# Leveraging Machine Learning for Disease Diagnoses based on Wearable Devices: A Survey

Zhihan Jiang, *Graduate Student Member, IEEE*, Vera van Zoest, Weipeng Deng, Edith C.H. Ngai*, *Senior Member, IEEE*, Jiangchuan Liu, *Fellow, IEEE*

***Abstract*—Many countries around the world are facing a shortage of healthcare resources, especially during the post-epidemic era, leading to a dramatic increase in the need for self-detection and self-management of diseases. The popularity of smart wearable devices, such as smartwatches, and the development of machine learning bring new opportunities for the early detection and management of various prevalent diseases, such as cardiovascular diseases, Parkinson's disease, and diabetes. In this survey, we comprehensively review the articles related to specific diseases or health issues based on small wearable devices and machine learning. More specifically, we first present an overview of the articles selected and classify them according to their targeted diseases. Then, we summarize their objectives, wearable device and sensor data, machine learning techniques, and wearing locations. Based on the literature review, we discuss the challenges and propose future directions from the perspectives of privacy concerns, security concerns, transmission latency and reliability, energy consumption, multi-modality, multi-sensor, multi-devices, evaluation metrics, explainability, generalization and personalization, social influence, and human factors, aiming to inspire researchers in this field.**

***Index Terms*—Wearable devices, smart watches, physical health, machine learning, disease diagnoses.**

## I. Introduction

VARIOUS chronic diseases, such as cardiovascular diseases, Parkinson's disease, and diabetes, can be early detected and managed by continuous and real-time monitoring of patients' vital signs [1]. Numerous people around the world are suffering from these diseases, which have adversely influenced patients' well-being and imposed heavy burdens on healthcare providers. In particular, the worldwide pandemic, coronavirus disease 2019 (COVID-19), has brought great challenges to clinical care [2]. Medical institutions have to deal with more patients with less staff, and healthcare resources are quite limited during pandemic periods. The effect can last for a long time in the post-epidemic era, especially in many developing countries [3], [4]. Upon this basis, it is important to facilitate the self-detection and self-management of chronic diseases.

*Corresponding author: Edith C.H. Ngai.*

Z. Jiang, W. Deng, and E. Ngai are with IoT Lab, Department of Electrical and Electronic Engineering, Faculty of Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China (e-mail: zhjiang@connect.hku.hk, dengf330@connect.hku.hk, chngai@eee.hku.hk).

V. van Zoest is with the Department of Information Technology, Uppsala University, Uppsala, Sweden (e-mail: vera.van.zoest@it.uu.se).

J. Liu is currently a professor in the Department of Computing at Simon Fraser University, Canada (e-mail: jcliu@cs.sfu.ca).



In recent years, wearable devices (e.g., smart watches, smart rings, and smart glasses) are getting more and more popular all around the world. Multiple sensors, such as accelerometers, gyroscopes, and heart rate sensors, have been integrated into small wearable devices, making it feasible to continuously monitor various physiological data of users in real time. By analyzing these health-related data, many chronic diseases can be detected at an early stage, and patients' conditions can be assessed. In tradition, this kind of analysis should be conducted by educated specialists, which is costly and inconvenient. Fortunately, the advance in artificial intelligence, especially machine learning, has significantly enhanced the effectiveness and efficiency of healthcare applications for wearable devices, providing personalized healthcare guidance for users without prior knowledge.

There have been numerous works on healthcare based on wearable devices and machine learning. Many of them focus on human activity recognition to record users' physical activities and calories burned, encouraging a healthier lifestyle. Vital signs, such as pulse rate and body temperature, are sensed and visualized for reference. Specifically, in this survey, we focus on the papers that are related to specific diseases or health issues rather than generally monitoring or visualization of physiological information, which can directly benefit the early detection and management of chronic diseases for patients.

