and usability.

Through this question, the preliminary study aims to build a comprehensive picture of the *current technological landscape* of AI-enabled IoT in Smart Health and to identify structural limitations that must be addressed by future research.

The performance of algorithms employed in smart health applications was evaluated with particular attention to their accuracy and reliability in the analysis, processing, and classification of data collected by IoT devices in healthcare contexts.

The first step consisted in examining the AI methodologies adopted in the reviewed studies, with the goal of assessing the complexity and depth of the proposed models.

The choice between ML and DL approaches depends on multiple factors, including data availability, since DL training typically requires large, well-annotated datasets, whereas ML techniques can operate effectively even with more limited data. Another determining factor is the availability of computational resources: DL models generally demand higher computational power both during training and inference, while ML methods tend to be more lightweight and computationally efficient.

RQ_2: What are the privacy and security issues taken into account while using health data collected from IoT devices?

A crucial dimension of Smart Health concerns the security and privacy of medical data, as health and medical data is highly sensitive and, therefore, its security and privacy are of paramount importance. Therefore, in order not to overlook the issues related to these approaches, we focused on the privacy and security issues faced in the collection and transmission of health data by IoT devices.

The adoption of IoT devices for continuous monitoring introduces new vulnerabilities throughout the data lifecycle, from acquisition to transmission, storage, and algorithmic processing. For this reason, particular attention was dedicated to examining the privacy and security issues associated with IoT-based health data, thereby addressing RQ_2.

Health IoT ecosystems typically rely on wearable sensors, mobile gateways, wireless communication protocols, and cloud or edge infrastructures. Each of these layers introduces potential risks. At the device level, limited computational capabilities often prevent the implementation of strong encryption, making IoT sensors susceptible to unauthorized access, tampering, or data injection attacks. During transmission, wireless channels can be exposed to eavesdropping, replay attacks, or man-in-the-middle attacks, especially when low-power communication protocols lack robust authentication mechanisms. On the server side, large-scale storage of medical records raises concerns about data breaches, identity theft, and misuse of personal information.

In parallel, the integration of AI models into Smart Health systems introduces additional privacy considerations. ML and DL algorithms require substantial amounts of data for training, raising the risk of overexposure of personal health information.

Beyond technical vulnerabilities, there are also broader ethical concerns. Patients may be unaware of the extent of data collected by wearable devices, the entities that have access to such data, or the purposes for which it is processed. Furthermore, the frequent transmission of raw biometric signals may increase the attack surface, while poor configuration, outdated firmware, or lack of secure update mechanisms can further compromise privacy.

2.2 Analysis of IoT Health Data Using Machine Learning Techniques

The second phase of this research focused on an exploratory analysis conducted on a real IoT-based biomedical dataset related to paroxysmal atrial fibrillation (PAF). This phase served as a bridge between the theoretical insights obtained from the systematic review and the design of the final AI model, and its main purpose was to empirically assess whether ML techniques could extract clinically meaningful information from physiological signals acquired through wearable and IoT systems.

In contrast with the first phase, which provided a conceptual and methodological understanding of the state of the art, this second phase adopted a data-driven perspective. The goal was to test, in practice, how traditional ML algorithms perform when applied to real sensor data characterized by noise, temporal variability, missing segments, and the typical imperfections of IoT acquisition environments.

The exploratory study involved several steps:

- **Data inspection and preprocessing**: analysis of signal distribution, detection of noise sources and artifacts, and application of filtering, normalization, and feature extraction techniques suited for IoT-collected physiological traces.

- **Feature engineering**: extraction of time-domain and frequency-domain descriptors commonly used in arrhythmia detection, including statistical metrics, morphological characteristics, and spectral components.

- **Application of ML models**: implementation and evaluation of various ML classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Random Forests, and Gradient Boosting methods, selected based on their suitability for small-to-medium datasets and interpretable decision boundaries.

- **Performance assessment**: evaluation of accuracy, sensitivity, specificity, and F1-score to determine whether these models could effectively discriminate between PAF and non-PAF segments under realistic IoT conditions.

The findings from this phase highlighted both the opportunities and the limitations of using classical ML techniques in Smart Health scenarios. On one hand, several models demonstrated good discriminatory power, confirming that IoT-generated physiological data contain meaningful diagnostic information that can be extracted even with lightweight algorithms. On the other hand, the performance variability across models and the sensitivity to data preprocessing revealed the intrinsic complexity of working with real-world IoT signals, suggesting the need for more robust and adaptable approaches.

For this reason, the outcomes of this phase directly motivated the transition toward the third stage of this thesis, which investigates DL models capable of learning representations directly from raw physiological signals. The empirical insights obtained here were essential to understand the strengths and weaknesses of ML in IoT-based health analytics and to guide the architectural choices of the final AI solution.

Research Questions

The exploratory analysis conducted in this second phase played a crucial role in shaping the research direction of the thesis. While the systematic review carried out in the first phase highlighted a fragmented scientific landscape, characterized by heterogeneous methodologies, limited real-world validations, and scarce ML IoT integrations, the hands on analysis of the IRIDIA-AF dataset provided a concrete opportunity to evaluate how ML behaves when confronted with real IoT-generated physiological signals.

This combination of theoretical insight and empirical experimentation enabled the identification of the key challenges that arise in practice when applying AI to wearable-derived biomedical data. Issues such as signal noise, temporal irregularities, inter-patient variability, and the sensitivity of classical ML models to preprocessing steps became evident during this exploratory phase. At the same time, several models demonstrated promising discriminatory power, confirming the presence of clinically meaningful information in IoT-acquired ECG data.

Building on these observations, the research questions for this phase were defined with the objective of assessing the concrete potential of ML for Smart Health and of understanding its role in diagnostic and predictive contexts. The following RQs therefore summarize the core investigative goals pursued in this phase.

***RQ$_3$*: What is the potential of ML techniques in extracting and analyzing clinically meaningful information from IoT-generated health data?**

This question aims to determine whether traditional ML algorithms are capable of processing physiological signals acquired through wearable IoT devices in a reliable and clinically relevant manner. The exploratory study sought to evaluate:

- the extent to which handcrafted temporal and spectral features extracted from ECG signals can capture patterns associated with paroxysmal atrial fibrillation (PAF);
- whether classical ML models (e.g. SVM, k-NN, Random Forests, Gradient Boosting) can effectively learn from these features despite the inherent imperfections of IoT-acquired data;
- the sensitivity of model performance to preprocessing quality, noise removal, and feature

engineering choices;

- the overall feasibility of employing lightweight ML solutions in Smart Health scenarios where computational resources and power consumption may be constrained.

The purpose of RQ_3 is to establish a baseline understanding of how much clinically meaningful information can be extracted using conventional ML pipelines before advancing toward more sophisticated DL architectures.

RQ_4: What extent can Ml models enhance diagnostic accuracy, predictive capabilities, and decision support in Smart Health?

While RQ_3 focuses on the technical feasibility of feature extraction and pattern recognition, RQ_4 addresses the broader clinical relevance and operational impact of ML in Smart Health. Specifically, this question investigates:

- how accurately ML models can discriminate between PAF and non-PAF episodes in long-term wearable ECG recordings;
- whether these models can support early detection, risk stratification, or continuous monitoring tasks typical of IoT-based healthcare systems;
- the potential of ML to reduce diagnostic delays or provide real-time alerts in wearable monitoring scenarios;
- the positioning of ML as a decision-support tool in remote cardiac monitoring, especially in comparison with traditional rule-based or clinician-driven approaches.

By evaluating the diagnostic performance of ML models within the constraints of IoT-based signal acquisition, RQ_4 aims to determine whether classical ML approaches can represent a meaningful step toward intelligent, continuous, and patient-centered cardiovascular monitoring systems.

Together, these research questions provide a structured analytical framework for understanding the strengths and limitations of ML in Smart Health and pave the way for the third phase of the thesis, which investigates whether DL can overcome the challenges observed in this exploratory study.

2.3 Model Design and Implementation

The third phase of this research focused on the development, implementation, and evaluation of a DL model specifically tailored for the automated detection of paroxysmal atrial fibrillation (PAF) using the IRIDIA–AF dataset.

Building upon the insights gained from the exploratory ML study conducted in the second phase, this part of the work aimed to investigate whether DL architectures, which can automatically learn hierarchical representations from raw or minimally processed signals, could overcome the limitations observed in classical ML approaches when applied to IoT-generated ECG data.

While the ML based analysis confirmed that meaningful diagnostic information could be extracted from wearable ECG traces, it also highlighted several challenges intrinsic to IoT physiological data, such as temporal variability, noise, morphological distortions, and the high sensitivity of model performance to preprocessing choices and handcrafted features. These observations motivated the transition toward a DL approach capable of learning feature representations directly from the data, without relying exclusively on manual feature extraction.

Research Questions

The third phase of this research represents the natural continuation of the exploratory work conducted in the previous stage. After assessing the feasibility and limitations of classical ML techniques applied to IoT generated physiological signals, it became necessary to investigate whether more advanced Artificial Intelligence approaches, specifically *DL* architectures, could address the shortcomings observed in the ML based analysis and offer a substantial improvement in performance, robustness, and clinical relevance.

In contrast to traditional ML models, which rely heavily on hand crafted features and domain specific preprocessing, DL models are capable of learning hierarchical representations directly from raw or minimally processed data. This property is particularly valuable in Smart Health contexts, where IoT generated signals are often noisy, complex, and highly variable across time and individuals. Moreover, physiological signals such as ECG recordings contain rich temporal and morphological structures that DL models, especially convolutional and recurrent neural networks, are inherently designed to capture.

The experiences gained from the second phase highlighted several open questions: ML techniques showed promising results, but their performance was strongly influenced by feature engineering choices and preprocessing pipelines. Additionally, the presence of motion artefacts, irregular heart rhythms, missing samples, and the natural variability of IoT acquired ECG traces exposed the limits of traditional approaches. These considerations motivate the need for more expressive and resilient modelling techniques capable of automatically extracting clinically meaningful information and generalizing effectively in real world IoT scenarios.

For these reasons, the third phase focuses on the design, training, and evaluation of a DL model built on the same IRIDIA-AF dataset used in the ML based exploratory study. This enables a direct, controlled comparison between ML and DL approaches, allowing us to quantify the actual contribution of deep architectures to Smart Health applications and to understand whether they can offer concrete advantages in terms of diagnostic accuracy, predictive capability, and decision support.

RQ_5: What is the potential of DL techniques in extracting and analyzing clinically meaningful information from IoT-generated health data?

This research question aims to assess to what extent DL architectures can learn robust and clinically meaningful representations from physiological signals acquired through wearable IoT devices, while minimizing the reliance on handcrafted feature engineering. The exploratory study seeks to evaluate:

- the ability of DL models to discover and encode complex, non-linear relationships within ECG signals, potentially overcoming the limitations of handcrafted temporal and spectral features in characterizing arrhythmic phenomena and distinguishing between PAF and non-PAF episodes;

- the robustness of these models to noise, motion artifacts, missing samples, and inter- and intra-subject variability, which are typical of long-term monitoring with wearable sensors, by analyzing the impact of different preprocessing, denoising, and data augmentation strategies;

- the sensitivity of model performance to architectural design choices (network depth, kernel size, attention mechanisms, normalization layers) and training hyperparameters (learning rate, batch size, regularization schemes, early stopping criteria);

- the feasibility of designing and deploying lightweight or compressed DL models (for example, via pruning, quantization, knowledge distillation, or inherently efficient architectures for edge computing) that comply with the memory, computational, and energy constraints of Smart Health and IoT environments.

This question aims to assess whether DL architectures, thanks to their ability to learn directly from raw ECG signals, can capture subtle temporal and morphological patterns that classical ML methods may overlook. It explores the suitability of different DL paradigms for processing IoT acquired physiological data and evaluates their ability to generate meaningful, stable, and clinically interpretable representations.

RQ_6: To what extent can DL models enhance diagnostic accuracy, predictive capabilities, and decision support in Smart Health?

Whereas RQ_5 focuses on the technical feasibility of representation learning and pattern discovery, RQ_6 addresses the broader clinical relevance and operational impact of DL in Smart Health scenarios. In this regard, the question investigates how far DL based systems can improve diagnostic performance, enable advanced predictive functionalities, and support clinical decision making in remote cardiac monitoring and continuous ECG analysis. More specifically, this research question examines:

- how accurately DL models can discriminate between PAF and non-PAF segments in long-term wearable ECG recordings, considering clinically meaningful metrics such as sensitivity, specificity, precision, recall, and area under the ROC curve, even in imbalanced and noisy conditions;

- whether such models can support early detection and risk stratification by forecasting impending arrhythmic events or continuously estimating indicators of cardiovascular risk, thereby enabling more timely and personalized interventions;

- the potential of DL–based systems to provide near real-time alerts or notifications within IoT-enabled telemedicine platforms, and the associated trade-offs between responsiveness, false alarm rates, and the cognitive workload imposed on clinicians and patients;

- the role of DL as a decision-support tool in remote cardiac monitoring, including its integration with explainability methods (e.g. saliency maps, attention-based visualization, post-hoc explainable AI techniques) and its comparative advantages or limitations with respect to rule-based systems and traditional ML approaches in terms of usability, clinical trust, and impact on workflow.

By evaluating diagnostic performance, predictive power, and integration into real world Smart Health infrastructures, RQ_6 aims to determine whether DL models can represent a substantive step toward intelligent, continuous, and patient centered cardiovascular monitoring solutions, and to what extent they can pave the way for increasingly autonomous and personalized digital health systems.

Together, these research questions define the analytical framework of the third phase and provide a structured basis for evaluating the real contribution of DL in IoT-driven Smart Health applications.

2.4 Workflow and Background

The methodology adopted to address the research questions was common to all of them, this is because they are all the result of the application of an artificial intelligence process to medical data. Therefore, regardless of the type of data considered, the steps to follow are the same, therefore this paragraph describes them briefly.

The design of an AI model in Smart Health requires a structured and iterative workflow that spans from understanding the data to the development, optimization, and evaluation of the final predictive system. While standard AI pipelines define a set of canonical stages, in practice these steps must be adapted to the characteristics of the data, the specific clinical task, and the constraints of IoT-based acquisition.

In this thesis, the methodological process has been tailored to the analysis of physiological signals collected from wearable devices and to the development of a DL model capable of detecting paroxysmal atrial fibrillation (PAF). As established in the previous phases, the predictive task addressed in this work is a supervised classification problem: given physiological time-series acquired in an IoT setting, the goal is to predict the disease.

The experience gained during the exploratory ML phase demonstrated the feasibility and limitations of traditional approaches, paving the way for the design of a more expressive and robust DL solution that can more effectively capture temporal and morphological patterns embedded in raw or minimally processed IoT data.

The main phases of the development pipeline, each of which will be detailed in the following sections, are summarised below:

- Data collection and understanding
- Preprocessing and segmentation
- Model design and training
- Hyperparameter optimisation
- Model evaluation

- Explainability

2.4.1 Data Collection and Understanding

In any artificial intelligence development pipeline, the quality of the resulting model is strongly influenced by the characteristics of the data on which it is trained. For this reason, the first essential step concerns the selection of a dataset that is both relevant to the problem at hand and obtained from a reliable source, an aspect that is particularly important when dealing with sensitive clinical information.

Once the dataset has been acquired, it becomes necessary to perform an *Exploratory Data Analysis* (EDA) to gain a comprehensive understanding of its structure and content. This phase includes the examination of variable distributions, the identification of noise, inconsistencies, or missing values, and the assessment of potential correlations that may inform subsequent modelling choices. Through descriptive statistics and signal inspection, EDA provides valuable insight into the overall quality of the data and helps determine which transformations or feature engineering strategies may be required.

A critical component of this process is evaluating the representativeness of the available data with respect to the clinical task. Datasets affected by class imbalance or insufficient variability may hinder the model's ability to generalize to new, unseen cases. Therefore, special attention must be devoted to analysing the distribution of the target classes and ensuring that the selected sample adequately reflects the population of interest, particularly when addressing classification problems.

2.4.2 Preprocessing and Segmentation

Data pre-processing is a fundamental step in the development of any artificial intelligence model, as it transforms raw data into a form suitable for analysis and model training. Sensor measurements, IoT data streams, and clinical repositories frequently contain noise, missing values, inconsistent formats, or redundant information; if left untreated, these issues can hinder the learning process and negatively affect the interpretability of the final results. For this reason, data preparation is often one of the most time-consuming phases of an AI pipeline and involves a structured set of operations aimed at improving data quality and coherence. The pre-processing workflow typically includes several key activities:

- **Data Integration:** combining information from heterogeneous sources—such as clinical databases, wearable devices, or online platforms—while enforcing a consistent schema across all inputs. During this phase, it is crucial to avoid unnecessary duplication and ensure semantic consistency among merged datasets.

- **Data Reduction:** producing a more compact representation of the data without compromising relevant information. Common strategies include dimensionality reduction techniques (e.g. PCA, SVD), data compression, aggregation, or discretization, with the aim of simplifying the dataset and reducing computational overhead.

- **Data Cleaning:** detecting and correcting errors, inconsistencies, duplicated entries, missing values, or anomalous samples. This step is especially critical in medical datasets, which often exhibit irregularities due to acquisition errors, physiological variability, or interruptions in sensor recordings.

- **Data Transformation:** applying operations such as normalization, standardization, smoothing, encoding of categorical variables, or feature engineering. These transformations help create features that are more informative and better aligned with the requirements of the selected learning algorithm.

When the number of available instances is limited—a common scenario in biomedical applications—*data augmentation* techniques may be introduced to artificially expand the dataset. This approach is particularly valuable in DL, as it increases variability in the training set, reduces the risk of overfitting, and improves model generalization to unseen data.

Another frequent issue concerns **class imbalance**, which may bias the model toward the majority class. To address this problem, oversampling techniques such as SMOTE (Synthetic Minority Oversampling Technique) can be employed. SMOTE generates synthetic samples of the minority class by identifying the k nearest neighbours of each selected instance and creating new points through linear interpolation in the feature space.

2.4.3 Model Design and Training

The choice of the modelling approach is fundamentally driven by the nature of the data and by the specific objectives of the task. Once the dataset has been pre-processed and its characteristics sufficiently understood, it becomes possible to determine which category of algorithms is most suitable whether traditional ML techniques or more complex DL architectures, and whether the task should be framed within supervised or unsupervised learning paradigms.

After selecting the appropriate algorithmic family, the next step consists of training the model on a designated training subset of the data. During this phase, the model progressively learns the underlying relationships, structures, and patterns present within the input features. This learning process is iterative and involves the adjustment of internal parameters (such as weights in neural networks or splitting criteria in decision trees) so as to minimize prediction errors with respect to the ground-truth labels.

The training step also requires defining several design choices, such as the loss function, the optimisation algorithm, and the evaluation protocol adopted for monitoring the learning progress. In the context of healthcare data, particularly physiological signals acquired through IoT systems, this phase becomes crucial, as the model must learn to recognise subtle and often noisy patterns that may indicate clinically meaningful conditions.

Ultimately, the goal of the training process is to build a model that generalises well to unseen data, preserving robustness and reliability when applied to new patient recordings or real-world scenarios. To ensure this, the training phase is typically followed by rigorous validation procedures, which guide the selection of hyperparameters and help mitigate risks of overfitting.

Decision Tree-based Algorithms

Decision tree methods constitute a family of supervised learning techniques widely employed in both classification and regression problems. These algorithms operate by recursively partitioning the feature space into increasingly homogeneous subsets, producing a hierarchical tree structure in which each internal node represents a decision rule and each leaf corresponds to a predicted class or numerical output.

A decision tree is built starting from a root node that represents the entire dataset. At each step, the algorithm selects a feature and a corresponding threshold to divide the data into subgroups that are more uniform with respect to the target variable. This subdivision is guided by impurity measures such as Gini index or entropy for classification, and metrics like Mean Squared Error (MSE) for regression. The recursive splitting continues until a stopping criterion is met, producing a set of leaf nodes that ideally contain samples belonging to the same class or with similar numerical values.

One of the distinctive characteristics of decision trees is their high degree of interpretability: each prediction can be traced back to a sequence of simple rules, normally easy to understand even for non-experts. This aspect makes them particularly appealing in domains such as medicine, where transparent decision processes are often required. Additional advantages include their ability to deal with both categorical and continuous variables, their limited need for extensive pre-processing (e.g. no requirement for feature scaling), and their capacity to handle missing values by redirecting samples through alternative branches.

Nevertheless, decision trees also present several drawbacks. They tend to overfit the training data if grown without constraints, and even small perturbations in the dataset can lead to substantial changes in the final structure, reducing model stability. For these reasons, tree-based models are often used within ensemble approaches, which combine multiple trees to produce more robust and accurate predictions. Among the most common algorithms based on decision trees, we recall:

- **Decision Tree Classifier (DTC):** Often associated with early algorithms such as ID3, this model builds a tree by selecting, at each iteration, the feature that maximizes an impurity reduction measure. The classifier proceeds greedily, choosing the best split available at the local level without exploring alternative global configurations. New samples are classified by following the path determined by the decisions encoded in the internal nodes until a leaf is reached.

- **Random Forest Classifier (RFC):** This ensemble method constructs numerous decision trees, each trained on a bootstrap sample of the original dataset. For every node, only a random subset of features is considered when computing the best split. The final prediction is obtained via majority voting (classification) or averaging (regression). The use of multiple diversified trees mitigates overfitting and generally increases predictive performance and generalization.

- **Extra Trees Classifier (EXTC):** Similar in structure to Random Forests, Extra Trees employ even greater randomization. Instead of searching for the optimal splitting threshold, the algorithm chooses random cut points for each candidate feature, which accelerates training time and reduces variance. The increased randomness typically results in more stable and often more generalizable models, although at the cost of reduced interpretability compared to a single decision tree. This method is particularly useful when robustness and execution speed are priorities.

Boosting Algorithms

Boosting algorithms constitute a family of supervised ML methods designed to enhance predictive performance by combining multiple weak learners into a single, more accurate model. Unlike ensemble strategies such as bagging—where base models are trained independently and aggregated in parallel—boosting follows a *sequential* learning scheme. Each learner is added one at a time, and every new model is specifically trained to reduce the errors produced by the previous ones. This iterative refinement allows boosting techniques to focus progressively on the most difficult instances within the dataset.

One of the main strengths of boosting approaches lies in their ability to achieve high predictive accuracy even when individual learners, typically shallow decision trees, have limited capacity. Algorithms such as AdaBoost, Gradient Boosting, and XGBoost are particularly effective at modeling nonlinear relationships and complex interactions between variables, making them suitable for a wide range of tasks, including classification, regression, and ranking problems. In addition, boosting methods generally offer good robustness against overfitting, especially when appropriate regularization mechanisms and early stopping criteria are adopted. However, these advantages come at the cost of increased computational load: boosting models often require longer training times and careful tuning of hyperparameters to ensure stable performance. Interpretability is another challenge, as boosted ensembles are considerably more complex than single decision trees. The principal boosting algorithms considered in this work are summarised below:

- **AdaBoost Classifier (Adaptive Boosting, ABC:** AdaBoost constructs an ensemble by sequentially adding simple learners, typically decision stumps (e.g. trees with depth equal to one). At each iteration, training samples that were incorrectly classified receive a higher weight, forcing the next learner to prioritise these difficult cases. The final prediction is obtained through a weighted combination of all weak learners, where more accurate learners exert greater influence. Although AdaBoost is easy to implement and can substantially improve the performance of weak classifiers, it is sensitive to noisy data and outliers, since misclassified samples progressively gain higher importance in the learning process.

- **Gradient Boosting Classifier (GBC):** Gradient Boosting extends the boosting principle by constructing new learners to directly minimize the residual errors left by the previous ensemble. Instead of reweighting samples as in AdaBoost, each new tree is trained to model the gradient of the loss function, thus iteratively correcting prediction discrepancies. This approach provides high flexibility and the ability to optimize a wide variety of differentiable loss functions. Nevertheless, Gradient Boosting demands careful tuning of parameters such as learning rate, number of trees, and tree depth to prevent overfitting, and it can be computationally expensive on large datasets.

- **Extreme Gradient Boosting Classifier (XGBoost, XGBC):** XGBoost is an optimized and regularized implementation of Gradient Boosting, designed to offer superior computational efficiency and scalability. It incorporates several engineering improvements, including parallelized tree construction, efficient memory usage, and explicit regularization terms to control model complexity. Moreover, XGBoost naturally handles missing values and can manage very large datasets with high speed. Thanks to its strong predictive performance and robustness, XGBoost has become one of the most widely adopted boosting algorithms in ML competitions and real-world applications.

- **Categorical Boost Classifier (CatBoost, CBC):** CatBoost is a boosting algorithm specifically developed to process categorical variables more effectively than traditional methods. Instead of relying on external encoding techniques such as one-hot encoding, CatBoost introduces a target-based encoding strategy that transforms categorical attributes while reducing the risk of target leakage. Its symmetric tree construction and built-in regularization mechanisms contribute to reducing bias and overfitting, enabling the model to perform reliably even on small or noisy datasets. CatBoost is particularly advantageous when working with data that contain many categorical features, offering competitive accuracy with lower preprocessing requirements compared to other boosting methods.

Probabilistic and Distance-Based Algorithms

Probabilistic and distance-based methods represent a class of supervised learning algorithms that rely on statistical principles or geometric reasoning to perform classification or regression. These approaches differ substantially from tree-based or ensemble models, as they make explicit assumptions about data distributions or rely on similarity measures between observations.

- **Naive Bayes Classifier (NBC):** The Naive Bayes classifier is founded on Bayes' theorem and assumes that the features used for prediction are conditionally independent given the target class. This "naive" assumption simplifies computation, making the model extremely lightweight and fast to train. NBC estimates the probability that an input belongs to each class by combining prior probabilities with the likelihood of the observed features, subsequently assigning the sample to the class with the highest posterior probability.

  Despite its simplicity, Naive Bayes is remarkably effective in many scenarios, particularly when dealing with large, noisy, or partially incomplete datasets. It is widely used in text classification, spam detection, and medical filtering tasks. Its primary limitation stems from the independence assumption: when strong correlations exist among features, its predictive performance may decline. Nonetheless, in contexts where dependencies are weak or the dimensionality is high, NBC often provides robust and interpretable results.

- **K-Nearest Neighbors Classifier (KNNC):** The K-Nearest Neighbors algorithm is a non-parametric method used for both classification and regression. Unlike other models, KNNC does not build an explicit predictive function during training; instead, it stores the dataset and bases predictions on the distance between samples. To classify a new observation, the algorithm identifies the K closest points, according to a distance metric such as Euclidean distance, and assigns the most frequent class among those neighbors.

  KNNC is intuitive and easy to implement, leveraging the principle that samples close in feature space tend to share the same label. However, it has several drawbacks: it is computationally expensive during inference, sensitive to outliers and noisy data, and strongly influenced by the choice of K. A very small K may lead to highly unstable predictions, while a large K may obscure local structure in the data. Additionally, performance degrades significantly when dealing with high-dimensional data, due to the well-known "curse of dimensionality."

- **Quadratic Discriminant Analysis (QDA):** Quadratic Discriminant Analysis is a probabilistic classifier based on modeling each class as a multivariate Gaussian distribution. Unlike Linear Discriminant Analysis (LDA), QDA allows each class to have its own covariance matrix, enabling the construction of nonlinear decision boundaries. Using Bayes' rule, QDA computes posterior class probabilities by combining prior information with the likelihood computed from Gaussian densities.

  The flexibility of QDA makes it suitable for complex classification tasks where class boundaries are curved or irregular. However, this flexibility comes at the cost of estimating a greater number of parameters compared to LDA, which increases the risk of overfitting especially when the dataset is small. QDA also assumes that data follow a Gaussian distribution, and violations of this assumption can reduce performance. Despite these constraints, QDA remains a valuable tool in fields such as medical diagnosis, pattern recognition, and financial modeling, where nonlinear separability is often present.

Deep Learning Networks

DL represents a paradigm shift in artificial intelligence. Rather than relying on manually designed features, DL architectures automatically learn hierarchical representations directly from raw data. This is achieved through deep neural networks composed of multiple layers of nonlinear transformations, where each successive layer extracts increasingly abstract features.

- **Convolutional Neural Networks (CNNs):** Convolutional Neural Networks are a class of deep neural networks originally designed for image analysis but now widely applied to diverse data modalities, including onedimensional signals such as speech, sensor readings, and electrocardiograms. Their key innovation lies in the convolutional layer, where learnable filters (also called kernels) are applied across local regions of the input, producing feature maps that capture spatially or temporally localized patterns. Formally, the convolution operation for a one-dimensional input x with kernel w of size k is expressed as:

$$s(t) = (x * w)(t) = \sum_{i=0}^{k-1} w_i \cdot x_{t+i},$$

where *s(t)* represents the activation at position t. During training, both the kernel weights w and bias terms are learned via backpropagation, allowing the filters to adapt to taskspecific features. CNNs are organized in hierarchical blocks, each performing a sequence of transformations:

1. ***Convolutional layers:*** apply multiple filters to detect local patterns such as edges, peaks, or oscillations. In the case of biomedical signals, these filters often capture recurring motifs like QRS complexes or periodic oscillatory patterns.
2. ***Activation functions:*** nonlinear mappings, most commonly the Rectified Linear Unit (ReLU), are applied to the convolution outputs, introducing nonlinearity and improving representational capacity.

3. ***Pooling layers***: reduce the dimensionality of feature maps by downsampling, typically via max pooling or average pooling. This introduces invariance to small translations and reduces computational cost, while preserving the most salient features.

4. ***Normalization layers:*** such as batch normalization, which stabilize training by normalizing intermediate activations, improving both convergence speed and generalization.

5. ***Fully connected layers:*** at the final stage, feature maps are flattened and connected to dense layers that integrate local features into global representations. The output layer typically uses a sigmoid (for binary classification) or softmax (for multi-class problems) activation.

One of the strengths of CNNs is their ability to build hierarchical feature representations: lower layers capture low-level features (short-term patterns, sharp transitions), while deeper layers combine them into high-level abstractions (complex shapes, rhythm irregularities). This property is particularly advantageous for structured data where both local details and global structures are relevant.

Another important aspect is **weight sharing**. Unlike fully connected networks, where each parameter is unique to a connection, CNN filters are applied across the entire input domain. This drastically reduces the number of parameters, improving efficiency and mitigating overfitting, especially when training data is limited.

In addition, CNNs benefit from **translation invariance**: a pattern recognized in one part of the signal can also be detected elsewhere with the same filter. For one-dimensional biomedical signals such as ECG, this property is crucial because pathological patterns (e.g. fibrillatory waves) may appear at arbitrary positions in the input segment. Over time, several architectural refinements have been introduced, including **Deeper CNNs** with many stacked layers (e.g. VGG, ResNet) to capture complex hierarchical features, **Dilated convolutions**, which expand the receptive field without increasing the number of parameters, useful for modeling long temporal dependencies and **1D CNNs**, specifically tailored to sequential data, which process signals along the temporal axis instead of spatial grids.

Thanks to these properties, CNNs have become a cornerstone in DL for pattern recognition, offering a powerful compromise between expressiveness, computational efficiency, and scalability. Their ability to directly learn task-specific filters makes them particularly effective in domains where manual feature extraction is challenging or suboptimal.

- **Recurrent Neural Networks (RNNs):** Recurrent Neural Networks (RNNs) extend traditional feedforward architectures by introducing feedback connections that allow information to persist across time steps. Given a sequence of inputs $\{x_1, x_2, \ldots, x_T\}$, an RNN updates its hidden state h_t at each time step according to:

  $$h_t = \sigma(W_h h_{t-1} + W_x x_t + b),$$

  where W_h and W_x are weight matrices, b is a bias term, and is a nonlinear activation (e.g. tanh). This recurrent formulation makes RNNs suitable for tasks where order and temporal context matter, such as speech recognition, natural language processing, and physiological time series. At each step, the hidden state acts as a compressed memory of all previous inputs.

  However, training RNNs on long sequences is challenging due to the problem of **vanishing and exploding gradients**. As the error signal is backpropagated through many time steps (Backpropagation Through Time, BPTT), gradients may diminish to near zero or grow uncontrollably, preventing the effective learning of long-term dependencies.

  Consequently, vanilla RNNs tend to capture only short-term patterns, limiting their performance in complex sequential modeling tasks. Despite these limitations, RNNs introduced the key idea of modeling temporal dynamics within neural networks, paving the way for more advanced recurrent architectures.

- **Long Short-Term Memory (LSTM):** Long Short-Term Memory (LSTM) networks were developed to overcome the vanishing gradient problem inherent in RNNs. An LSTM introduces a memory cell that explicitly maintains information over time, together with gating mechanisms that regulate the flow of information.

$$
\begin{aligned}
f_t &= \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\
\tilde{c}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \\
h_t &= o_t \odot \tanh(c_t).
\end{aligned}
$$

  Through this design the **forget gate** discards irrelevant information, the **input gate** adds new information to the memory cell, the **output gate** controls what is exposed to the next hidden state.

  LSTMs thus maintain a stable memory over long sequences, enabling the modeling of complex temporal dependencies. They have been widely adopted in language modeling, speech recognition, machine translation, and biomedical time series forecasting. While computationally more demanding than vanilla RNNs, their ability to capture longterm context makes them a standard baseline for sequence learning.

- **Bidirectional Long Short-Term Memory (BiLSTM):** Bidirectional LSTMs (BiLSTMs) extend LSTMs by processing input sequences in both forward and backward directions. This design allows the network to exploit not only past information but also future context, producing richer sequence representations. Formally, a BiLSTM computes two hidden states for each time step:

$$
\vec{h}_t = \text{LSTM}_f(x_t, \vec{h}_{t-1}), \qquad h_t = \text{LSTM}_b(x_t, h_{t+1}),
$$

and concatenates them into:

$$h_t = [\vec{h}_t \,;\, h_t].$$

The main benefits of BiLSTMs include:

- **Comprehensive context modeling:** by integrating past and future states, BiLSTMs better capture dependencies spanning both directions.
- **Noise robustness:** bidirectional context smooths over irregularities in sequential data.
- **Superior performance in classification:** many sequential tasks (speech recognition, named entity recognition, ECG analysis) benefit from having the entire sequence before decision-making.

The main drawback of BiLSTMs is their increased computational cost and memory footprint, as they require approximately twice the parameters of standard LSTMs. Nonetheless, in many applications, the performance gains justify the added complexity.

• **Temporal Convolutional Networks (TCNs):** Temporal Convolutional Networks (TCNs) provide an alternative to recurrent architectures for sequence modeling. Instead of sequential processing, TCNs employ **causal onedimensional convolutions**, ensuring that predictions at time t depend only on present and past inputs. By stacking layers of dilated convolutions, the receptive field grows exponentially with depth, allowing TCNs to capture longterm dependencies efficiently. Key components of TCNs include:

  - **Causal convolutions**, which preserve the temporal ordering of input data.
  - **Dilated convolutions**, which insert gaps between kernel elements to expand the receptive field without increasing parameters.
  - **Residual connections**, often employed to stabilize training in deep TCNs and facilitate gradient flow.

  Compared to RNNs and LSTMs, TCNs offer several advantages:

- Parallel computation across time steps, improving training efficiency.
- More stable gradients, avoiding vanishing/exploding issues.
- Competitive or superior accuracy in tasks such as speech synthesis, machine translation, and time series forecasting.

A limitation of TCNs is their higher memory usage for very long sequences, as intermediate convolutions must be stored. Nonetheless, they have emerged as a strong alternative to recurrent models, combining the benefits of convolutional architectures with the ability to model longrange dependencies.

2.4.4 Hyperparameter Optimisation

Optimizing the hyperparameters of a model represents a fundamental step in the development of any ML pipeline, as these parameters govern the model's behavior, capacity, and ability to generalize beyond the training data. Hyperparameters influence aspects such as model complexity, learning dynamics, regularization strength, and the structure of the predictive function itself. An inappropriate hyperparameter configuration may lead to phenomena such as *overfitting*, where the model adapts excessively to training data and fails to generalize, or *underfitting*, in which the model is too simple to capture the intrinsic structure of the data.

For these reasons, the optimization phase aims to systematically explore alternative configurations to identify those that maximize performance metrics while maintaining a balance between expressiveness and generalization. This process often involves the evaluation of multiple parameter combinations, enabling the construction of a more robust model than one relying on default or arbitrarily chosen values. Additionally, hyperparameter tuning can play a key role in improving computational efficiency: an optimally configured model may require fewer training iterations, converge faster, and achieve better predictive stability.

Among the most widely adopted techniques, *Grid Search* stands out as a systematic and exhaustive approach: it explores all possible combinations within a predefined parameter grid to identify the configuration that maximizes a selected performance metric. Although computationally intensive, Grid Search offers an effective and reproducible mechanism for identifying optimal hyperparameters, especially when the search space is limited or well-defined.

In the context of this thesis, two different strategies were adopted depending on the type of model. For traditional ML classifiers, the default hyperparameter configurations provided by the respective libraries were employed. This choice reflects the objective of establishing a solid and unbiased baseline across heterogeneous algorithmic families such as tree-based, linear, probabilistic, kernel-based, and shallow neural models while avoiding extensive tuning procedures that could lead to overfitting specific characteristics of the dataset.

Conversely, for DL architectures, a dedicated hyperparameter optimization process was carried out to ensure stable convergence, prevent divergence during training, and improve the model's generalization capability. The DL training pipeline relies on the binary cross-entropy loss function optimized with the Adam optimizer. Furthermore, training stability and generalization are enhanced through the use of callbacks such as *Early Stopping* and *Model Checkpointing*, while Dropout layers and Max Pooling contribute additional regularization.

Overall, the methodology integrates two complementary paradigms: on one hand, feature-based ML models offer interpretability, computational efficiency, and operational stability, qualities that are essential for clinical applicability and edge deployment. On the other hand, DL architectures leverage hierarchical representation learning to capture complex spatio-temporal patterns directly from raw data, achieving state-of-the-art predictive performance. The synergy of these two approaches provides a comprehensive perspective on the potential and limitations of AI techniques applied to IoT-derived health data.

2.4.5 Model Evaluation

Validating a ML model is a fundamental step to ensure that its predictions are reliable and applicable beyond the data used for training. Building a model is not sufficient: its performance must be evaluated rigorously to verify that it generalizes well and that the results can be trusted in real clinical or operational contexts. Only in very rare cases, typically when extremely large and representative datasets are available, can validation be considered less critical. In most practical scenarios, however, the available data only partially reflect the true population, making systematic validation indispensable.

The choice of an appropriate validation strategy strongly influences the robustness and credibility of the resulting model. Different techniques exist, each with advantages and limitations, and selecting the method that best fits the characteristics of the data is crucial for developing transparent, unbiased, and dependable AI systems.

Once preprocessing has been completed, the dataset is typically divided into two main subsets: a *training set*, used to fit the model, and a *test set*, used exclusively to assess its final performance. In many cases, an additional *validation set* is also created to fine-tune hyperparameters and guide intermediate evaluations during model development.

One of the simplest and most widely used strategies for this division is the **Holdout** method, which splits the dataset into two or three portions for example, two-thirds for training and one-third for testing. Although intuitive, this approach suffers from a key limitation: the resulting subsets might not be fully representative. In classification tasks, for instance, minority classes may be underrepresented or even completely absent in one split, leading to biased training and unreliable test results. A refined version, known as **stratified holdout**, preserves the original proportion of classes across splits, partially mitigating this issue; however, when the dataset is small, underrepresentation problems may persist.

A more robust alternative is **K-Fold Cross-Validation**. In this method, the dataset is partitioned into k equally sized folds. At each iteration, one fold acts as the validation set, while the remaining $k-1$ folds are used for training. This process is repeated k times so that every sample appears once in the validation set and $k-1$ times in the training set. The final performance is then computed as the average across all iterations, yielding a more stable and reliable estimate and reducing the impact of random sampling variability typically associated with a single train/test split.