Previous surveys of wearable device applications in healthcare that related to specific diseases or health issues based on artificial intelligence usually only focus on a single disease. For example, Koumpouros *et al.* [5] reviewed the literature on the detection of autism spectrum disorders (ASD) using wearable and mobile devices. Channa *et al.* [6] discussed works on COVID-19 symptom diagnoses using wearable devices. Pereira *et al.* [7] reviewed works on atrial fibrillation detection based on Photoplethysmography (PPG). Furthermore, previous surveys usually include a wide range of wearable devices, while some large portable devices (e.g., backpacks) and professional instruments (e.g., blood pressure meters) are not pervasive enough. In this survey, we comprehensively reviewed the research articles for specific diseases or health issues based on more prevalent small wearable devices (e.g., smart watches) and machine learning, and present an overview of the articles selected. We classify them according to their corresponding disease types, summarize the disease categories that have received the most attention (i.e., Parkinson's Disease, Cardiovascular Diseases, Sleep Issues, Diabetes, Respiratory and Pulmonary Diseases), and also discuss those in the minority (i.e., Epileptic Seizure, Frailty Syndrome, Sarcopenia, and Abdominal Aortic Aneurysm). For each selected study, we present their objectives, wearable devices and sensor data, and machine learning methods used. It is a promising area undergoing an increase currently and in the coming years. Furthermore, based on the review, we discussed the challenges and future directions, showing that although current works have achieved great success, there are still many issues to be addressed. New research opportunities can be found by exploring these issues.

The remainder of this paper is organized as follows. Section II introduces the method used for paper searching and the criteria for paper selection. We also presented a brief summary of the related papers in this section. Then, in Section III, we generally introduce the machine learning algorithms frequently used in the reviewed papers. Subsequently, in Section IV, we classify the reviewed papers according to their corresponding diseases and elaborate on their objectives, related wearable devices and sensor data, and machine learning algorithms. Finally, we discuss the challenges and future directions in Section V and conclude our survey in Section VI.

## II. Methods

For the literature review, we used the Web of Science Core Collection database. The initial literature search was conducted on the 15th of December 2022 using the search string (”smart watches” OR ”smart wearables” OR ”smart watch” OR ”wearable devices”) AND (”health” OR ”disease” OR ”healthcare”) AND (”machine learning” OR ”deep learning”). Only research articles and proceeding papers (including early access) written in English were included. This resulted in 602 papers. As shown in Fig. 1, there is no result before 2015, and the number of papers drastically increases after 2018. It should be noted that some papers accepted or published in 2022 may not have been included in the database. Thus, the number of papers is slightly fewer than that in 2021. According to the Web of Science categories (TABLE I), most of the papers belong to Engineering Electrical Electronic and Computer Science Information Systems. TABLE II summarizes the distribution across countries or regions, showing that The United States, China, and India have published the most papers.

In general, identifying the types and intensities of users' physical activities automatically is a fundamental function of many wristbands. Among these papers, many of them studied human activity recognition [8]–[13] since the amount of physical activities is an important health indicator [14]. There are also many works on health monitoring systems. Various sensors embedded on wearable devices, such as accelerometer, gyroscope, and PPG sensor, were leveraged to collect users' physiological signals to proactively monitor their health status to avoid serious diseases [15]–[17] and facilitate rehabilitation of patients [18], [19]. Besides, some researchers proposed to estimate stress [20]–[22], detect mood disorders [23], [24], predict depression [25]–[27], etc., to measure users' mental health using wearable devices [28]–[30]. In this work, we focus on *specific physical health issues*, and the definition of wearable devices is limited to small devices that are not permanently fixed to the body. More specifically, we further selected the papers based on the following criteria.

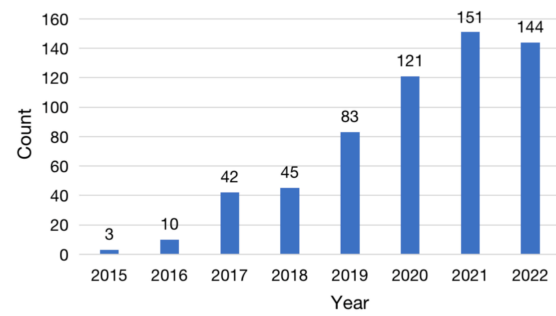


Fig. 1. The number of relevant papers across years.