Despite its advantages, K-Fold Cross-Validation requires higher computational effort, as the training process must be repeated k times. Nevertheless, it remains one of the most effective and widely adopted validation techniques when the goal is to obtain a trustworthy and generalizable evaluation of model performance.

The confusion matrix, therefore, describes the complete performance of the model assessed using accuracy, precision, recall, F1-score, and the confusion matrix, which allows quantifying both correctly and incorrectly classified segments. The confusion matrix distinguishes true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), thus enabling the computation of the following measures.

Accuracy (A) measures the proportion of correctly classified instances over the total number of predictions. Defining true positives (TP) and true negatives (TN) as the correctly classified instances and false positives (FP) and false negatives (FN) as the incorrectly classified instances, accuracy is defined as follows:

$$A = \frac{TP + TN}{TP + FP + TN + FN}.$$

Precision (P) quantifies the fraction of true positives among all instances predicted as positive. More specifically, Precision is the ability of a classifier not to label a positive instance that is actually negative; it is, therefore, the percentage of correct positive predictions (TP) on the total positive predictions of the model (correct TP or wrong FP). It is defined as follows:

$$P = \frac{TP}{TP + FP}.$$

Recall (R) referred to as sensitivity, evaluates the proportion of correctly identified positive cases over the total number of actual positive cases.

$$R = \frac{TP}{TP + FN}.$$

F1-Score (F1) computed as the harmonic mean of precision and recall, which gives more weight to small values. This causes a classifier to achieve a high F1-Score only when both Precision and Recall are high.

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}.$$

In addition to these metrics, for DL models we also monitored the **loss function**, which quantifies the discrepancy between the predicted and the true labels. In this study, binary cross-entropy was adopted as the probabilistic loss function, optimized with Adam. Learning curves of accuracy and loss across epochs were analyzed to assess generalization performance and detect potential signs of overfitting or underfitting.

2.4.6 Explainability

Explainability represents a crucial aspect in the development of AI systems for healthcare, as it aims to provide transparency regarding how a model produces its predictions. In clinical contexts, the ability to justify and interpret decisions is fundamental for fostering trust among medical professionals and enabling the safe integration of AI solutions into real-world workflows. Unlike classical ML classifiers, whose decision boundaries can often be inspected through feature importance analyses or rule-based structures, DL architectures inherently operate as black-box models, making their internal reasoning less transparent.

For this reason, explainability was considered as an important methodological dimension of the proposed framework, particularly with regard to the potential clinical deployment of the CNN+BiLSTM architecture developed in the third phase of this work. In particular, saliency-based approaches and the analysis of the contribution of convolutional filters and recurrent temporal dynamics were identified as relevant strategies to investigate which parts of the ECG signal most influence the model's predictions.

Although explicit explainability outputs are not included among the experimental results reported in this thesis, preliminary qualitative inspections of activation patterns suggested that the model focuses on physiologically relevant ECG regions. The integration of interpretability techniques remains a key direction for future development. Such analyses would be valuable to: (i) assess the clinical plausibility of the learned representations; (ii) identify possible model biases or artefact-driven behaviours; and (iii) support a more transparent comparison between feature-based ML classifiers and end-to-end DL solutions.

In this perspective, explainability is not treated as a secondary add-on, but as an essential component for strengthening the reliability, transparency, and adoption potential of AI models in Smart Health applications.

CHAPTER

3 Approach and Methodologies

The research presented in this thesis was developed through a structured, multi phase methodological framework designed to progressively explore, analyse, and model the integration of Artificial Intelligence techniques within IoT based Smart Health systems. Rather than relying on a single analytical perspective, the work follows a sequential approach in which each phase builds upon the findings of the previous one, ensuring both coherence and methodological rigour.

The first phase focused on a systematic exploration of the existing scientific literature, aimed at understanding the state of the art in IoT and AI integration for Smart Health. This stage provided the conceptual foundations of the thesis, identifying the principal technological approaches, recurring challenges, open limitations, and privacy or security issues associated with the processing of IoT generated health data. The outcomes of this analysis enabled the formulation of the first set of research questions and guided the definition of the experimental directions pursued in the subsequent phases.

The second phase shifted to an empirical perspective by selecting a real biomedical dataset and conducting a comprehensive evaluation of traditional ML techniques. Using the IRIDIA–AF long term ECG dataset as a testbed, this phase assessed the ability of classical ML algorithms to extract clinically meaningful information from IoT acquired physiological signals and to support diagnostic tasks such as atrial fibrillation detection. This investigation provided valuable insights into the strengths and limitations of feature based ML approaches when applied to real, noisy, and irregular IoT health data.

Building on these results, the third phase developed and evaluated a DL model capable of operating directly on raw ECG segments. This stage aimed to determine whether end–to–end neural architectures, such as CNN–BiLSTM hybrids, could surpass the performance of ML classifiers and offer improved diagnostic accuracy, predictive capabilities, and generalisation. By comparing the results of DL models with those obtained in the previous phase, this final step provided an evidence based answer to the research questions regarding the potential of DL techniques in Smart Health.

Together, these three phases constitute an integrated methodological pipeline that progressively moves from theoretical understanding, to empirical ML evaluation, and finally to advanced DL modelling. This structure ensures that the final conclusions are grounded in both literature evidence and rigorous experimental analysis across different AI paradigms.

3.1 Preliminary Study and Application of AI Techniques

The methodological approach adopted to conduct this research [10] was inspired by the principles formalised by Kitchenham (2004) [56] for evidence-based software engineering. In line with these guidelines, the study followed a structured and replicable process aimed at identifying, selecting, and analysing the most relevant scientific contributions related to the integration of Internet of Things (IoT) technologies and Artificial Intelligence (AI) within Smart Health systems.

The objective of the methodology was to gather consolidated knowledge regarding both (i) the technological approaches, challenges, and limitations surrounding IoT and AI integration in healthcare (RQ_1), and (ii) the privacy and security issues arising from the collection and processing of health data through IoT devices (RQ_2). To this end, the following sequential phases were defined:

- **Formulation of research questions**: identification of the two guiding questions focusing on technological advances and security concerns in IoT-based Smart Health.

- **Definition of search keywords and query construction**: extraction of the most representative terms associated with IoT, AI, Smart Health, privacy, and security, then combined into a structured query suitable for scientific search engines.

- **Selection of electronic databases**: identification of the most authoritative sources in the field (e.g. IEEE Xplore, PubMed, ACM Digital Library, Scopus), ensuring adequate coverage of both technological and healthcare-oriented literature.

- **Application of initial filtering criteria**: definition of temporal boundaries, language constraints, and relevance requirements to eliminate clearly unsuitable publications.

- **Screening of titles and abstracts**: exclusion of duplicated or contextually irrelevant studies through a first-level assessment.

- **Full-text eligibility assessment**: evaluation of the remaining documents using more specific inclusion criteria related to IoT architectures, AI methodologies, data management, and security/privacy considerations.

- **Structured analysis of selected studies**: extraction and synthesis of the information necessary to answer RQ_1 and RQ_2, focusing on technological solutions, methodological approaches, limitations, and identified risks in the handling of health data.

To select the articles on which our research work focused, the following two main databases were used:

- IEEEXplore: a digital library and research database for discovering and accessing journal articles, conference proceedings, technical standards, and related materials in computer science, electrical engineering, electronics, and related fields;

- ScienceDirect: which provides access to a large bibliographic database of scientific and medical publications from the Dutch publisher Elsevier.

Following these phases, the research questions were translated into a specific query used across the selected databases:

Q: (*Wearable*) **AND** (*Artificial Intelligence*) **AND** (*Smart Health*)

This systematic approach ensured methodological rigour and provided a solid foundation for analysing the technological landscape and the associated privacy and security issues emerging in IoT-driven Smart Health ecosystems.

Figure 3.1 presents the temporal distribution of the works definitively considered. This shows the years on the x-axis and the number of contributions on the y-axis, showing the total number of contributions with the blue bars, and the percentage with the orange bars. As can be seen from the figure, the number of articles addressing the topic of our research has become significant since 2017 and has increased over the years.

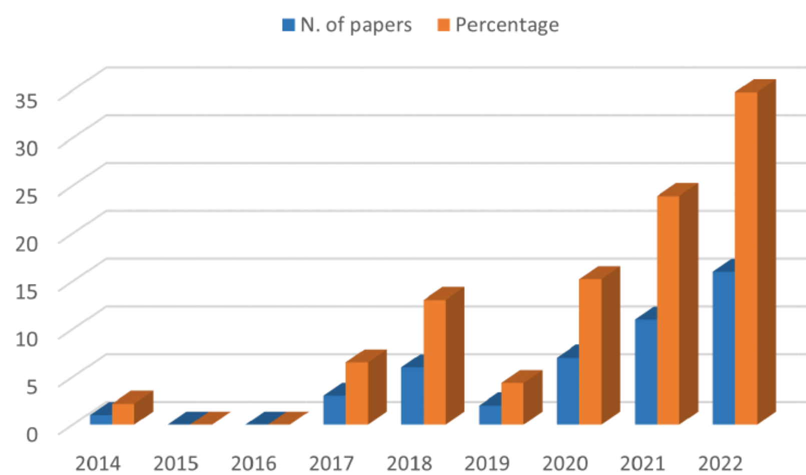

Figure 3.1: Yearly distribution of the analyzed papers.

3.2 Analysis of IoT Health Data Using Machine Learning Techniques

The analytical pipeline developed in this study [8] was designed to investigate the effectiveness of ML techniques in identifying clinically meaningful patterns within IoT-acquired physiological signals. The objective was twofold: (i) to assess the potential of ML methods in extracting and analysing clinically relevant information from IoT-generated health data (RQ_3), and (ii) to evaluate the extent to which these models can enhance diagnostic accuracy, predictive capabilities, and decision-support mechanisms within Smart Health systems (RQ_4).

To address these questions, a structured workflow was applied to the IRIDIA-AF dataset, which contains long-term ECG recordings with annotated paroxysmal atrial fibrillation (PAF) episodes. The methodological steps were organized into a modular and lightweight pipeline, summarized as follows:

1. **Data Acquisition and Preparation.** Raw ECG traces and annotation files were imported from the IRIDIA-AF database. Metadata such as sampling frequency, recording duration, and the temporal boundaries of annotated AF episodes were parsed and standardized. This ensured consistent preprocessing across all recordings and enabled a uniform experimental setting.

2. **Coalescence-Based Segmentation.** To prevent the fragmentation of AF annotations caused by short gaps or noise, a coalescence window strategy was introduced. Similar temporal merging approaches are commonly adopted in ECG analysis to smooth fragmented annotations and obtain clinically meaningful arrhythmia intervals [91]. In this study, a temporal threshold was defined to merge temporally adjacent AF annotations separated by short gaps. The optimal value of this threshold was determined through a sensitivity analysis employing the *KneeLocator* algorithm, which identifies the point of maximum curvature in the evaluation curve and thus provides a principled trade-off between excessive merging and over-fragmentation of AF episodes. The resulting consolidated intervals define segmentation windows used to extract the corresponding raw ECG portions for downstream analysis.

3. **Segment Labeling.** Each extracted segment was labeled as either *AF* or *NSR* (Normal Sinus Rhythm) based on its temporal overlap with manually annotated arrhythmic events. All segments were exported in human readable `.txt` format, ensuring reproducibility and straightforward traceability of each sample.

4. **Feature Extraction.** For every labeled ECG segment, two distinct feature sets were

computed to evaluate the discriminative potential of traditional ML compatible descriptors:

- **Statistical time-domain features:** mean, standard deviation, variance, skewness, kurtosis, and related summary statistics capturing morphological changes induced by arrhythmic patterns.
- **Frequency-domain features:** spectral descriptors derived using both the Fourier Transform and the Discrete Wavelet Transform (DWT), enabling the characterization of oscillatory and transient behaviors typical of PAF.

These features operationalize the ECG signal into a structured numerical form suitable for classical ML models, thereby enabling the evaluation required for RQ3.

5. **Model Training and Evaluation.** The extracted features were used to train a Random Forest classifier, chosen for its robustness, interpretability, and suitability for medium-scale datasets. Model performance was assessed in terms of accuracy, sensitivity, specificity, and F1-score to determine its capacity for distinguishing AF from NSR segments. This step directly addresses RQ_4 by quantifying the diagnostic and predictive benefits achievable through ML based analysis of IoT acquired signals.

Overall, this methodological framework enables a rigorous empirical assessment of the capabilities and limitations of ML in Smart Health environments. By combining principled segmentation, handcrafted feature extraction, and interpretable classification models, the pipeline provides meaningful insights into how ML can support automated AF detection, clinical decision making, and large scale IoT healthcare monitoring.

3.2.1 Dataset

IRIDIA-AF is a large publicly available long-term electrocardiogram (ECG) monitoring dataset developed for the study of paroxysmal atrial fibrillation (PAF) [42][1]. The dataset signals included time-series data acquired through wearable devices, reflecting cardiac electrical

[1]Available at: https://zenodo.org/records/8405941

activity and other relevant vital parameters. Such data provided an ideal testbed for evaluating the applicability of ML methods in early detection and classification tasks related to PAF episodes. The dataset, thoroughly described in a *Nature Scientific Data* publication, was developed to support the automated detection and classification of AF episodes based on expert annotated ground truth.

The dataset includes 167 recordings from 152 patients, some of whom underwent multiple Holter recording sessions on different dates. These signals, often collected over 24 to 72 hours through Holter monitors or similar devices, provide a realistic view of paroxysmal AF dynamics, which may not be captured during short term monitoring. Metadata describing recording sessions, sampling rate, and patient information are stored alongside the signals to enable downstream segmentation and annotation. Each record contains ECG data sampled at 200 Hz, split into 24-hour files, with AF episodes manually annotated by clinical experts. The dataset is organised into several components:

1. **A metadata file** describing the 167 recordings, each associated with a patient. Since some patients underwent multiple acquisition sessions, the number of recordings exceeds the number of patients. The file includes detailed information for each record: acquisition date and time, calculated duration, number of files, number of samples, and effective recording time in seconds. The effective recording time may differ from the calculated duration due to interruptions or overlaps in the data.

2. **An ECG annotation CSV file,** which provides the temporal details of each AF episode and periods of normal sinus rhythm (NSR), including start and end timestamps, file indices, QRS complex indices, episode duration, and duration of NSR before onset.

3. **An RR interval annotation CSV file,** which contains the start and end indices of RR cycles (the intervals between consecutive heartbeats) and the corresponding file indices.

4. **A folder of HDF5 files** for each recording, containing:

    - The raw ECG signals are stored as a two-dimensional array $[samples \times 2]$, where each row represents one sample across two ECG leads (e.g. two electrode positions);

- The RR interval data, expressed in milliseconds, is useful for heart rate variability analysis.

The recordings vary in length, ranging from a minimum duration of 71,635 seconds ($\sim$ 20 hours) to a maximum of 345,599 seconds ($\sim$ 95 hours). Accordingly, each record is composed of a variable number of HDF5 files, depending on its total duration. The dataset provides annotated long term Holter ECG recordings, where AF events have been manually labelled by expert cardiologists. These annotations were used as ground truth to segment the recordings and extract labelled AF and NSR episodes. The IRIDIA-AF dataset has been widely used in studies focusing on atrial fibrillation detection and is particularly suitable for evaluating automated analysis techniques on long-term ECG recordings.

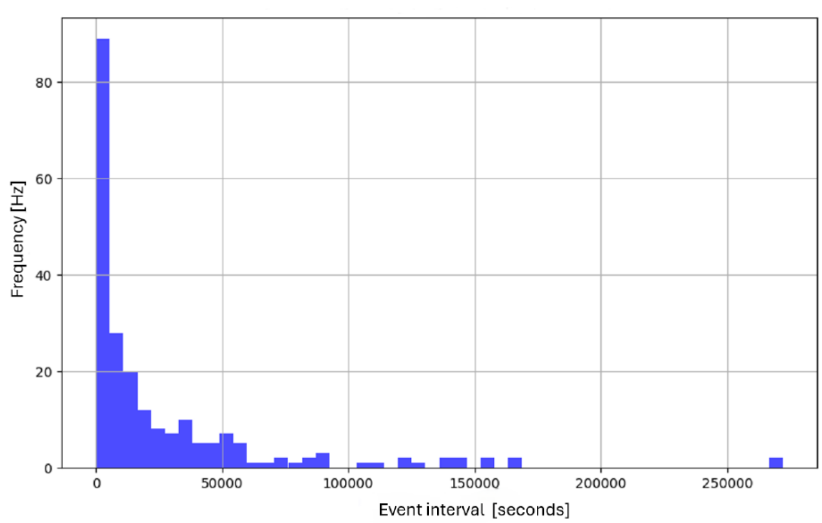

Figure 3.2: Distribution of inter-event intervals between AF episodes across all recordings. A high concentration of short intervals is observed, suggesting frequent AF recurrences within short time spans.

The target classification task considered in this study is binary and consists of discriminating between atrial fibrillation (AF) and normal sinus rhythm (NSR) segments. For the purpose of model training, the class labels were encoded numerically as AF = 1 and NSR = 0.

Since the original dataset may exhibit class imbalance, particular attention was devoted to balancing the data before training. In the feature-based ML pipeline, class imbalance was addressed through random undersampling of the majority class.[2] In the raw-signal DL pipeline, a balanced subset was constructed by matching the number of NSR segments to the number of AF segments after filtering by signal length.[3] Therefore, the experimental datasets used in this work were balanced through undersampling strategies, in order to reduce classification bias toward the majority class and ensure a fairer comparison between models.

3.2.2 Preprocessing

Preprocessing represents a fundamental step of the proposed pipeline, ensuring that atrial fibrillation (AF) episodes are consistently detected and that both feature based and DL approaches operate on standardised input data. The preprocessing workflow is articulated into several stages, as detailed below. The preprocessing pipeline was implemented in Python.[4]

Coalescence Window Analysis

Paroxysmal atrial fibrillation (AF) episodes generally occur in temporal clusters rather than as isolated events. To properly capture this behaviour, a *coalescence window* (CW) was introduced, defined as a temporal threshold used to merge consecutive AF annotations into coherent and clinically meaningful groups. Two episodes are assigned to the same cluster whenever the time interval between the end of the first and the onset of the second is shorter than the selected CW. This procedure mitigates excessive fragmentation caused by brief interruptions, annotation artefacts, or transient noise within the ECG signal. To determine an appropriate CW value, the distribution of inter-event intervals was examined and a sensitivity analysis was carried out by progressively increasing the size of the window. This process yields a monotonically decreasing curve that relates the CW value to the number of resulting AF clusters. The *KneeLocator* algorithm was then employed to identify the elbow point of this curve, which reflects the best compromise between two problematic scenarios:

[2]The ML preprocessing and classification pipeline was implemented in Python using NumPy, pandas, scikit-learn, imbalanced-learn, XGBoost, CatBoost, and joblib.

[3]The DL models were implemented in Python using TensorFlow/Keras. Additional libraries included keras-tcn, kneed, statsmodels, and Matplotlib for auxiliary analyses and visualizations.

[4]The preprocessing and feature engineering steps were implemented in Python using NumPy, pandas, statsmodels, kneed, and Matplotlib, together with custom utility functions developed for signal segmentation, labeling, and feature extraction.

- **Event collisions:** distinct AF occurrences may be incorrectly merged when using an excessively large CW;

- **Segment truncation:** short or incomplete segments may be produced when the CW is too small, particularly for episodes located near the boundaries of the recording.

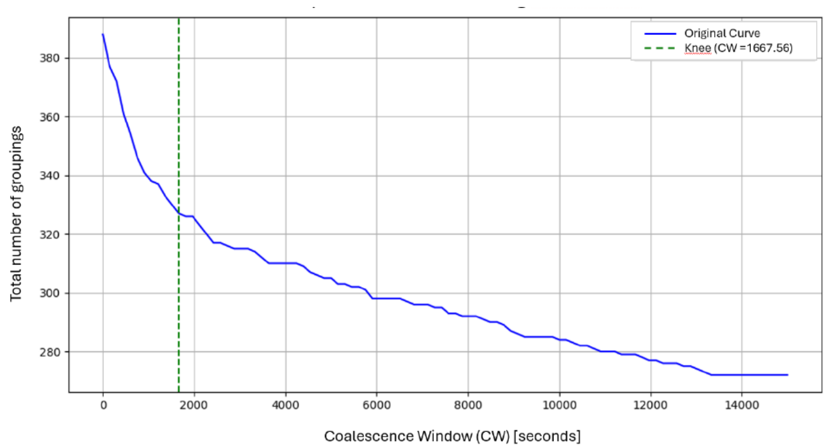

Figure 3.3: Determination of the optimal coalescence window using the KneeLocator algorithm. The curve shows the number of AF episode clusters obtained as a function of the coalescence window size. As the window increases, temporally adjacent episodes are progressively merged, reducing the number of distinct groups. The green dashed line marks the elbow point *(CW 1667.56 seconds)*, which represents the optimal trade-off between over-fragmentation and over-grouping.

A preliminary enrichment of the dataset included assigning AF/NSR labels at the segment level, obtained through the segmentation process driven by the CW analysis. The distribution of inter-event times Figure 3.2 showed a high concentration of short gaps between AF episodes, indicating the necessity of temporal grouping. For this reason, the number of grouped episodes was evaluated across different CW values, generating the curve shown in Figure 3.3. The application of the KneeLocator allowed the identification of the elbow point, representing the CW that provides an optimal trade-off between over-grouping and excessive fragmentation.

This approach ensures that the segmentation strategy not only remains computationally consistent but also reflects the underlying clinical behaviour of paroxysmal AF episodes, thus enhancing the representativeness of the data used for model training.

Segmentation

Since raw long duration ECG recordings cannot be directly employed for classification tasks, the signals were partitioned into fixed-length segments of $L = 166{,}755$ samples. This length corresponds to the average duration obtained after applying the optimal coalescence window, estimated at approximately 1667.56 seconds (~ 27.8 minutes). Each resulting segment was assigned a label, either AF or NSR based on the expert annotations provided with the IRIDIA-AF dataset. Through this segmentation strategy, continuous ECG recordings were transformed into a collection of structured, labelled samples suitable for both traditional ML and DL workflows. Using fixed size windows also ensures consistency across samples and enables efficient batch processing during model training.

To guarantee compatibility with subsequent models, all extracted segments were normalized to the same temporal length. Segments shorter than the target duration were symmetrically zero padded to preserve temporal alignment, whereas segments exceeding the limit were truncated at the nearest valid boundary. Each processed segment was then exported as a plain text (`.txt`) file, a format chosen to maintain transparency, allow manual inspection, and simplify experimental reproducibility. The outcome of this procedure is a well organized and labelled dataset of ECG segments forming the basis for both the feature-engineered ML models and the raw signal DL architectures explored in this thesis.

Feature Extraction

In the first branch of the pipeline, each ECG segment is processed through a set of handcrafted feature extraction procedures designed to capture the temporal and spectral properties of the cardiac signal. Feature computation is carried out across multiple domains to obtain a comprehensive numerical description of the underlying rhythm.

- **Time-domain features**: statistical descriptors such as mean, standard deviation, variance, skewness, and kurtosis, together with complexity-related metrics including entropy measures and Hjorth parameters (activity, mobility, and complexity). These quantities summarize the amplitude variability and structural organization of the signal.

- **Frequency-domain features**: obtained through the Fast Fourier Transform (FFT), which enables the extraction of spectral entropy, dominant frequency components, and energy distributions across predefined frequency bands. These characteristics help distinguish AF patterns from normal sinus rhythm by analyzing oscillatory behaviour.
- **Wavelet-domain features**: derived using multi-resolution wavelet decomposition, allowing the detection of transient and nonstationary patterns. This representation is particularly suitable for ECG signals, whose morphology can vary rapidly during arrhythmic events.

Prior to model training, all extracted features undergo z-score normalization to ensure consistent scaling across dimensions. To mitigate the class imbalance inherent in the dataset, random undersampling was applied to reduce the dominance of majority class instances. Segments exhibiting more than 10% missing samples were excluded from the feature extraction process, while short gaps were reconstructed using linear interpolation to preserve signal continuity.

The resulting feature matrix provides a stable and compact numerical representation of the ECG dynamics, well suited for classical ML classifiers. This feature based branch prioritizes interpretability and computational efficiency, making the resulting models suitable for deployment in clinical environments or edge devices with limited processing capabilities.

3.2.3 Machine Learning Pipeline

To assess the discriminative capability of the features extracted from each ECG segment, a broad set of supervised ML algorithms was explored. The selected models cover a wide spectrum of methodological families, allowing a comprehensive evaluation of how different learning paradigms perform on the atrial fibrillation classification task.

- **Tree based approaches**: This group includes Decision Tree, Random Forest, Extra Trees, Gradient Boosting, AdaBoost, XGBoost, and CatBoost. These algorithms are particularly effective for tabular data, tend to be resilient to feature scaling issues, and often exhibit strong performance even in the presence of noise or missing values.

Their interpretability and reliability make them widely adopted in biomedical research. Decision tree methods constitute a family of supervised learning techniques widely employed in both classification and regression problems [17, 84]. Random Forest further extends this paradigm by aggregating multiple decision trees trained on bootstrap samples [16], while ensemble boosting approaches such as Gradient Boosting and AdaBoost iteratively combine weak learners to improve predictive accuracy [39, 40]. More recent implementations such as XGBoost and CatBoost introduce optimized and regularized gradient boosting frameworks that improve scalability and predictive performance on structured datasets [22, 82].

- **Linear and probabilistic models**: Logistic Regression and Gaussian Naive Bayes were selected to establish baseline performance using simple, easily interpretable classifiers. Logistic Regression assumes a linear decision boundary and estimates class probabilities through a logistic function [49], while Gaussian Naive Bayes relies on the assumption of normally distributed features and conditional independence among them [32].

- **Distance based method**: The K-Nearest Neighbors (KNN) classifier assigns a label by examining the closest samples in the feature space. Although conceptually simple and non parametric, its performance strongly depends on the choice of distance metric and proper feature normalization. The method is rooted in early work on nearest neighbor classification [30].

- **Kernel based model**: The Support Vector Classifier (SVC) was employed to evaluate the ability of kernel-based techniques to separate classes through nonlinear decision boundaries defined in high-dimensional spaces. Support Vector Machines maximize the margin between classes and can implicitly map inputs into higher dimensional feature spaces using kernel functions [29].

- **Neural model**: A Multi-Layer Perceptron (MLPClassifier) was integrated as a lightweight neural baseline, enabling the assessment of simple feed-forward architectures for learning nonlinear relationships among the extracted features. MLP networks represent one of the earliest forms of artificial neural networks and are trained through backpropagation to learn nonlinear decision boundaries [92].

All classifiers were trained on the same collection of labeled segments and feature vectors to ensure methodological consistency and a fair comparison across models.

3.3 Model Design and Implementation

This section describes the proposed methodology to build the AI model for atrial fibrillation (AF) detection, that integrates traditional ML with handcrafted features and modern DL architectures operating directly on raw signals [8]. Figure 3.4 provides a high-level overview of the experimental pipeline.

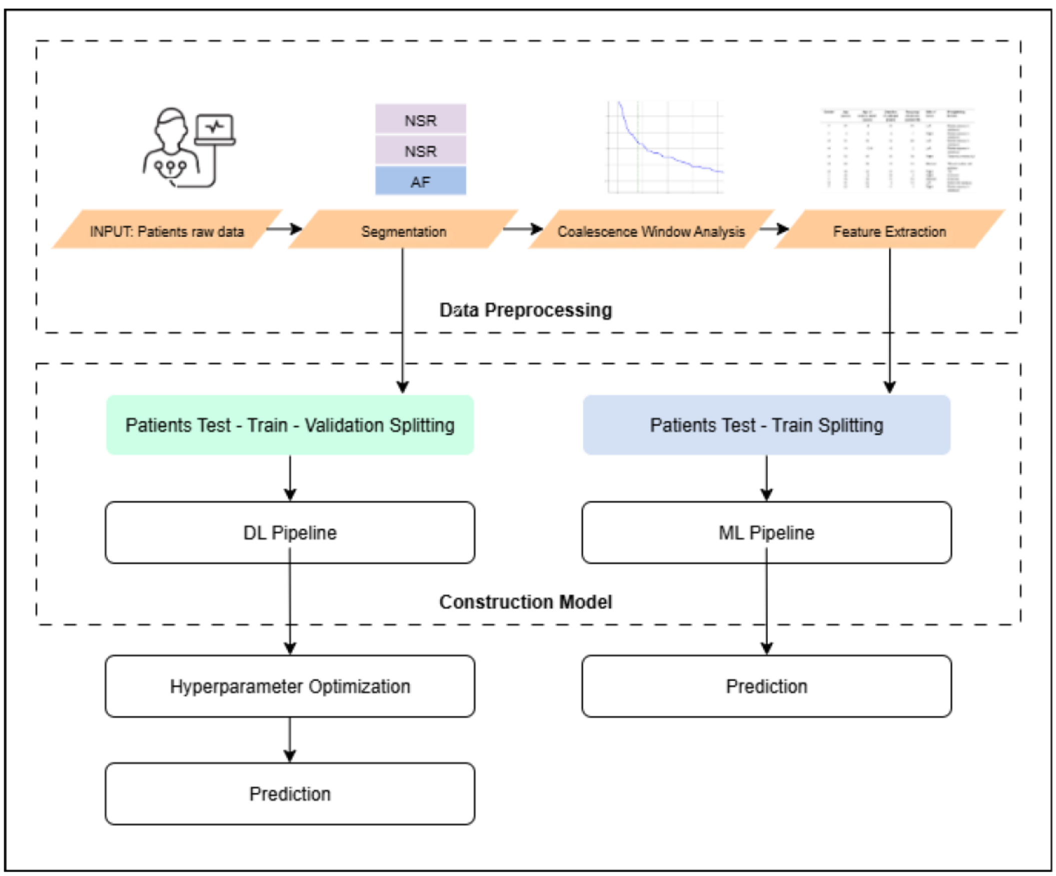

Figure 3.4: General pipeline of the proposed AI model for AF detection.

The proposed approach consists of two main stages: Data preprocessing and Model Training and Evaluation. The features extracted in the data preprocessing stage serve as input to the classifiers, trained to distinguish AF from NSR segments. The proposed approach is organized into two main stages: *Data Preprocessing* and *Model Training and Evaluation*. The features extracted during the preprocessing phase serve as input for the classical classifiers, which are trained to distinguish AF from NSR segments. In parallel, raw signals are processed by DL models specifically designed to capture temporal and morphological patterns relevant to AF detection.

Analyzing the workflow in more detail:

3.3.1 Preprocessing

Preprocessing represents a fundamental step of the proposed pipeline, ensuring that atrial fibrillation (AF) episodes are consistently detected and that both feature-based and DL approaches operate on standardised input data.

The methodological workflow adopted in this phase builds directly upon the procedures already established in the second stage of the research. Consequently, the *preprocessing pipeline* illustrated in Figure 3.4, including data extraction, coalescence-based segmentation, signal normalization, and segment labeling corresponds exactly to the one previously described. Since these operations have already been thoroughly detailed, they are not reiterated here. Instead, this section focuses exclusively on the additional methodological elements introduced to extend the analysis toward DL models, highlighting the adaptations required to support end-to-end learning from raw ECG signals.

Building on the previously generated dataset of labelled ECG segments, the third phase expands the analytical pipeline by introducing two essential components: *data splitting* and model *hyperparameter configuration*. These steps ensure methodological consistency across experiments and allow a fair comparison between traditional ML methods and the proposed DL architectures.

3.3.2 Data Splitting

To ensure a correct evaluation of both ML and DL models, the dataset was divided into training, validation, and testing subsets following strict criteria to avoid data leakage and preserve class balance. Since the initial dataset exhibited a significant imbalance between AF and NSR segments, random undersampling of the majority class (NSR) was applied before partitioning. This choice reduced bias during model training and ensured that the proportion of AF and NSR samples remained consistent across splits.

For the ML classifiers, a stratified split was applied by allocating 30% of the data to the test set, while the remaining 70% was used for training. DL models required an additional validation set: after an initial 70/30 split between training and testing, the test portion was further subdivided into 70% for testing and 30% for validation. This produced three non overlapping subsets, enabling proper hyperparameter tuning without contaminating the final evaluation.

To ensure fairness and comparability across experiments, the splitting procedure was executed once and reused identically for all models under consideration.

Hyperparameter	**Description**	**Values**
filters	Dimensionality of the output space, i.e., number of output filters	32, 64, 128
kernel_size	Length of the 1D convolution window	5
pool_size	Size of the max pooling window	2, 4
LSTM units	Number of hidden units in the BiLSTM layer	32
dropout_rate	Probability of cutting a node out during training	0.2, 0.3
batch_size	Number of samples per gradient update	8
n_epochs	Number of epochs to train the model	20
optimizer	Optimization algorithm used to minimize loss	Adam
loss	Function used to evaluate predictions	Binary cross-entropy
last_activation	Activation function of the last layer	Sigmoid

Figure 3.5: Considered hyperparameters and values for DL models.

3.3.3 Hyperparameter Configuration

For traditional ML classifiers, the default hyperparameter configurations from the corresponding libraries were intentionally adopted. This strategy allowed the establishment of a neutral and reproducible baseline across heterogeneous algorithmic families (tree-based, linear, probabilistic, kernel-based, and shallow neural models) without introducing biases resulting from dataset-specific tuning.

Conversely, DL models required an explicit optimization of several key hyperparameters to guarantee stable convergence and robust generalization. Figure 3.5 summarizes the main hyperparameters explored for the proposed CNN+BiLSTM architecture. Training was performed using binary cross entropy loss and the Adam optimizer, combined with regularization techniques such as dropout, max pooling, early stopping, and model checkpointing. These mechanisms prevented overfitting and ensured that the architecture learned meaningful spatio temporal patterns directly from the raw ECG traces.

This dual methodology reflects the complementary strengths of the two paradigms: feature based ML models offer transparency, interpretability, and low computational cost particularly suitable for edge or clinical environments while DL models exploit end-to-end learning to capture complex morphological and temporal dependencies, ultimately enabling higher diagnostic accuracy and improved generalization.

3.3.4 DL Pipeline

In parallel to the feature engineering pipeline, a dedicated DL framework was developed to operate directly on the raw ECG segments, thereby removing the dependency on handcrafted features. After preprocessing, each segment was reshaped into a three dimensional tensor of the form `(batch_size, sequence_length, 1)`, allowing the model to ingest the temporal signal in its original form.

The architecture adopted in this study follows a hybrid paradigm in which **convolutional neural layers (CNNs)** are combined with a **Bidirectional Long Short-Term Memory (BiLSTM)** module, as illustrated in Figure 3.6. The convolutional blocks serve as local feature extractors, capturing morphological structures and short-range temporal dynamics inherent in ECG waveforms. Their output is subsequently processed by BiLSTM units, which are specifically designed to model sequential dependencies in both forward and backward temporal directions. This dual directional processing enables the network to capture the recurrent patterns and temporal irregularities that characterise paroxysmal atrial fibrillation.

Following the temporal modelling stage, fully connected layers equipped with dropout regularisation perform the final classification into AF and NSR classes. To assess robustness, two architectural variants were additionally explored: a CNN only configuration and a CNN combined with Temporal Convolutional Networks (TCN). These alternative designs allowed a systematic comparison between different temporal modelling strategies.

Training procedures were carefully regularised to promote generalisation. Dropout layers were inserted throughout the network to mitigate overfitting, and early stopping in conjunction with model checkpointing was applied to retain the best-performing set of weights. All models were implemented using Keras with TensorFlow as backend and optimised through the Adam algorithm with binary cross entropy as the loss function.

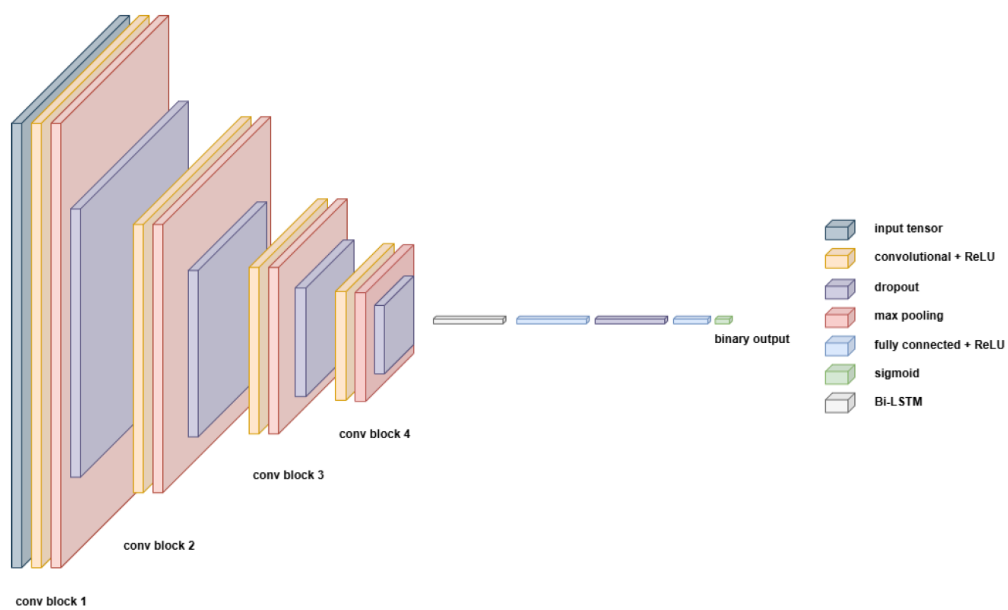

Figure 3.6: Architecture of the DL model: CNN layers extract local features, BiLSTM captures temporal dependencies, and dense layers perform final classification.

Model performance was assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrices. The DL pipeline demonstrated a strong capacity to learn highly discriminative representations directly from raw ECG data, capturing subtle temporal dependencies that feature based methods could not fully encode. Compared with the ML branch, the DL models exhibited superior generalisation ability, underscoring the effectiveness of end-to-end neural architectures for reliable detection of atrial fibrillation.

CHAPTER

4 Study Result

In this Chapter, the predictive results obtained to address each Research Question defined in Chapter 2 are discussed.

4.1 Preliminary Study and Application of AI Techniques

The continuous acquisition of physiological signals, behavioural indicators, and clinical parameters from wearable and ambient sensors provides unprecedented opportunities for data driven healthcare. At the same time, this evolution introduces new methodological challenges related to data processing, model selection, interpretability, and the protection of sensitive health information.

This section presents a detailed analysis of the methodologies and approaches adopted throughout this research project. The state of the art is investigated to understand current trends, technological solutions, and open issues in IoT based Smart Health [9].

RQ_1: What are the current technological approaches, challenges, and limitations in integrating IoT and AI within Smart Health systems?

Different methodological strategies can be adopted to analyse health-related data, with ML and DL representing the most widely employed approaches in the scientific literature.

The variety of available algorithms, ranging from traditional statistical learners to modern neural architectures, together with the presence of numerous customised variants, makes the landscape of possible solutions extremely broad.

Given this heterogeneity, an initial objective of our analysis was to determine which methodological paradigm (ML, DL, or both) is most frequently adopted in the selected studies. Understanding this distribution provides insight into the level of model complexity typically pursued and whether studies tend to rely on one family of algorithms or compare multiple approaches. Furthermore, for works adopting both ML and DL, the number of algorithms tested offers an indication of the depth and breadth of the experimental evaluation.

Figure 4.1 summarises these findings by reporting, on the x-axis, the number of studies employing ML alone, DL alone, or both approaches, together with the absolute counts (blue) and relative percentages (orange). The majority of contributions (59.62%) rely exclusively on ML techniques, whereas 25% adopt DL models. Only a limited subset of studies (15.38%) evaluates both paradigms within the same work.