TABLE I
Web of Science categories with more than 10 papers. One paper can be classified into more than one category.

| Web of Science Categories | Count | % |
|---|---|---|
| Engineering Electrical Electronic | 206 | 34.2% |
| Computer Science Information Systems | 149 | 24.8% |
| Telecommunications | 97 | 16.1% |
| Computer Science Theory Methods | 95 | 15.8% |
| Engineering Biomedical | 93 | 15.4% |
| Computer Science Artificial Intelligence | 88 | 14.6% |
| Computer Science Interdisciplinary Applications | 82 | 13.6% |
| Medical Informatics | 73 | 12.1% |
| Instruments Instrumentation | 68 | 11.3% |
| Health Care Sciences Services | 48 | 8.0% |
| Chemistry Analytical | 43 | 7.1% |
| Physics Applied | 31 | 5.2% |
| Mathematical Computational Biology | 29 | 4.8% |
| Computer Science Software Engineering | 23 | 3.8% |
| Computer Science Cybernetics | 21 | 3.5% |
| Computer Science Hardware Architecture | 21 | 3.5% |
| Automation Control Systems | 18 | 3.0% |
| Engineering Multidisciplinary | 16 | 2.7% |
| Materials Science Multidisciplinary | 11 | 1.8% |
| Multidisciplinary Sciences | 11 | 1.8% |
| Neurosciences | 11 | 1.8% |
| Public Environmental Occupational Health | 11 | 1.8% |

TABLE II
Countries with more than 10 papers. One paper can be classified into more than one country.

| Countries | Count | % |
|---|---|---|
| The United States | 144 | 23.9% |
| China | 130 | 21.6% |
| India | 72 | 12.0% |
| Italy | 54 | 9.0% |
| England | 50 | 8.3% |
| South Korea | 40 | 6.6% |
| Spain | 27 | 4.5% |
| Australia | 26 | 4.3% |
| Canada | 24 | 4.0% |
| Japan | 23 | 3.8% |
| Saudi Arabia | 21 | 3.5% |
| Singapore | 20 | 3.3% |
| Switzerland | 19 | 3.2% |
| Germany | 16 | 2.7% |
| Greece | 13 | 2.2% |
| Turkey | 13 | 2.2% |
| United Arab Emirates | 12 | 2.0% |
| Finland | 11 | 1.8% |
| Pakistan | 11 | 1.8% |

**Criterion 1. Three Key Components**. Only those papers directly related to smart wearable devices, machine learning, and physical health were included.

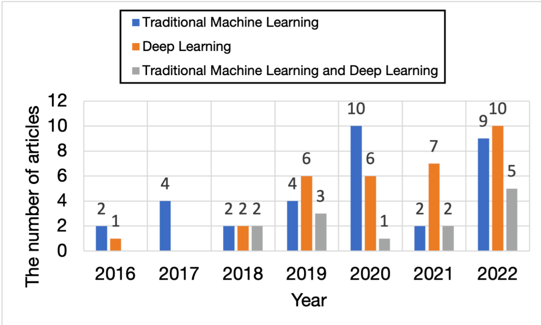


Fig. 2. The distribution of articles across years with regard to traditional machine learning, deep learning, and both traditional machine learning and deep learning.

**Criterion 2. The Coverage of Wearable Devices**. Wearable devices in this paper refer to small devices that are not permanently fixed to the body and have a function as wearable. We excluded large portable devices such as backpacks used to carry around and larger unwearable devices.

**Criterion 3. The Coverage of Health Issues**. The papers on system design or assisting in sports and training without targeting any specific disease or health issue were excluded. The papers on mental health were also excluded.

Based on the above criteria, we selected 78 articles and further classified and analyzed them in terms of the targeted diseases, the machine learning methods used, and their main conclusions. Fig. 2 shows the distribution of articles across years with regard to different categories of machine learning techniques (i.e., traditional machine learning, deep learning, and both traditional machine learning and deep learning). We can find that the number of articles using deep learning techniques is increasing. This can result from the increasing popularity of deep learning techniques. Meanwhile, there are still many articles using traditional machine learning methods, since traditional methods are usually more effective than deep learning methods when the amount of data is limited, especially in the healthcare domain.

## III. Machine learning techniques

The popularity of the use of machine learning (ML) has grown over the past decades. ML allows computers to make use of statistical techniques to learn and improve tasks after seeing data and gaining experience. Rather than being specifically programmed for certain actions, the systems can train themselves by recognizing patterns in the data. ML can be considered a subfield of Artificial Intelligence (AI).