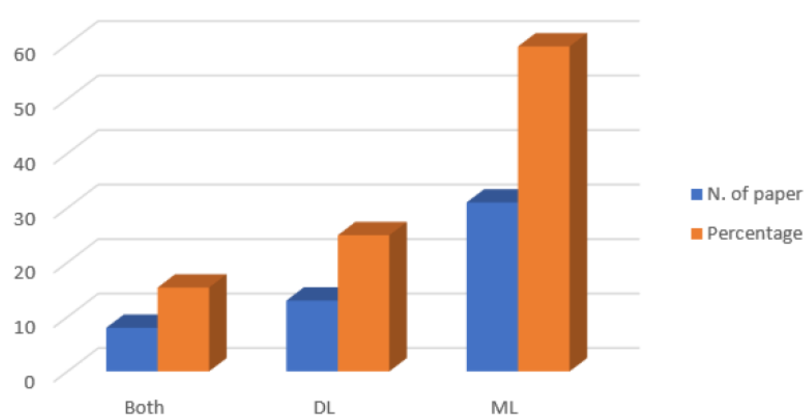

Figure 4.1: Distribution of the use of ML and DL techniques in the selected studies.

To provide a more detailed perspective, Figure 4.2 illustrates the distribution of the number of algorithms evaluated per study.

In panel (a), the distributions refer to papers using exclusively ML or exclusively DL. ML based works typically assess a larger number of classifiers, often more than three (median of the blue distribution). This reflects the relative computational efficiency of ML methods, which facilitates broader comparative testing. In contrast, DL focused studies usually evaluate a single neural architecture, likely due to the substantial computational time and memory requirements associated with training deep networks.

Panel (b) of Figure 4.2 reports the distribution for studies employing both ML and DL approaches. Here again, ML models outnumber DL ones, with the latter usually represented by a single architecture. This pattern confirms a recurring trend: even when researchers explore hybrid methodological frameworks, DL experimentation remains computationally constrained and therefore less extensive than ML evaluation.

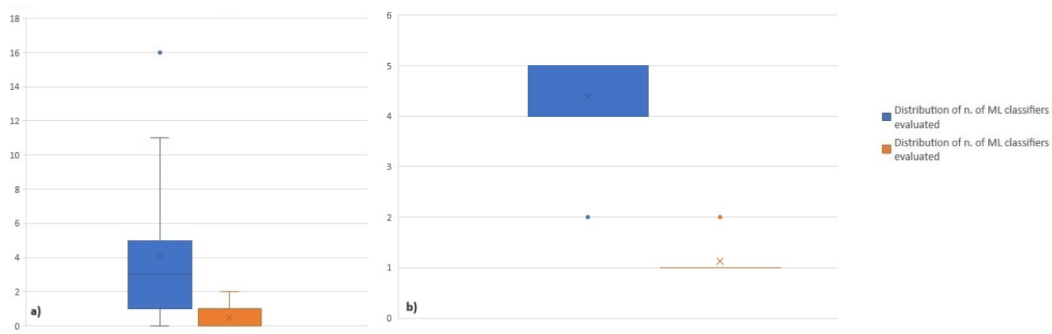

Figure 4.2: a) Distributions of the number of ML/DL classifiers evaluated in the studies b) Distributions of the number of ML/DL classifiers evaluated in the studies that used both methodologies

DL architectures play a key role in IoT based Smart Health systems, supporting the development of advanced applications for continuous monitoring, diagnosis, and personalised patient management. When combined with suitable data preprocessing and model optimisation strategies, these architectures enable early disease detection and improved handling of complex clinical conditions.

The reviewed literature shows that several neural network architectures are regularly adopted in IoT Smart Health studies, each addressing specific data modalities:

- **Convolutional Neural Networks (CNNs):** Widely used for image based diagnostics such as X-rays, CT scans, MRIs, and real-time video analysis. CNNs are effective in detecting patterns, anomalies, and clinically relevant visual structures.

- **Recurrent Neural Networks (RNNs):** Applied to time-series health data from wearable and IoT sensors, including continuous heart rate or blood pressure monitoring and activity tracking.

- **Long Short-Term Memory Networks (LSTMs):** A specialised type of RNN designed to capture long range dependencies in sequential medical data, making them suitable for vital sign monitoring and physiological time series analysis.

- **Densely Connected Convolutional Networks (DenseNet):** A CNN variant where each layer is connected to all subsequent layers, improving feature propagation. DenseNets are effective for medical image analysis, particularly when training datasets are limited.

- **Deep Neural Networks (DNNs):** Commonly used in remote patient monitoring systems, where data from wearable IoT devices must be integrated to identify anomalies and support real time interventions.

Figure 4.3 illustrates the distribution of medical conditions addressed in the analysed studies. The blue bars represent the number of publications, while the orange bars show their corresponding percentages.

The results indicate that the most frequently studied diseases are cardiovascular conditions (21.15%), followed by mental and neurodegenerative disorders (17.31%), and respiratory diseases (11.54%).

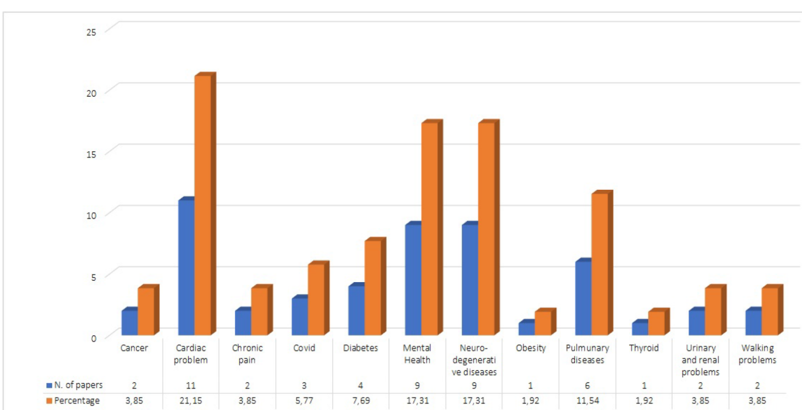

Figure 4.3: Distribution of pathologies analyzed in the selection of articles

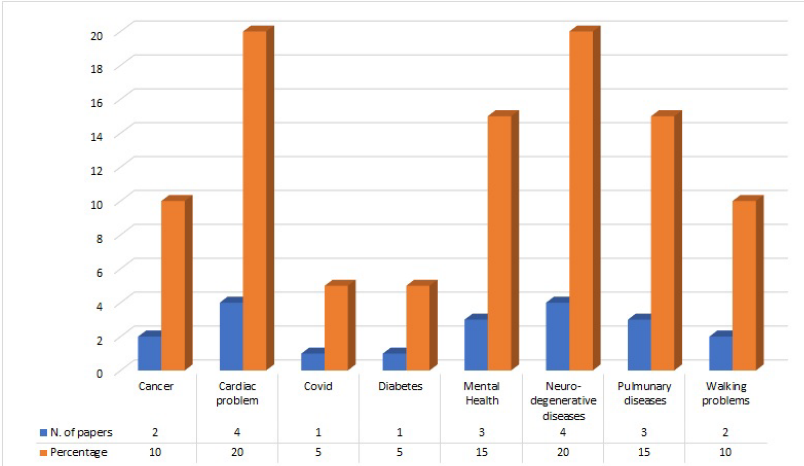

Figure 4.4: Distribution of pathologies analyzed with a DL approach

A detailed focus on works exclusively employing DL approaches is presented in Figure 4.4. These results reveal that complex architectures are particularly adopted for cardiological and neurodegenerative disorders (20%), as well as mental health and respiratory conditions (15%), reflecting the intrinsically multifaceted nature of these pathologies.

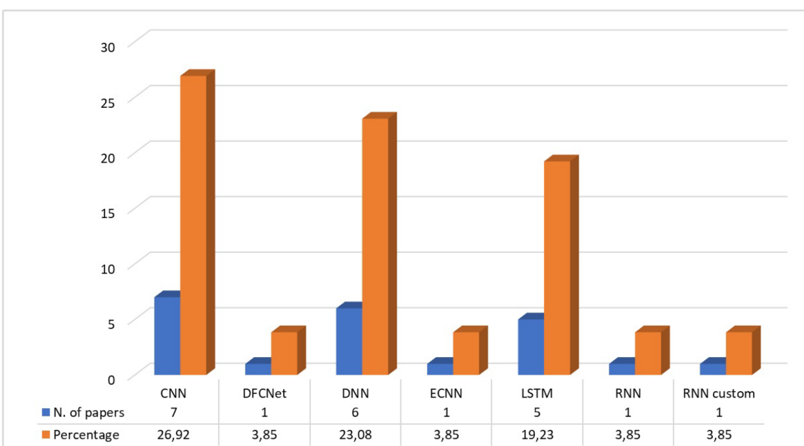

Figure 4.5: Number of papers and relative percentage, based on the type of Neural Network architecture

To understand which architectures are most commonly applied to IoT acquired medical data, Figure 4.5 reports the distribution of neural network models used across the selected studies. CNNs, DNNs, and LSTMs together account for nearly 60% of the architectures, confirming their prominence in IoT based Smart Health research.

Finally, Figure 4.6 provides a cross analysis linking neural network architectures to specific diseases.

CNNs are the most frequently used for neurodegenerative, mental health, diabetic, cardiovascular, and oncological conditions.

DNNs are more common in gait analysis, respiratory conditions, and neurodegenerative diseases, whereas LSTMs are preferred for cardiac monitoring, COVID-19 analysis, mental health assessment, pulmonary conditions, and gait related data.

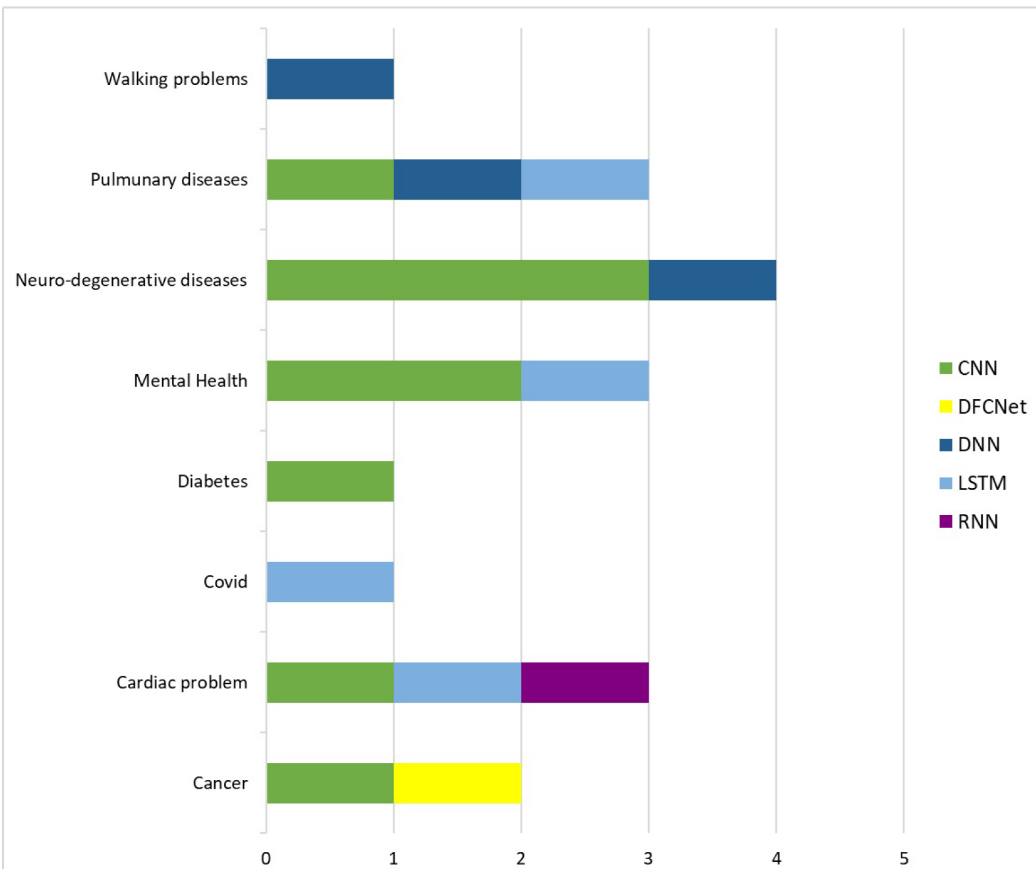

Figure 4.6: Distribution of Neural Network architectures based on the pathology analyzed by the different types of Neural Network architectures

ML and DL techniques play a fundamental role in enhancing both the accuracy and reliability of analyses performed on health data acquired through IoT monitoring devices. By learning patterns from historical data, these algorithms are able to identify clinically relevant trends, detect anomalies, and support continuous monitoring of vital signs such as heart rate, blood pressure, or glucose levels. Their ability to model complex relationships among physiological parameters makes them valuable tools for long term assessment, risk stratification, and the development of personalised treatment strategies aimed at improving patient outcomes and overall quality of life.

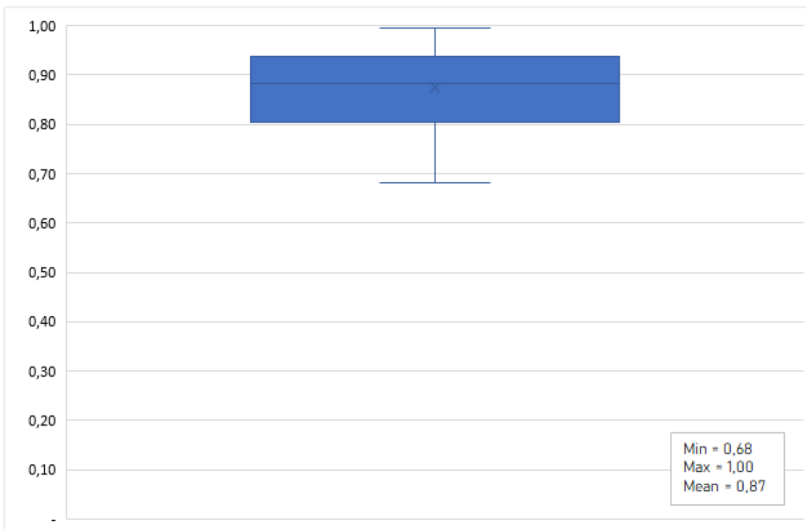

Figure 4.7: Distribution of values of accuracy metrics obtained with the approaches used in the selected papers

Given the importance of obtaining trustworthy predictions from these systems, it becomes essential to assess how effectively ML and DL algorithms perform on IoT generated health data. For this purpose, an empirical analysis was carried out to examine the distribution of classification accuracies reported across the selected studies.

The results, summarised in Figure 4.7, show a wide performance range: the lowest accuracy observed among the 46 analysed works was 68% (multinomial logistic regression), whereas the highest reached 100% (Cubic SVM).

Across all studies, the average accuracy was approximately 87%, with a confidence interval of [79.80%, 96%], indicating generally strong predictive performance across different methodological choices.

To further compare the effectiveness of ML versus DL techniques, Figure 4.8 reports the average accuracy achieved by each algorithm type. The three top performing approaches were:

- **Cubic SVM (100% average accuracy)** — a Support Vector Machine variant employing a cubic kernel to project data into higher dimensional spaces, enabling the separation of complex, non linear patterns typically found in biomedical signals.
- **J48 (99% average accuracy)** — a well established decision tree algorithm that recursively partitions the dataset based on information gain or Gini index optimisation. Its interpretability and solid performance explain its frequent adoption in clinical decision support systems.
- **Custom classifiers (97% average accuracy)** — models tailored from existing algorithms to better fit the specific characteristics of the health data under study. While potentially powerful, these approaches require careful validation to avoid overfitting or unintended biases.

Overall, the analysis reveals that both ML and DL techniques can achieve high levels of accuracy when applied to IoT-acquired health data.

However, ML models (especially kernel based or tree based approaches) often provide competitive performance with lower computational cost, whereas DL models offer greater modelling capacity at the expense of higher complexity and resource requirements.

RQ_2: What are the privacy and security issues taken into account while using health data collected from IoT devices?

The analysis also investigated to what extent the selected studies addressed the privacy and security challenges associated with the use of health data collected through IoT devices. Given the highly sensitive nature of physiological and clinical information, ensuring adequate protection during acquisition, transmission, and storage is essential. For this reason, specific attention was dedicated to identifying whether the reviewed works explicitly discussed these issues and how they proposed to mitigate them.

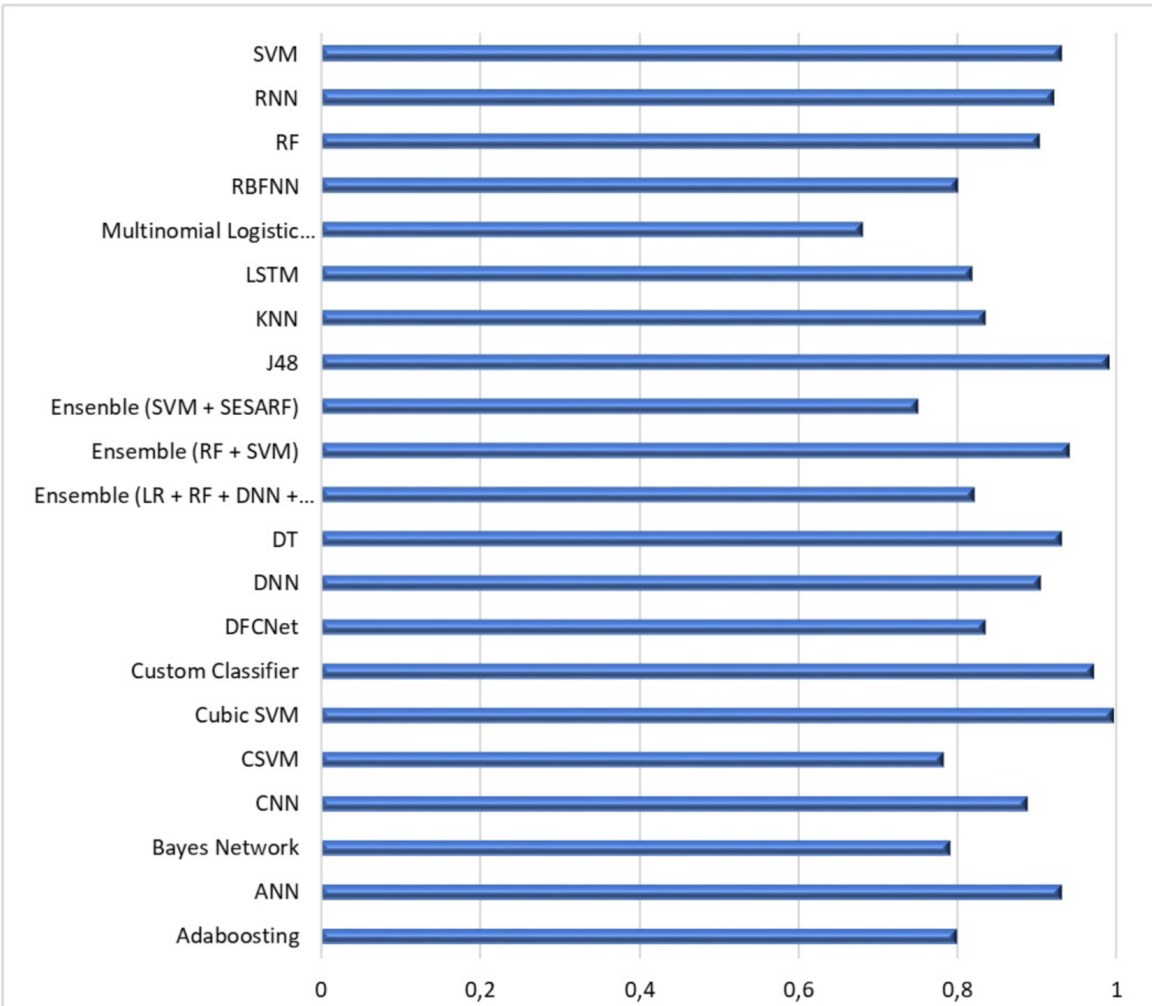

Figure 4.8: Average accuracy obtained by the approaches used in the selected papers

Overall, the results showed that privacy and security remain only partially considered in the current literature. Indeed, only about 24% of the analyzed contributions explicitly addressed either privacy or security aspects. In practical terms, this means that fewer than one in three studies using clinical or health related data reflected on whether the adopted IoT solutions ensured sufficient user protection or complied with established data handling standards.

Among the works that did consider such aspects, the most frequently examined issue was privacy during the data acquisition phase. Conversely, only 13.60% of the studies extended their analysis to include security of stored data, revealing a substantial gap in the overall treatment of these topics.

Several strategies were reported to mitigate privacy risks during data collection:

- in 2.20% of the studies, only derived features were stored instead of raw sensor data;
- in 4.30% of the works, video or audio recordings containing facial or vocal information were removed from the dataset before analysis;
- another 2.20% replaced video-based monitoring with a *Radio Frequency Identification* (RFID) system, which enables user detection without capturing visual data.

For studies addressing data security during and after storage, the main solutions adopted included:

- the use of edge–fog–cloud architectures in 4.30% of the contributions, distributing storage and reducing vulnerability to centralized attacks [19, 94];
- the application of encryption techniques, ensuring protection of data both in transit and at rest, also in 4.30% of the studies;
- the adoption of high security cloud storage protocols in 2.20% of the works;
- and, in the remaining 2.20%, the execution of regular security audits and continuous monitoring to detect vulnerabilities or unauthorized access.

These findings highlight that, despite the rapid evolution of IoT enabled Smart Health systems, privacy and security considerations remain insufficiently addressed in the current literature. Strengthening these aspects is essential to support safe, reliable, and ethically grounded deployment of IoT based healthcare technologies.

Conclusions

ML techniques demonstrate solid performance while maintaining high transparency, lower computational requirements, and easier interpretability, properties that make them suitable for clinical environments, remote monitoring systems, and edge based deployments. On the other hand, DL architectures exhibit superior capacity to capture complex temporal and morphological patterns directly from raw physiological signals, thereby achieving state of the art accuracy in tasks such as arrhythmia and atrial fibrillation detection. However, their effectiveness often depends on the availability of large datasets and adequate computational resources, which can limit their applicability in certain real world scenarios.

The results obtained across the different methodological phases highlight a clear trend: AI driven analysis of IoT generated health data represents a powerful tool to enhance diagnostic accuracy, support personalised treatment, and enable continuous patient monitoring. At the same time, the integration of such technologies must carefully consider privacy, data protection, and system reliability, especially when sensitive physiological information is continuously collected and processed.

Overall, the analysis we conducted highlighted the following results:

- most of the research papers reviewed deal with ML methodologies rather than DL techniques, and only a small fraction of both;
- most articles using ML algorithms use multiple classifiers, while those using DL algorithms typically use a single network, and this pattern remains true even in cases where articles use both approaches;
- the most used neural network architectures for IoT Smart Health problems are (i) CNNs when dealing with neurodegenerative diseases, mental problems, problems heart problems, diabetes and cancer, (ii) LSTM when it comes to heart problems, covid, mental health, lung disease diseases and walking problems, (iii) DNNs are mostly used in case of walking problems, heart diseases, diseases pulmonary and neuro degenerative diseases;
- the minimum accuracy value obtained is 68% while the maximum accuracy obtained is 100% and on average the studies reach an accuracy of 87%;

- privacy and security issues were evaluated in only 24% of the works analyzed.

In conclusion, the findings of this work emphasise both the potential and the challenges associated with the adoption of AI in Smart Health ecosystems. Future developments in this field should aim to design models that are not only accurate, but also interpretable, computationally efficient, and compliant with modern privacy standards. By advancing in these directions, AI powered healthcare systems will be able to support clinicians more effectively and provide patients with safer, more personalised, and proactive care.

4.2 Analysis of IoT Health Data Using Machine Learning Techniques

The integration of Internet of Things (IoT) technologies with Artificial Intelligence (AI) represents one of the most transformative trends in modern digital healthcare. IoT enabled Smart Health systems leverage interconnected sensors, wearable devices, and distributed computing infrastructures to continuously acquire physiological, behavioral, and environmental data. AI techniques ranging from traditional ML to advanced DL enable these data streams to be translated into meaningful clinical insights. The ML experiments were implemented using the Python ecosystem.[1] Despite the significant progress made, several technological approaches, challenges, and limitations still shape this domain. This section presents the results obtained from the classification experiments performed on the ECG signal segments extracted through the preprocessing pipeline [10].

RQ$_3$: What is the potential of ML techniques in extracting and analyzing clinically meaningful information from IoT generated health data?

The primary objective of this analysis was to distinguish atrial fibrillation (AF) from

[1] The classical ML models were implemented using scikit-learn, imbalanced-learn for random undersampling, XGBoost for gradient-boosted trees, CatBoost for categorical boosting, and joblib for model serialization.

normal sinus rhythm (NSR) using handcrafted features, with particular emphasis on frequency domain descriptors.

As an initial step, a set of preliminary experiments was conducted to evaluate the contribution of different feature families to the overall classification performance. When only descriptive statistical features, such as mean, standard deviation, skewness, and kurtosis, were used, the best performing model reached an accuracy of 0.77. Although this result was satisfactory, it also revealed that simple amplitude and distribution based descriptors are not sufficient to capture the complexity and variability typical of paroxysmal AF.

To investigate whether richer representations could enhance discrimination performance, the feature set was expanded to include temporal descriptors, spectral measures, nonlinear variability indices, and selected wavelet based components. The inclusion of these global handcrafted features produced a notable improvement, raising the accuracy to 0.83. This gain confirms the importance of integrating multiple feature domains to better describe the underlying morphology and dynamics of the ECG signals.

The observed improvement can be attributed to the complementary information provided by the various feature categories. Together, these results demonstrate that ML classifiers can effectively exploit handcrafted features to extract clinically meaningful information from IoT generated ECG data, establishing a solid performance baseline before transitioning to DL approaches.

RQ_4: What extent can Ml models enhance diagnostic accuracy, predictive capabilities, and decision support in Smart Health?

The family of traditional ML classifiers exhibited heterogeneous performance levels across the evaluated metrics, reflecting the trade off between model complexity, generalisation capability, and predictive power.

Each algorithm was assessed using standard performance indicators (accuracy, precision, recall, and F1-score) computed both on a held out test set and through a 5-fold cross validation scheme. This evaluation strategy was adopted to reduce the effects of class imbalance and to mitigate the risk of overfitting. The detailed results for the test set are reported in Figure 4.9, whereas Figure 4.10 presents the average cross validated metrics, providing a more stable estimate of model performance.

Across all evaluated classifiers, *XGBoost* consistently emerged as the best performing model, particularly with respect to recall and F1-score. This confirms the capability of gradient boosting methods to capture complex, nonlinear relationships within structured physiological data, making them highly suitable for early detection of paroxysmal AF.

When compared with state of the art DL solutions operating directly on raw ECG data (e.g. multi scale CNNs or hybrid CNN–BiLSTM architectures), the results reveal that feature based pipelines can achieve competitive performance while requiring significantly fewer computational resources.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-score**
Decision Tree (DT)	0.746	0.6724	0.6748	0.6736
Random Forest (RF)	0.8336	0.8456	0.7043	0.7365
Extra Trees (ET)	0.8292	0.8434	0.6948	0.7263
Gradient Boosting (GB)	0.7984	0.7933	0.6422	0.6631
XGBoost (XGB)	**0.8331**	**0.8131**	**0.7259**	**0.7524**
CatBoost (CB)	0.8320	0.8311	0.7086	0.7391
AdaBoost (AD)	0.7711	0.7194	0.6078	0.6186
KNN	0.7509	0.6646	0.6091	0.6194
Gaussian NB (NB)	0.3061	0.5456	0.5152	0.2794
Logistic Regression	0.7479	0.6993	0.5319	0.4966
Support Vector (SVC)	0.7518	0.7489	0.5340	0.4975
MLP Classifier	0.7842	0.7235	0.6680	0.6845

Figure 4.9: Global performance of the evaluated classifiers on the holdout test set

Classifier	Accuracy	Precision	Recall	F1-score
Decision Tree (DT)	0.7558	0.6831	0.6832	0.6831
Random Forest (RF)	0.8324	0.8416	0.7022	0.7340
Extra Trees (ET)	0.8311	0.8462	0.6966	0.7286
Gradient Boosting (GB)	0.8026	0.7970	0.6494	0.6720
XGBoost (XGB)	**0.8346**	**0.8108**	**0.7312**	**0.7565**
CatBoost (CB)	0.8357	0.8330	0.7157	0.7464
AdaBoost (AD)	0.7698	0.7160	0.6026	0.6120
KNN	0.7303	0.6256	0.5877	0.5941
Gaussian NB (NB)	0.7280	0.6026	0.5489	0.5424
Logistic Regression	0.7449	0.6621	0.5295	0.4952
Support Vector (SVC)	0.7487	0.7169	0.5269	0.4853
MLP Classifier	0.5691	0.4797	0.5200	0.4138

Figure 4.10: Mean cross validation performance of tested classifiers

For example, Xia et al. [73] report an F1-score of 0.848 using a dilated CNN on PhysioNet AFDB, whereas Terry et al. [115] obtain 0.84 with a CNN+BiLSTM architecture on 30 second ECG–PPG segments. In contrast, the best performing model in our study XGBoost reached an F1-score of 0.84 using only handcrafted features, highlighting its suitability for lightweight, interpretable AF detection in mobile or embedded settings.

Among all evaluated models, XGBoost achieved the highest macro averaged F1-score (0.7565), followed closely by CatBoost (0.7464) and Random Forest (0.7340). These ensemble based classifiers offered a good balance between sensitivity and specificity, confirming their suitability for biomedical classification tasks involving subtle signal variations. Extra Trees and Gradient Boosting also produced solid results (macro F1-scores of 0.7286 and 0.6720, respectively). By contrast, linear approaches such as Logistic Regression and Support Vector Classifier struggled to capture the nonlinear structure of ECG morphology, resulting in F1-scores below 0.53.

Gaussian Naive Bayes showed the weakest performance (F1-score 0.5424), primarily due to violations of its assumptions regarding feature independence and Gaussian distributed inputs. Kernel based and distance based models such as SVC and KNN performed moderately better (accuracy around 0.66–0.68), but their precision–recall balance remained suboptimal, with KNN particularly sensitive to noise and overlapping class regions.

Ensemble tree based classifiers Random Forest and Extra Trees demonstrated the most robust performance overall, with accuracies exceeding 0.80 and balanced metrics across both AF and NSR. However, even these models reached a performance ceiling, with no algorithm surpassing an accuracy of 0.81, showing the inherent limitations of relying solely on handcrafted features to capture the rich spatio temporal dynamics that characterize AF.

The MLPClassifier achieved competitive but lower performance (accuracy 0.74), outperforming linear and kernel based models but still falling short of ensemble approaches. Overall, the ML models provided efficient and interpretable baselines with reasonable discriminative power, yet their inability to exceed the performance obtained from handcrafted features alone motivates the exploration of DL architectures capable of learning more expressive representations directly from raw ECG signals.

Conclusions

This work proposes an interpretable and computationally efficient ML framework for the detection of paroxysmal atrial fibrillation (PAF) from long term ECG recordings. The methodology combines coalescence based segmentation, handcrafted feature extraction in the time, frequency, and nonlinear domains, and a range of supervised classifiers, with a particular emphasis on ensemble tree based models. Among all the algorithms evaluated, XGBoost and CatBoost consistently achieved the most competitive results, reaching macro F1-scores above 0.75 and demonstrating strong robustness in distinguishing AF from NSR segments.

The experiments showed that augmenting the feature set with frequency domain descriptors yielded approximately a 1% improvement over statistical features alone, underscoring their relevance for capturing subtle periodic and morphological patterns associated with AF episodes. To maintain realistic evaluation conditions and avoid introducing synthetic bias, no oversampling or data augmentation techniques were applied. Instead, the assessment relied on a combination of hold out testing and macro averaged metrics computed through cross validation, ensuring fairness under the naturally imbalanced class distribution typical of real world clinical data.

The resulting framework is lightweight, transparent, and suitable for deployment on resource constrained platforms such as wearable monitors or edge devices, where computational efficiency and explainability are essential. Nevertheless, several limitations should be noted. First, the evaluation is based solely on the IRIDIA-AF dataset. Although this dataset is extensive and well annotated, external validation on independent cohorts would be necessary to fully assess generalizability across different acquisition conditions and patient populations.

Second, the current pipeline treats each ECG segment as an isolated instance. This prevents the model from capturing long range temporal dependencies or patient specific trends. Integrating sequential learning approaches such as Temporal Convolutional Networks, or recurrent architectures represents a promising direction for future refinement.

Third, the reliance on handcrafted features inherently limits the capacity of the pipeline to learn complex, high level abstractions present in ECG morphology. While this choice enhances interpretability, complementary DL approaches could extract richer representations directly from raw signals, potentially improving performance in more challenging scenarios.

Finally, although the proposed system is well suited for offline analysis and periodic screening, future work should explore real time implementation constraints, energy efficient execution, and adaptive mechanisms that can maintain performance during prolonged continuous monitoring.

4.3 Deep Learning Model Design and Experimental Evaluation for AF Detection

This section details the DL pipeline developed to extend the analytical framework introduced in the previous phase of the study. After establishing a baseline with feature engineered ML classifiers, we shift the focus toward end-to-end neural architectures designed to operate directly on raw ECG segments extracted from IoT based monitoring systems. The DL models were implemented in Python using TensorFlow/Keras.[2]

The purpose of this phase [8] is to evaluate whether DL models can overcome the limitations of handcrafted representations by exploiting hierarchical feature extraction, temporal

[2]The deep learning architectures were implemented using TensorFlow/Keras. Additional libraries included keras-tcn for Temporal Convolutional Networks and Matplotlib for visualization of learning curves and performance plots.

context modelling, and non linear pattern recognition. Several architectural variants, including shallow CNNs, deeper convolutional stacks, and hybrid CNN–BiLSTM designs, were systematically tested to quantify their impact on diagnostic accuracy, recall, and overall generalization ability.

RQ_5: What is the potential of DL techniques in extracting and analyzing clinically meaningful information from IoT generated health data?

DL techniques demonstrate substantial potential for extracting, modelling, and interpreting clinically meaningful information from IoT generated health data, particularly in scenarios characterised by high data complexity, temporal variability, and subtle diagnostic patterns. In the context of this work, the development and evaluation of DL models applied to long term ECG recordings illustrate several key advantages offered by DL approaches in Smart Health applications.

First, DL models operate directly on raw physiological signals, eliminating the need for handcrafted feature design and enabling the automatic discovery of discriminative patterns. For example, the hybrid *CNN–BiLSTM* architecture explored in this study effectively captures both local morphological characteristics of ECG waveforms (via convolutional layers) and long range temporal dependencies associated with paroxysmal atrial fibrillation (via bidirectional recurrent units). This ability to learn hierarchical and multi scale representations gives DL models a clear advantage in detecting clinically subtle or transient events that may be missed by traditional ML pipelines.

Second, DL techniques show improved capacity for modelling the non linear and non stationary behaviour typical of biological signals acquired from wearable IoT devices. By leveraging temporal context, recurrent and convolutional architectures can analyse continuous data streams and adapt to variability in heart rhythm, noise conditions, and recording settings conditions that are inherently challenging for classical ML classifiers.

Third, despite requiring structured preprocessing, the DL pipeline used in this work demonstrated competitive performance compared to state of the art models in the literature, reaching accuracy and F1-score levels comparable to more computationally intensive architectures.

This result highlights the suitability of DL approaches for supporting diagnostic tasks such as arrhythmia detection in real world, long term monitoring scenarios.

Finally, DL techniques open the path toward more advanced applications, including real time risk prediction, personalised modelling, cross signal fusion (e.g. ECG–PPG), and adaptive monitoring systems capable of learning from continuous data streams. These capabilities make DL a powerful tool for improving diagnostic accuracy, enhancing predictive performance, and enabling proactive decision support in Smart Health ecosystems.

In summary, the results obtained in this work confirm that DL techniques hold significant potential for extracting clinically relevant information from IoT acquired health signals, providing robust, scalable, and high performing solutions for next generation intelligent healthcare systems.

RQ_6: To what extent can DL models enhance diagnostic accuracy, predictive capabilities, and decision support in Smart Health?

DL models demonstrated a substantial ability to improve diagnostic precision, predictive robustness, and decision support reliability when applied to ECG signals collected through IoT based monitoring systems. Unlike traditional ML approaches, whose performance is constrained by the expressiveness of handcrafted features, DL architectures learn hierarchical, data driven representations that capture both the morphological structure of ECG waveforms and the long range temporal dependencies characteristic of atrial fibrillation (AF).

Across the experimental evaluation, DL models consistently outperformed all feature based classifiers, surpassing the accuracy ceiling of approximately 81% observed in traditional approaches.

The Figure 4.11 presents a row for each experiment performed, while the columns describe respectively the phase to which the experiment refers, the experiment number, the length of the ECG segment, whether the BiLSTM block was considered and if so the related neurons, how many convolutional blocks were considered and their relative sizes, how many pooling blocks were considered and their relative sizes, the number of epochs designated for training, the number of epochs actually performed, the patience set for early stopping, and the last 4 columns concern the validation metrics previously described (Accuracy, Precision, Recall and F1-Score).

The results highlight a clear four phase evolution, where each architectural and parametric change contributed to progressively improving the model's performance. In the first phase *(exp1-4)*, we used segments with original length ($\approx$ 166 k samples) and a relatively simple CNN architecture, consisting of two Conv1D blocks (32 and 64 filters) followed by MaxPooling (pool size 2), GlobalAveragePooling, and a fully connected layer of 32 neurons. The main hyperparameters were gradually varied, increasing the maximum number of epochs (from 20 to 50 to 100) and the early stopping patience (from 5 to 10). Performance progressively improved from an accuracy of 0.76 in *exp1* to 0.90 in *exp4*, indicating that a greater number of epochs allowed for more stable convergence, albeit with very long training times and an increasing risk of overfitting. In this case, the model tended to learn noisy local patterns, unable to effectively exploit the long term temporal relationships present in the ECG signal. In phase two *(exp5)*, to address the training time issue, the segment length was reduced to approximately 83k samples while keeping the architecture unchanged. This choice resulted in a significant reduction in time per epoch and greater training stability, while still maintaining good performance (accuracy 0.85). This indicates that shorter segments still contain sufficient discriminative information for AF/NSR classification and may be preferable when computational resources are limited.

Phase	Experiment	segment length	BiLSTM	Conv1D	Pool-size	Epochs	Epochs performed	Patience	Accuracy	Precision	Recall	F1-score
1	exp1	original	x	2 (32-64)	(2-2)	20	20	5	0.7678	0.7702	0.7679	0.7673
1	exp2	original	X	2 (32-64)	(2-2)	50	39	5	0.8334	0.8346	0.8334	0.8333
1	exp3	original	X	2 (32-64)	(2-2)	100	59	5	0.8496	0.8506	0.8495	0.8494
1	exp4	original	X	2 (32-64)	(2-2)	100	100	10	0.9011	0.9011	0.9011	0.9011
2	exp5	halved	X	2 (32-64)	(2-2)	150	62	10	0.8505	0.8507	0.8507	0.8505
3	exp6	halved	✓(32)	4 (32-64-64-128)	(2-2-2-2)	10	10	10	0.8759	0.8766	0.8763	0.8759
4	exp7	halved	✓(32)	4 (32-64-64-128)	(4-4-2-2)	20	20	10	**0.965**	**0.9652**	**0.9649**	**0.965**

Figure 4.11: Results obtained from performing the four different phases of DL pipeline experiments

In phase three *(exp6)*, to overcome the intrinsic limitations of CNNs alone, which capture predominantly local patterns, a 32-unit Bidirectional LSTM block was introduced after four convolutional blocks (32–64–64–128 filters). This allowed the network to also model long term temporal dependencies, resulting in a significant improvement in metrics (accuracy 0.88). The synergistic effect between local feature extraction (Conv1D) and sequential modelling (BiLSTM) represented the first real qualitative leap. Finally, in phase four *(exp7)*, the *(exp6)* architecture was optimised by increasing the pooling aggressiveness in the first two blocks (pool size 4 instead of 2) to reduce dimensionality and computation time, and increasing the number of epochs to 20. This produced the best overall result, with an accuracy of 0.965 and virtually identical precision, recall, and F1 scores, indicating a well balanced and generalizable model. The stability of the validation curves and the low loss also show that the dropout setting (0.2/0.3) is sufficient to prevent overfitting, even with a more complex model. Figure 4.12 provides the detailed classification report; the results show very high overall performance and balance between the two classes.