AI focuses on the simulation of human intelligence in machines to perform tasks that normally require human intelligence [31]. In general, it can be divided into the following two categories [32]:

- **Narrow AI**, also known as Weak AI, is designed to perform specific tasks, such as face recognition or recommendation systems. Most of the existing AI systems belong to this category.
- **General AI**, also known as Strong AI, refers to the systems that have generalized human cognitive abilities. They have the ability to understand, learn, adapt, and implement knowledge in different domains.

ML is a significant subset of AI, where machines learn and improve from experience or data [33]. ML models learn to identify patterns based on the provided data (i.e., training data) to make predictions with new data. This ability to learn from data makes ML apart from traditional rule-based approaches [34]. There are some basic key concepts in ML [35]:

- **Features and Labels**. In a dataset for ML, features are input variables, and labels are the corresponding outputs that the model aims to predict.
- **Training and Testing**: To build an ML model, the dataset is typically split into training and testing sets. The training set is used to train the model, and the testing set is used to evaluate the performance of the trained model.
- **Overfitting and Underfitting**: Overfitting occurs when a model performs very well on the training data but fails to generalize to testing data. Underfitting means a machine learning model is unable to capture the underlying patterns of the data, failing to achieve satisfactory performance on both training and testing data.
- **Evaluation Metrics**: Various metrics are used to evaluate the ML model performance for different tasks. In classification tasks, metrics such as accuracy, precision, recall, and F1 score, are usually used. In regression tasks, metrics such as mean absolute error, mean square error, root mean squared error, and $R^2$ score, are usually used.

ML algorithms have been used in a wide variety of applications, such as email filtering and computer vision [36]. It has also achieved great success in disease diagnoses [37]. Within ML, we distinguish between traditional machine learning and deep learning approaches. Compared with traditional machine learning, deep learning usually requires a larger amount of data and more computing power for training. Whereas traditional machine learning usually requires more domain expertise and effort in feature engineering, deep learning uses artificial neural networks to mimic the learning process of humans and requires less human intervention. With sufficient data and computing power, deep learning can learn complex correlations and has been applied to various domains. However, traditional machine learning approaches are easier to understand and explain. Many deep learning models are "black box" models, meaning that we do not know how the models make decisions.

Furthermore, the explainability of machine learning models is of vital importance in health-related studies. Firstly, increased model explainability makes the decision-making process more transparent, increasing the soundness and reliability of the decisions. Secondly, increased explainability supports the identification of risk factors in medical studies, which can greatly benefit the improvement of people's health.

Among the 78 articles selected in this survey, there are 33 articles using traditional machine learning approaches, 32 articles using deep learning approaches, and 13 articles using both traditional machine learning and deep learning approaches, as shown in Fig. 3. However, only one article [38] among them considered the explainability of models.

In this section, we introduce the popular traditional machine

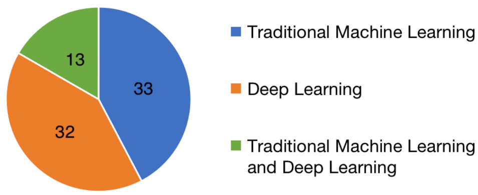


Fig. 3. The distribution of techniques used in the selected papers.

learning and deep learning methods used in disease diagnoses and further discuss the explainability of machine learning.

### A. Traditional Machine Learning

For traditional machine learning, we discussed the following algorithms frequently used with data collected from smart wearable devices in disease diagnoses: Support Vector Machine, Discriminant Analysis, Naïve Bayes, K-Nearest Neighbors, tree-based methods, logistic regression, k-Means, and Hidden Markov Model.

The Support Vector Machine (SVM) algorithm optimizes the separation boundary (hyperplane) in n-dimensional space between feature values of different classes. The separation boundary can be a single straight line, in Linear SVM, or a non-linear line in Non-Linear SVM. SVM can be used for both classification and regression tasks. The training process involves solving an optimization problem as follows:

**Problem**: (Support Vector Machine)

$$\begin{aligned} &minimize \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} max(0, 1 - y_i(w^T x_i + b)) \\ &subject\ to \quad \forall i, y_i(w^T x_i + b) \geq 1, \end{aligned} \tag{1}$$

where $w$ represents the weight vector; $n$ is the number of training samples; $b$ is the bias; $x_i$ and $y_i$ are the input data vector and corresponding label, and $C$ is a regularization parameter. The time complexity is typically between $O(dn^2)$ and $O(dn^3)$, where $d$ is the number of features, depending on the implementation and kernel functions used.

In Discriminant Analysis (DA), we use a discriminant function to classify groups based on a set of numerical variables. In two-group discriminant analysis, i.e., when we only have two categories, this boils down to a linear regression function with an outcome variable that can take values zero and one depending on the category (i.e., a dummy variable). In multiple-group discriminant analysis, we have several discriminant functions to narrow down the outcome variable to a specific class. Linear Discriminant Analysis (LDA) finds a linear combination of features that maximizes the separation between classes, while Quadratic Discriminant Analysis (QDA) extends LDA by allowing quadratic decision boundaries. Training LDA and QDA involves computing covariance matrices and solving linear equations. The time complexity for training is approximately $O(nd^2)$ for LDA and $O(nd^3)$ for QDA.

Naive Bayes (NB) is a probabilistic classifier based on Bayes' theorem. The probability of a certain class depends on the probability of a set of features occurring in a class in a training data set. NB assumes, however, independence between the features. There are many different variants of NB, which may be suitable depending on the type of data at hand. The posterior probabilities of each class given the input features using Bayes' theorem are calculated as follows:

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}, \tag{2}$$

where $P(C_i|x)$ is the posterior probability of class $C_i$ given the input $x$, $P(x|C_i)$ is the likelihood of $x$ given class $C_i$, $P(C_i)$ is the prior probability of class $C_i$, and $P(x)$ is the evidence probability. Training an NB model involves estimating class priors and likelihood probabilities. The time complexity is typically linear, approximately $O(nd)$.

The k-Nearest Neighbors (KNN) algorithm is used to predict the target value based on the nearest $k$ number of neighbors in the training dataset. For $k = 1$, this simply boils down to the one nearest neighbor. For $k > 1$, the classifier chooses the class that is most present in the $k$ number of neighbors in the training dataset. Similarly, KNN can be used for regression tasks, where the classifier takes the average of the k nearest neighbors. The computational complexity of KNN is influenced by the number of training samples $n$ and the number of features $d$. Building the KNN model involves indexing the training data, which has a time complexity of approximately $O(ndlog(n))$. Making predictions with KNN has a time complexity of approximately $O(dn)$.

Tree-based methods are popular in classification tasks within smart healthcare but can be used for regression tasks as well. There are many different tree-based methods, starting from a single decision tree (DT), which splits branches at different feature values to combine different combinations of feature values to match patterns in the output. The computational complexity depends on the depth of the trees and the number of training samples $n$. Constructing a decision tree typically has a time complexity of $O(ndlog(n))$, where $d$ is the number of features. Random Forests (RF) combine and average over a larger set of trees to improve the performance of the algorithm. Several extensions exist, such as bagging and gradient boosting. Random Forests train multiple decision trees in parallel, resulting in a higher computational complexity compared to individual decision trees.

For the classification of a positive vs. negative diagnosis or for estimating the probability of an event to occur, logistic regression (LR) is a traditional technique that has often been used. Although it is a regression-based technique, it can be used for classification tasks where the outcome can be classified based on a probability of occurrence. It models the relationship between the input features and the probability of belonging to a particular class using a logistic function. The algorithm estimates the parameters of the logistic regression model by maximizing the likelihood function. The logistic regression model uses the logistic function to model the probability of the positive class:

$$P(y = 1|x) = \frac{1}{1 + e^{-w^T x}}, \tag{3}$$

where $w^T x$ represents the linear combination of input features $x$ and corresponding weights $w$.

Clustering methods are used to detect patterns and similarities to form clusters. Clustering has many similarities to classification, with the biggest difference being that it is an unsupervised learning approach, and thus no output data is available in training. The use of unsupervised learning approaches is not very common in the analysis of data from smart wearable devices in healthcare. Instead, often a diagnosis is made by a doctor to create an output in the training dataset.