Class	Precision	Recall	F1-score	Support
NSR (0)	0.97	0.96	0.96	929
AF (1)	0.96	0.97	0.97	957
Accuracy			**0.97**	**1886**
Macro Avg	**0.97**	**0.96**	**0.96**	**1886**
Weighted Avg	**0.97**	**0.97**	**0.96**	**1886**

Figure 4.12: Classification report of the best performing DL model

Specifically, for the NSR (0) class, the model achieves a precision of 0.97 and a recall of 0.96, while for the AF (1) class, it achieves virtually symmetric values (precision 0.96, recall 0.97). The F1 score, which balances precision and recall, is 0.96–0.97 for both classes, confirming balanced and generalizable behaviour. The overall accuracy is 0.97 out of a total of 1,886 tested samples.

The confusion matrix (Figure 4.13) confirms this result: out of 929 NSR samples, 888 are correctly classified and only 41 are false positives (classified as AF); out of 957 AF samples, 932 are correct and 25 are false negatives. In both cases, the number of errors is very low and distributed symmetrically between the two classes, indicating the absence of residual bias.

	Predicted NSR	Predicted AF
True NSR	888	41
True AF	25	932

Figure 4.13: Confusion matrix of the best performing DL model

These results significantly outperformed all feature based classifiers, underscoring the ability of DL to uncover discriminative spatio temporal representations that handcrafted features could not capture. Figure 4.14 shows the accuracy and loss comparison during training and validation for the first *(exp1)* and last *(exp7)* DL experiments. In the case of *exp1*, based on a simple CNN architecture with two convolutional blocks and long segments (166k samples), the validation accuracy increases slowly and tends to stabilise after just a few epochs at values around 0.75, while the validation loss remains high (0.53) and shows fluctuations, indicating unstable convergence and limited predictive capacity. In contrast, in *exp7*, which uses a deeper network with four convolutional blocks and a BiLSTM layer trained on halved segments ($\approx$ 83 k samples), we observe a rapid increase in validation accuracy to values close to 0.97 and a decrease in validation loss to approximately 0.08. Furthermore, the training and validation curves are very close, indicating no overfitting and excellent generalisation. This comparison highlights how the introduction of recurrent components (BiLSTM), the increase in convolutional depth, and the reduction in segment length enabled faster and more stable convergence, dramatically improving the model's overall performance.

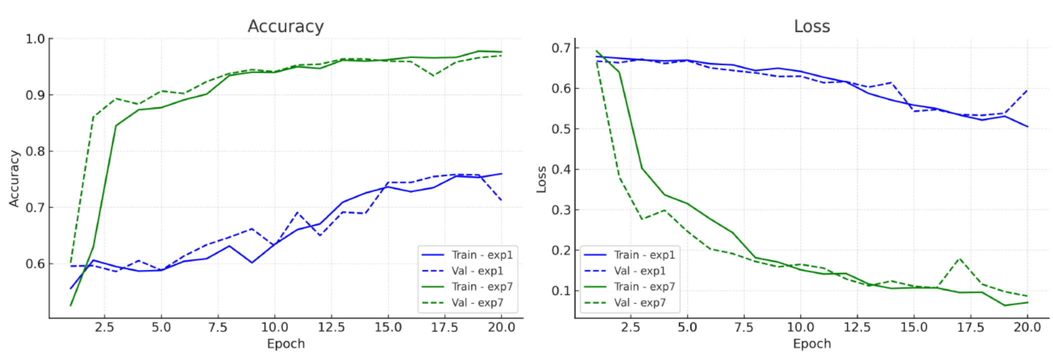

Figure 4.14: Representative learning curves (accuracy and loss vs. epochs) for the comparison between *exp1* and *exp7*

The progressive refinement of the architecture, from shallow convolutional models to deeper CNNs enriched with Bidirectional LSTM units, led to marked improvements in model capability. The most advanced configuration achieved a validation accuracy of 0.97, with precision, recall, and F1-scores all in the range 0.96–0.97 for both AF and NSR classes. These results highlight the capacity of deep networks to extract discriminative spatio temporal patterns that remain inaccessible to handcrafted feature pipelines.

The introduction of recurrent layers, specifically BiLSTM units, played a decisive role in enhancing predictive performance. By modelling sequential dynamics in both temporal directions, the network was able to capture subtle transitions between rhythms and better recognise paroxysmal AF episodes. Furthermore, optimisation of the convolutional depth and pooling configurations not only reduced training time but also improved generalisation, as reflected by stable and closely aligned training and validation curves.

The resulting model displayed excellent classification balance, with very low false-positive and false-negative rates, confirming its reliability for clinical decision making scenarios in which missed detections or erroneous alerts may have critical implications. The confusion matrix further validated this behaviour, showing symmetric error distribution and strong discrimination between AF and NSR.

Taken together, these findings show that DL models significantly enhance the capacity of Smart Health systems to support automated diagnosis, long term condition monitoring, and early event detection. Their ability to learn meaningful and generalisable representations directly from raw physiological signals enables a level of diagnostic accuracy and predictive insight that exceeds that of traditional ML models, making DL a powerful asset for next generation, IoT driven healthcare solutions.

Conclusions

The comparison between ML and DL models highlights a clear advantage of end-to-end neural architectures in the context of AF detection from long term ECG recordings. While tree based ML classifiers such as XGBoost and CatBoost achieved solid baseline performance, reaching approximately 76% accuracy with balanced precision recall profiles, their effectiveness is inherently limited by the quality and expressiveness of handcrafted features. These models struggle to fully capture the irregular, non stationary, and multi scale temporal dynamics characteristic of paroxysmal AF.

In contrast, the proposed DL architecture, combining multiple convolutional blocks with a Bidirectional LSTM module, demonstrated substantially superior performance. By learning directly from raw ECG segments, the network was able to extract both local morphological descriptors and long range temporal dependencies, achieving an accuracy of 97% and near perfect precision, recall, and F1-scores. The resulting confusion matrix confirmed highly balanced classification across both AF and NSR classes, underscoring the robustness and clinical reliability of the DL solution.

Overall, the results clearly show that although ML classifiers remain valuable for lightweight or resource constrained deployments, DL models provide the most accurate, reliable, and clinically meaningful predictions. Their ability to autonomously learn hierarchical representations from raw biosignals positions them as the most promising paradigm for next generation Smart Health diagnostic and decision support systems. Future developments should further explore generalization across diverse patient populations, incorporate transfer learning strategies, and integrate explainability frameworks to strengthen clinical interpretability.

Conclusions

This thesis investigated the problem of atrial fibrillation (AF) detection from long-term ECG recordings through a structured, multi-phase research approach.

In the first phase, a systematic literature review was conducted to analyze the current state of the art in AF detection, with particular attention to the methodological strategies adopted in both traditional ML and DL approaches. This review allowed the identification of prevailing trends, common limitations, and open challenges in the analysis of long-term ECG signals, providing a solid theoretical and methodological foundation for the subsequent experimental work.

The second phase focused on an exploratory study of the available dataset, restricted to traditional ML applications based on handcrafted features. Several ML classifiers were evaluated in order to assess whether meaningful patterns related to AF episodes could be captured from the data and to establish reliable baseline performances. While these models, with XGBoost in particular, demonstrated reasonable effectiveness, their ability to model the complex temporal and morphological characteristics of ECG signals remained inherently limited.

In the third phase, the study shifted toward a DL-based modeling approach aimed at overcoming the limitations observed in traditional ML methods. An end-to-end architecture combining convolutional layers with bidirectional LSTM units was designed to operate directly on raw ECG segments. This approach proved significantly more effective, achieving superior performance across all evaluation metrics and reaching an accuracy of up to 97% in the best configuration. The analysis of confusion matrices and classification reports confirmed not only a reduction in false positives but also an improved sensitivity to AF episodes, highlighting the robustness and clinical relevance of the proposed model.

Overall, the results of this thesis demonstrate the advantages of DL and end-to-end learning strategies over feature-based ML approaches for AF detection from long-term ECG recordings. While traditional ML models remain valuable for exploratory analysis and rapid prototyping, DL architectures are better suited to capture subtle, long-range dependencies in physiological signals and to extract clinically meaningful representations directly from raw biomedical data.

From a broader perspective, these findings contribute to the development of intelligent Smart Health ecosystems, in which wearable sensors, IoT infrastructures, and advanced analytical models cooperate to enable continuous and proactive healthcare monitoring. The ability of DL models to process large volumes of physiological signals and automatically identify clinically relevant patterns makes them particularly suitable for integration into IoT-based healthcare systems, where data are continuously generated by wearable and remote monitoring devices. In this context, the proposed framework demonstrates how advanced AI techniques can support early arrhythmia detection and long-term cardiac monitoring within next-generation digital healthcare infrastructures.

Study Limitations and Future Research Directions

Despite the promising results obtained, several methodological limitations should be acknowledged. The experimental validation conducted in this work relies on a single dataset, namely the IRIDIA-AF database, and focuses exclusively on a single clinical condition, atrial fibrillation detection. While this choice allowed for a controlled and in-depth investigation of the proposed methodology, it naturally limits the generalizability of the results across different patient populations, acquisition devices, and clinical scenarios.

In real-world Smart Health environments, physiological signals collected from wearable IoT devices may exhibit substantial variability due to sensor characteristics, environmental noise, motion artefacts, and patient-specific factors. Consequently, although the proposed DL architecture demonstrated strong performance within the considered dataset, further validation on additional datasets and across diverse clinical conditions will be necessary to assess the robustness and broader applicability of the proposed approach.

Another important aspect concerns model interpretability. Although explainability was identified as a relevant methodological dimension for the clinical adoption of the proposed DL framework, explicit interpretability outputs were not included among the reported experimental results. Future developments should therefore integrate saliency-based analyses or other explainable AI techniques in order to better understand which temporal regions and signal characteristics drive the model's predictions, thus improving transparency, clinical plausibility, and trust in automated decision support.

From a translational perspective, future work should also further investigate the implications of model errors in continuous monitoring scenarios. While the achieved performance metrics are encouraging, a deeper analysis of false positives and false negatives would be essential to evaluate the real-world impact of the proposed system in clinical practice. In long-term Smart Health monitoring, the balance between sensitivity and specificity is particularly critical: high sensitivity may support early detection of AF episodes, but an excessive false-positive rate may lead to unnecessary alerts, increased clinician workload, and alert fatigue. Conversely, false negatives may reduce the effectiveness of the system in identifying clinically relevant arrhythmic events in a timely manner.

For these reasons, future research should not only focus on improving predictive performance, but also on evaluating the proposed framework in more realistic deployment settings, where risk management, usability, and clinical workflow integration become central. Prospective validation, patient-level analysis, and the assessment of alert strategies adapted to different monitoring contexts would further strengthen the translational relevance of the work.

Future research will therefore focus on extending the proposed framework to larger and more heterogeneous datasets, exploring transfer learning and self-supervised pretraining strategies to improve model generalization. In addition, the integration of explainability techniques and multimodal physiological signals acquired from wearable IoT devices, such as ECG, PPG, and activity data, represents a promising direction for the development of more comprehensive, transparent, and clinically reliable Smart Health monitoring solutions.

Bibliography

[1] Sundus Abrar, Chu Kiong Loo, and Naoyuki Kubota. “A Multi-Agent Approach for Personalized Hypertension Risk Prediction”. In: *IEEE Access* 9 (2021), pp. 75090–75106.

[2] Rishiraj Adhikary et al. “SpiroMask: Measuring Lung Function Using Consumer-Grade Masks”. In: 4.1 (2023). DOI: 10.1145/3570167. URL: https://doi.org/10.1145/3570167.

[3] Sahar Ahmadzadeh, Jikui Luo, and Richard Wiffen. “Review on Biomedical Sensors, Technologies and Algorithms for Diagnosis of Sleep Disordered Breathing: Comprehensive Survey”. In: *IEEE Reviews in Biomedical Engineering* 15 (2022), pp. 4–22.

[4] Muhammad Mahtab Alam et al. “A Survey on the Roles of Communication Technologies in IoT-Based Personalized Healthcare Applications”. In: *IEEE Access* 6 (2018), pp. 36611–36631.

[5] Michael Armbrust et al. “A View of Cloud Computing”. In: *Communications of the ACM* 53.4 (2010), pp. 50–58.

[6] A. Asadi, M. Sharma, N. Ghasemi, et al. “Automatic AF Detection Using GAN-Augmented ECG and Neural Architecture Search”. In: *arXiv preprint* arXiv:2301.10173 (2023).

[7] Kevin Ashton. “That ’Internet of Things’ thing”. In: *RFID Journal* (2009).

[8] Name Author1, Name Author2, and Name Author3. "Artificial Intelligence Applications in Healthcare: Current Trends and Future Perspectives". In: *Applied Sciences* 16.5 (2026), p. 2390. DOI: 10.3390/app16052390. URL: https://www.mdpi.com/2076-3417/16/5/2390.

[9] Lerina Aversano et al. "A systematic Review on the use of Artificial Intelligence Approaches and Smart Health Devices". In: *PeerJ Computer Science* (2023).

[10] Lerina Aversano et al. "From Raw ECG to Classification: a Lightweight Approach to Paroxysmal AF Detection". In: *IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)* (2025).

[11] N. Bahador et al. "Multimodal spatio-temporal-spectral fusion for deep learning applications in physiological time series processing". In: *Inf. Fusion* 73 (2021), pp. 125–143.

[12] Christos C. Bellos et al. "Identification of COPD Patients' Health Status Using an Intelligent System in the CHRONIOUS Wearable Platform". In: *IEEE Journal of Biomedical and Health Informatics* 18.3 (2014), pp. 731–738.

[13] Hongliang Bi, Jiajia Liu, and Nei Kato. "Deep Learning-Based Privacy Preservation and Data Analytics for IoT Enabled Healthcare". In: *IEEE Transactions on Industrial Informatics* 18.7 (2022), pp. 4798–4807.

[14] Flavio Bonomi et al. "Fog computing and its role in the Internet of Things". In: *Proceedings of MCC Workshop* (2012).

[15] R. Bouhenguel and I. Mahgoub. "A Risk and Incidence Based Atrial Fibrillation Detection Scheme for Wearable Healthcare Computing Devices". In: *PervasiveHealth Conf.* 2012.

[16] Leo Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32.

[17] Leo Breiman et al. *Classification and Regression Trees*. Wadsworth, 1984.

[18] Yi Cai et al. “Machine-learning approaches for recognizing muscle activities involved in facial expressions captured by multi-channels surface electromyogram”. In: *Smart Health* 5-6 (2018), pp. 15–25.

[19] Chinmay Chakraborty and Amit Kishor. “Real-Time Cloud-Based Patient-Centric Monitoring Using Computational Health Systems”. In: *IEEE Transactions on Computational Social Systems* 9.6 (2022), pp. 1613–1623.

[20] Kévin Chapron et al. “Acti-DM1: Monitoring the Activity Level of People With Myotonic Dystrophy Type 1 Through Activity and Exercise Recognition”. In: *IEEE Access* 9 (2021), pp. 49960–49973.

[21] Omar Cheikhrouhou et al. “One-Dimensional CNN Approach for ECG Arrhythmia Analysis in Fog-Cloud Environments”. In: *IEEE Access* 9 (2021).

[22] Tianqi Chen and Carlos Guestrin. “XGBoost: A scalable tree boosting system”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785–794.

[23] Xianda Chen et al. “ApneaDetector: Detecting Sleep Apnea with Smartwatches”. In: 5.2 (2021). DOI: `10.1145/3463514`. URL: `https://doi.org/10.1145/3463514`.

[24] Xuhang Chen et al. “An IoT and Wearables-Based Smart Home for ALS Patients”. In: *IEEE Internet of Things Journal* 9.21 (2022), pp. 20945–20956.

[25] Po-Han Chiang and Sujit Dey. “Offline and Online Learning Techniques for Personalized Blood Pressure Prediction and Health Behavior Recommendations”. In: *IEEE Access* 7 (2019), pp. 130854–130864.

[26] Po-Han Chiang, Melissa Wong, and Sujit Dey. “Using Wearables and Machine Learning to Enable Personalized Lifestyle Recommendations to Improve Blood Pressure”. In: *IEEE Journal of Translational Engineering in Health and Medicine* 9 (2021), pp. 1–13.

[27] Prerna Chikersal et al. “Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach With Robust Feature Selection”. In: 28.1 (2021). ISSN: 1073-0516. DOI: 10.1145/3422821. URL: https://doi.org/10.1145/3422821.

[28] Willy Chou et al. “Design of Smart Brain Oxygenation Monitoring System for Estimating Cardiovascular Disease Severity”. In: *IEEE Access* 8 (2020), pp. 98422–98429.

[29] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine Learning* 20.3 (1995), pp. 273–297.

[30] Thomas Cover and Peter Hart. “Nearest neighbor pattern classification”. In: *IEEE Transactions on Information Theory* 13.1 (1967), pp. 21–27.

[31] M. Dhingra et al. “Edge2Analysis: A Novel AIoT Platform for Atrial Fibrillation Recognition and Detection”. In: *IEEE Internet Things J.* 9.12 (2022).

[32] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley, 2006.

[33] Shabnam Sadeghi Esfahlani et al. “Machine Learning role in clinical decision-making: Neuro-rehabilitation video game”. In: *Expert Systems with Applications* 201 (2022), p. 117165.

[34] Andre Esteva et al. “A guide to deep learning in healthcare”. In: *Nature Medicine* 25 (2019), pp. 24–29.

[35] European Parliament and Council. *Regulation (EU) 2016/679 (General Data Protection Regulation)*. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[36] Y. Faye et al. “Resource-Efficient CNN Based Atrial Fibrillation Detection Using P-wave and RR Interval Features for Edge-AI Cardiac Health Monitoring Devices”. In: *Sensors* 22.8 (2022).

[37] Rahatara Ferdousi, M. Anwar Hossain, and Abdulmotaleb El Saddik. “Early-Stage Risk Prediction of Non-Communicable Disease Using Machine Learning in Health CPS”. In: *IEEE Access* 9 (2021), pp. 96823–96837.

[38] Quentin Fleury et al. “A deep learning modular ECG approach for cardiologist assisted adjudication of atrial fibrillation and atrial flutter episodes”. In: *Heart Rhythm O2* 5.12 (2024), pp. 862–872. ISSN: 2666-5018. DOI: `https://doi.org/10.1016/j.hroo.2024.09.007`.

[39] Yoav Freund and Robert E. Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting”. In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139.

[40] Jerome H. Friedman. “Greedy function approximation: A gradient boosting machine”. In: *Annals of Statistics* 29.5 (2001), pp. 1189–1232.

[41] Ala Al-Fuqaha et al. “Internet of Things: A survey on enabling technologies”. In: *IEEE Communications Surveys & Tutorials* (2015).

[42] Cédric Gilon et al. “IRIDIA-AF, a large paroxysmal atrial fibrillation long-term electrocardiogram monitoring database”. In: *Scientific data* 10.1 (2023), p. 714.

[43] Martin Gjoreski et al. “Monitoring stress with a wrist device using context”. In: *Journal of Biomedical Informatics* 73 (2017), pp. 159–170.

[44] María Goñi et al. “Smartphone-Based Digital Biomarkers for Parkinson’s Disease in a Remotely-Administered Setting”. In: *IEEE Access* 10 (2022), pp. 28361–28384.

[45] Jayavardhana Gubbi et al. “Internet of Things (IoT): A vision, architectural elements, and future directions”. In: *Future Generation Computer Systems* 29.7 (2013), pp. 1645–1660.

[46] Divya Gupta, M. P. S. Bhatia, and Akshi Kumar. “Resolving Data Overload and Latency Issues in Multivariate Time-Series IoMT Data for Mental Health Monitoring”. In: *IEEE Sensors Journal* 21.22 (2021), pp. 25421–25428.

[47] Pegah Hafiz et al. “Wearable Computing Technology for Assessment of Cognitive Functioning of Bipolar Patients and Healthy Controls”. In: 4.4 (2020). DOI: 10.1145/3432219. URL: https://doi.org/10.1145/3432219.

[48] J. Han, L. Shi, and X. Zhang. “A novel deep neural network for detection of Atrial Fibrillation using spatial and frequency features”. In: *J. Biomed. Inform.* 112 (2021).

[49] David W. Hosmer, Stanley Lemeshow, and Rodney X. Sturdivant. *Applied Logistic Regression*. Wiley, 2013.

[50] SM Riazul Islam et al. “The Internet of Things for Health Care: A Comprehensive Survey”. In: *IEEE Access* 3 (2015), pp. 678–708.

[51] A. Jerritta et al. “Identifying Heart Arrhythmias Through Multi-Level Algorithmic ECG Signal Analysis”. In: *IEEE Access* 9 (2021).

[52] Dawoon Jung et al. “Classifying the Risk of Cognitive Impairment Using Sequential Gait Characteristics and Long Short-Term Memory Networks”. In: *IEEE Journal of Biomedical and Health Informatics* 25.10 (2021), pp. 4029–4040.

[53] V Kalidas and L Tamil. “Detection of atrial fibrillation using RR interval-based entropy features and Random Forest”. In: *Biomedical Signal Processing and Control* 34 (2017), pp. 33–40.

[54] Vignesh Kalidas and Lakshman S. Tamil. “Detection of atrial fibrillation using discrete-state Markov models and Random Forests”. In: *Computers in Biology and Medicine* 113 (2019), p. 103386. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2019.103386.

[55] Jung-Yoon Kim et al. “Flat-Feet Prediction Based on a Designed Wearable Sensing Shoe and a PCA-Based Deep Neural Network Model”. In: *IEEE Access* 8 (2020), pp. 199070–199080.

[56] Barbara Kitchenham. “Procedures for Performing Systematic Reviews”. In: *Keele, UK, Keele Univ.* 33 (Aug. 2004).

[57] B. Król-Józaga. “Atrial fibrillation detection using convolutional neural networks on different ECG representations”. In: *Biomed. Signal Process. Control* 71 (2022).

[58] Martin Kropf, Dieter Hayn, and Günter Schreier. “ECG classification based on time and frequency domain features using random forests”. In: *2017 Computing in Cardiology (CinC)*. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.168-168.

[59] H Kung and et al. “Resource-efficient AF detection on wearable ECG using lightweight classifiers”. In: *IEEE EMBC*. 2020.

[60] Bo-Han Kung et al. “An Efficient ECG Classification System Using Resource-Saving Architecture and Random Forest”. In: *IEEE Journal of Biomedical and Health Informatics* 25.6 (2021), pp. 1904–1914. DOI: 10.1109/JBHI.2020.3035191.

[61] Kaya Kuru et al. “Treatment of Nocturnal Enuresis Using Miniaturised Smart Mechatronics With Artificial Intelligence”. In: *IEEE Journal of Translational Engineering in Health and Medicine* 12 (2024), pp. 204–214.

[62] Ting-Ying Li et al. “A Fast and Low-Cost Repetitive Movement Pattern Indicator for Massive Dementia Screening”. In: *IEEE Transactions on Automation Science and Engineering* 17.2 (2020), pp. 771–783.

[63] Z. Li et al. “Robust ECG Signal Classification for Detection of Atrial Fibrillation Using a Novel Neural Network”. In: *Comput. Biol. Med.* 148 (2022).

[64] C.-T. Lin et al. “An Intelligent Telecardiology System Using a Wearable and Wireless ECG to Detect Atrial Fibrillation”. In: *IEEE Trans. Inf. Technol. Biomed.* 14.3 (2010), pp. 726–732.

[65] D. Lin et al. “An interpretable electrocardiogram-based model for predicting arrhythmia with clinical rules”. In: *Sci. Rep.* 12 (2022).

[66] Y. Liu, J. Zhao, and M. Xu. “A real-time atrial fibrillation detection algorithm using EEMD and XGBoost from single-lead ECG”. In: *Computers in Biology and Medicine* 140 (2022), p. 105094. DOI: 10.1016/j.compbiomed.2021.105094.

[67] R Mahajan and et al. “Design of an optimized ECG-based AF detection system using feature selection and ensemble classifiers”. In: *Journal of Medical Systems* 43.6 (2019), pp. 1–12.

[68] Ruhi Mahajan et al. “Cardiac rhythm classification from a short single lead ECG recording via random forest”. In: *2017 Computing in Cardiology (CinC)*. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.179-403.

[69] V. Manogaran et al. “Real-Time Personalized Atrial Fibrillation Prediction on Multi-Core Wearable Sensors”. In: *Comput. Biol. Med.* 148 (2022).

[70] Anum Masood et al. “Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images”. In: *Journal of Biomedical Informatics* 79 (2018), pp. 117–128.

[71] Brendan McMahan et al. “Communication-efficient learning of deep networks from decentralized data”. In: *AISTATS* (2017).

[72] Peter Mell and Timothy Grance. *The NIST Definition of Cloud Computing*. Tech. rep. Special Publication 800-145. National Institute of Standards and Technology, 2011.

[73] B. Misra et al. “Atrial Fibrillation Detection by Multi-Lead ECG Processing at the Edge”. In: *Microprocess. Microsyst.* 93 (2022).

[74] Huma Mughal et al. “Parkinson’s Disease Management via Wearable Sensors: A Systematic Review”. In: *IEEE Access* 10 (2022), pp. 35219–35237.

[75] Evangelia Myrovali et al. “Identifying patients with paroxysmal atrial fibrillation from sinus rhythm ECG using random forests”. In: *Expert Systems with Applications* 213 (2023), p. 118948. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2022.118948.

[76] Himadri Neog, Prayakhi Emee Dutta, and Nabajyoti Medhi. “Health condition prediction and covid risk detection using healthcare 4.0 techniques”. In: *Smart Health* 26 (2022), p. 100322.

[77] Pai Chet Ng et al. "Epidemic Exposure Tracking With Wearables: A Machine Learning Approach to Contact Tracing". In: *IEEE Access* 10 (2022), pp. 14134–14148.

[78] H. Nguyen and H. Tran. "Stacking segment-based CNN with SVM for recognition of atrial fibrillation". In: *Proc. IEEE Int. Conf. Biomed. Eng.* 2021.

[79] Alexandros Pantelopoulos and Nikolaos Bourbakis. "A survey on wearable sensor-based systems for health monitoring". In: *IEEE Transactions on Systems, Man, and Cybernetics* (2010).

[80] Banu Priya Prathaban, Ramachandran Balasubramanian, and R. Kalpana. "A Wearable ForeSeiz Headband for Forecasting Real-Time Epileptic Seizures". In: *IEEE Sensors Journal* 21.23 (2021), pp. 26892–26901.

[81] John Prince, Fernando Andreotti, and Maarten De Vos. "Multi-Source Ensemble Learning for the Remote Prediction of Parkinson's Disease in the Presence of Source-Wise Missing Data". In: *IEEE Transactions on Biomedical Engineering* 66.5 (2019), pp. 1402–1411.

[82] Liudmila Prokhorenkova et al. "CatBoost: unbiased boosting with categorical features". In: *Advances in Neural Information Processing Systems* 31 (2018).

[83] Waleed Bin Qaim et al. "Towards Energy Efficiency in the Internet of Wearable Things: A Systematic Review". In: *IEEE Access* 8 (2020), pp. 175412–175435.

[84] J. Ross Quinlan. "Induction of decision trees". In: *Machine Learning* 1.1 (1986), pp. 81–106.

[85] Md Juber Rahman et al. "Automated assessment of pulmonary patients using heart rate variability from everyday wearables". In: *Smart Health* 15 (2020), p. 100081.

[86] N. Rahul and A. Sharma. "Artificial intelligence-based approach for atrial fibrillation detection using time-frequency representation of ECG". In: *Comput. Methods Programs Biomed.* 223 (2022).

[87] Ramin Ramazi et al. "Predicting progression patterns of type 2 diabetes using multi-sensor measurements". In: *Smart Health* 21 (2021), p. 100206.

[88] Govindaraj Ramkumar et al. “IoT-based patient monitoring system for predicting heart disease using deep learning”. In: *Measurement* 218 (2023), p. 113235.

[89] Elham Rastegari and Hesham Ali. “A bag-of-words feature engineering approach for assessing health conditions using accelerometer data”. In: *Smart Health* 16 (2020), p. 100116.

[90] Daniele Ravì et al. “Deep Learning for Health Informatics”. In: *IEEE Journal of Biomedical and Health Informatics* 21.1 (2017), pp. 4–21.

[91] Daniele Ravì et al. “Deep learning for health informatics”. In: *IEEE Journal of Biomedical and Health Informatics* 21.1 (2017), pp. 4–21.

[92] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning representations by back-propagating errors”. In: *Nature* 323 (1986), pp. 533–536.

[93] Susan Sabra, Khalid Mahmood Malik, and Mazen Alobaidi. “Prediction of venous thromboembolism using semantic and sentiment analyses of clinical narratives”. In: *Computers in Biology and Medicine* 94 (2018), pp. 1–10.

[94] Shehjar Sadhu et al. “Towards a telehealth infrastructure supported by machine learning on edge/fog for Parkinson’s movement screening”. In: *Smart Health* 26 (2022), p. 100351.

[95] Simanta Shekhar Sarmah. “An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System Using Deep Learning Modified Neural Network”. In: *IEEE Access* 8 (2020), pp. 135784–135797.

[96] Atifa Sarwar et al. “PainRhythms: Machine learning prediction of chronic pain from circadian dysregulation using actigraph data — a preliminary study”. In: *Smart Health* 26 (2022), p. 100344.

[97] Breawn Schoun et al. “Non-contact tidal volume measurement through thin medium thermal imaging”. In: *Smart Health* 9-10 (2018), pp. 37–49.

[98] Klaus Schwab. *The Fourth Industrial Revolution*. World Economic Forum, 2016.

[99] M. Sekulić and D. Vlahović. "Automated detection of arrhythmias using a novel interpretable feature set". In: *Comput. Methods Programs Biomed.* 200 (2021).

[100] Qi Shen et al. "Multitask Residual Shrinkage Convolutional Neural Network for Sleep Apnea Detection Based on Wearable Bracelet Photoplethysmography". In: *IEEE Internet of Things Journal* 9.24 (2022), pp. 25207–25222.

[101] Weisong Shi et al. "Edge computing: Vision and challenges". In: *IEEE Internet of Things Journal* (2016).

[102] S. Shukla et al. "A Smart Wearable for Real-Time Cardiac Disease Detection Using Beat-by-Beat ECG Signal Analysis with an Edge Computing AI Classifier". In: *Biomed. Signal Process. Control* 78 (2022).

[103] Md. Nazmul Islam Shuzan et al. "A Novel Non-Invasive Estimation of Respiration Rate From Motion Corrupted Photoplethysmograph Signal Using Machine Learning Model". In: *IEEE Access* 9 (2021), pp. 96775–96790.

[104] Ashima Singh et al. "eDiaPredict: An Ensemble-based Framework for Diabetes Prediction". In: 17.2s (2021). ISSN: 1551-6857. DOI: 10.1145/3415155. URL: https://doi.org/10.1145/3415155.

[105] Digvijay Singh et al. "Monitoring and alerting the physicians related to trauma cases using behavioural DL models". In: *Measurement: Sensors* 29 (2023), p. 100890. ISSN: 2665-9174. DOI: https://doi.org/10.1016/j.measen.2023.100890. URL: https://www.sciencedirect.com/science/article/pii/S266591742300226X.

[106] Jagmeet P Singh et al. "Short-term prediction of atrial fibrillation from ambulatory monitoring ECG using a deep neural network". In: *European Heart Journal - Digital Health* 3.2 (Apr. 2022), pp. 208–217. DOI: 10.1093/ehjdh/ztac014.

[107] H. Sinha et al. "Atrial Fibrillation Detection by Means of Edge Computing on Wearable Device: A Feasibility Assessment". In: *Sensors* 21.9 (2021).

[108] Aditi Site, Jari Nurmi, and Elena Simona Lohan. "Systematic Review on Machine-Learning Algorithms Used in Wearable-Based eHealth Data Analysis". In: *IEEE Access* 9 (2021), pp. 112221–112235.

[109] Justyna Skibinska et al. “COVID-19 Diagnosis at Early Stage Based on Smartwatches and Machine Learning Techniques”. In: *IEEE Access* 9 (2021), pp. 119476–119491.

[110] J. Smith et al. “Spontaneous Brain Activity and EEG Microstates: A Novel EEG-fMRI Analysis Approach”. In: *Neuroimage* 207 (2020).

[111] Rahul Soangra et al. “Classifying Toe Walking Gait Patterns Among Children Diagnosed With Idiopathic Toe Walking Using Wearable Sensors and Machine Learning Algorithms”. In: *IEEE Access* 10 (2022), pp. 77054–77067.

[112] Dionisije Sopic et al. “Real-Time Event-Driven Classification Technique for Early Detection and Prevention of Myocardial Infarction on Wearable Systems”. In: *IEEE Transactions on Biomedical Circuits and Systems* 12.5 (2018), pp. 982–992.

[113] Sergio Staab et al. “Automated documentation of almost identical movements in the context of dementia diagnostics”. In: *Smart Health* 26 (2022), p. 100333.

[114] Fei Teng et al. “PDGes: An Interpretable Detection Model for Parkinson’s Disease Using Smartphones”. In: 19.4 (2023). ISSN: 1550-4859. DOI: `10.1145/3585314`. URL: `https://doi.org/10.1145/3585314`.

[115] Connor Terry, Qi Shi, and Xuefeng Wang. “Hybrid CNN–BiLSTM model for detection of atrial fibrillation using short single-lead ECG and PPG signals”. In: *Computer Methods and Programs in Biomedicine* 226 (2022), p. 107118. DOI: `10.1016/j.cmpb.2022.107118`.

[116] U.S. Congress. *Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. U.S. Department of Health & Human Services. 1996. URL: `https://www.hhs.gov/hipaa/index.html`.

[117] A. N. Uwaechia and D. A. Ramli. “A Comprehensive Survey on ECG Signals as New Biometric Modality for Human Authentication: Recent Advances and Future Challenges”. In: *IEEE Access* 9 (2021), pp. 97760–97791.

[118] X. Wang and Q. Liu. “Interpretable Independent Recurrent Neural Network for Atrial Fibrillation Detection”. In: *Expert Syst. Appl.* 191 (2022).

[119] Z. Wang, H. Li, and X. Zhou. "Symbolic Dynamics-Based RR Interval Analysis for Paroxysmal Atrial Fibrillation Detection". In: *Sensors* 23.2 (2023), p. 546. DOI: 10.3390/s23020546.

[120] Shweta Ware et al. "Automatic depression screening using social interaction data on smartphones". In: *Smart Health* 26 (2022), p. 100356.

[121] Patrick Henry Winston. *Artificial intelligence*. Addison-Wesley Longman Publishing Co., Inc., 1984.

[122] C. Wu and L. Zhang. "TFDGiniXML: A novel explainable machine learning model for ECG arrhythmia classification". In: *Comput. Biol. Med.* 144 (2022).

[123] Chia-Tung Wu et al. "A Precision Health Service for Chronic Diseases: Development and Cohort Study Using Wearable Device, Machine Learning, and Deep Learning". In: *IEEE Journal of Translational Engineering in Health and Medicine* 10 (2022).

[124] Xiaodan Wu et al. "Extracting deep features from short ECG signals for early atrial fibrillation detection". In: *Artificial Intelligence in Medicine* 109 (2020), p. 101896. ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2020.101896.

[125] Yuan Wu et al. "Ubi-Asthma: Toward Ubiquitous Asthma Detection Using the Smartwatch". In: *IEEE Internet of Things Journal* 10.13 (2023), pp. 11576–11587. DOI: 10.1109/JIOT.2023.3243188.

[126] Yifan Xia et al. "A multi-scale dilated convolutional network with auxiliary tasks for atrial fibrillation detection from wearable ECG signals". In: *Biomedical Signal Processing and Control* 83 (2023), p. 104676. DOI: 10.1016/j.bspc.2023.104676.

[127] Ming–Xia Xiao et al. "Toe PPG sample extension for supervised machine learning approaches to simultaneously predict type 2 diabetes and peripheral neuropathy". In: *Biomedical Signal Processing and Control* 71 (2022), p. 103236.

[128] Zhanglu Yan, Jun Zhou, and Weng-Fai Wong. "EEG classification with spiking neural network: Smaller, better, more energy efficient". In: *Smart Health* 24 (2022), p. 100261.

[129] Fan Yang et al. “Improving pain management in patients with sickle cell disease from physiological measures using machine learning techniques”. In: *Smart Health* 7-8 (2018), pp. 48–59.

[130] Hui Ye, Bo Wang, and Li Chen. “A Hybrid Deep Learning Model Based on CNN, BiLSTM, and XGBoost for Multiclass Arrhythmia Classification”. In: *Diagnostics* 15.7 (2023), p. 865. DOI: 10.3390/diagnostics15070865.

[131] Hongxu Yin and Niraj K. Jha. “A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles”. In: *IEEE Transactions on Multi-Scale Computing Systems* 3 (2017), pp. 228–241.

[132] Xiang Yu et al. “ResNet-SCDA-50 for Breast Abnormality Classification”. In: 18.1 (2021). ISSN: 1545-5963. DOI: 10.1109/TCBB.2020.2986544. URL: https://doi.org/10.1109/TCBB.2020.2986544.

[133] Chaoqun Yue et al. “Automatic depression prediction using Internet traffic characteristics on smartphones”. In: *Smart Health* 18 (2020), p. 100137.

[134] Morteza Zabihi et al. “Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier”. In: *2017 Computing in Cardiology (CinC)*. 2017, pp. 1–4. DOI: 10.22489/CinC.2017.069-336.

[135] Andrea Zanella et al. “Internet of Things for smart cities”. In: *IEEE Internet of Things Journal* (2014).

[136] J. Zhang et al. “Randomized Attention and Dual-Path System for Electrocardiogram Identity Recognition”. In: *IEEE Trans. Instrum. Meas.* 70 (2021).

[137] Qingxue Zhang et al. “A Machine Learning-Empowered System for Long-Term Motion-Tolerant Wearable Monitoring of Blood Pressure and Heart Rate With Ear-ECG/PPG”. In: *IEEE Access* 5 (2017), pp. 10547–10561. DOI: 10.1109/ACCESS.2017.2707472.

[138] Zhichao Zhang et al. “A sensor-based wrist pulse signal processing and lung cancer recognition”. In: *Journal of Biomedical Informatics* 79 (2018), pp. 107–116.

[139] Shuhan Zhong et al. “DYPA: A Machine Learning Dyslexia Prescreening Mobile Application for Chinese Children”. In: 7.3 (2023). DOI: `10.1145/3610908`. URL: `https://doi.org/10.1145/3610908`.

[140] M. Zihlif, Y. Himeur, and F. Bensaali. “A Novel Deep Arrhythmia-Diagnosis Network for Atrial Fibrillation Classification Using ECG Signals”. In: *Comput. Methods Programs Biomed.* 229 (2023).

List of Publications

[9] Aversano, L., Iammarino, M., Mancino, I., Montano, D. (2023) A systematic Review on the use of Artificial Intelligence Approaches and Smart Health Devices In PeerJ Computer Science

[10] Aversano, L., Iammarino, M., Mancino, I., Verdone, C. (2025) From Raw ECG to Classification: a Lightweight Approach to Paroxysmal AF Detection In 2025 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)

Aversano, L., Iammarino, M., Mancino, I., Verdone, C., Madau, A., Montano, D. (2025) Mining Community Health: Association Rules for the Design of Proximity Care Models In 2025 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)

[8] Aversano, L., Merengo, A., Mancino, I., Verdone, C. (2025) Comparative Analysis of Machine Learning and Deep Learning Models for Atrial Fibrillation Detection from Long-Term ECG In IEEE Open Journal of the Computer Society