The most traditional example of a clustering approach is k-means clustering, in which the observations are divided into $k$ clusters. Each observation belongs to the cluster to which it is closest to the mean of that cluster, where the mean of the cluster is determined such that the within-cluster variances between the observations are minimized. The algorithm iteratively minimizes the within-cluster sum of squared distances to converge to a locally optimal solution as follows:

1) Initialize $k$ cluster centers randomly.
2) Assign each data point to the nearest center based on the distance (e.g., Euclidean distance).
3) Update the centroids by computing the mean of the assigned data points for each cluster.
4) Repeat steps 2) and 3) until convergence, which occurs when the centers no longer change significantly or a maximum number of iterations is reached.

The time complexity is approximately $O(Inkd)$, where $I$ is the number of iterations required for convergence; $n$ the number of data points, and$d$ is the number of features.

Another method for clustering is the Hidden Markov Model (HMM), a probabilistic approach based on unknown (hidden) feature values. The objective of an HMM is to learn the states of the features by observing a sequence of outcome values. HMMs are used mainly for sequential data.

### B. Deep Learning

Artificial Neural Networks (ANN) form the basis of Deep Learning, a subfield of machine learning. ANNs can be used for both classification and regression tasks and are particularly suitable for non-linear relationships (regression) and non-linear decision boundaries (classification). The input data is passed through hidden layers with different weights, as well as an activation function that allows for non-linear patterns and complex relationships in the data to be discovered. A common type of ANN is the Multilayer Perceptron (MLP), a fully connected feedforward neural network. It consists of at least three layers, including an input layer, a hidden layer, and an output layer. It utilizes back-propagation for training. Each neuron in an MLP is a basic processing unit that computes a weighted sum of its inputs, applies an activation function to introduce non-linearity, and passes the output to the next layer. The equation for the output of a neuron in an MLP can be represented as follows:

$$y = f(w_1x_1 + w_2x_2 + ... + w_nx_n + b), \tag{4}$$

where $y$ is the output; $x_1$, $x_2$,..., $x_n$ are the input values; $w_1$, $w_2$,..., $w_n$ are the corresponding weights, and $b$ is bias; $f$ is the activation function.

Another commonly used type of neural network (NN) is the Convolutional Neural Network (CNN). It is a regularized MLP. The hidden layers of CNN usually include convolutional layers, pooling layers, and fully-connected layers. In convolutional layers, a kernel is used to generate convolved features based on the input from the previous layer, and the features are transmitted to the next layer. The main equation in a convolutional layer of a CNN is the convolution operation, which applies a set of learnable filters to the input image:

$$C[i] = f[\sum(W[k] * X[i-k])], \tag{5}$$

where $C[i]$ represents the output feature map at position $i$, $f$ is the activation function, $W[k]$ is the weights of $k$-th filter, $X[i-k]$ represents the input patch with the center $i$, and $*$ denotes the convolution operation. The pooling layers are used to reduce the parameters and prevent over-fitting. In a fully-connected layer, every neuron is connected to every neuron in another layer. CNNs can recognize spatial patterns of data and have achieved great performance, especially on image data.

Recurrent Neural Network (RNN) is another class of ANN, which is usually used for sequence data such as time series data, text, and audio. The circles in RNNs mean that the output from some neurons can influence the input of these neurons. In this way, RNNs can learn the temporal dependency of input sequence data. The main equation in an RNN is the recurrent connection, which updates the hidden state at each time step:

$$h[t] = f(W_h \cdot h[t-1] + W_x \cdot x[t] + b), \tag{6}$$

where $h[t]$ is the hidden state at time $t$, $x[t]$ is the input at time $t$, $W_h$ represents the recurrent weights, $W_x$ represents the input weights, $b$ is bias, and $f$ is the activation function. When training an RNN with many layers, vanishing and exploding gradient problem often occurs. Long Short-Term Memory Networks (LSTMs) are a type of RNNs that can address this problem. LSTMs have memory blocks connected into layers. Gates, including forget gate, input gate, and output gate, are used to manage the state and output of blocks in LSTMs. Gated Recurrent Units (GRUs) are a gating mechanism in RNNs. It updates the gate and resets the gate. Compared with LSTMs, GRUs require fewer parameters and are usually faster.

Based on the above-discussed networks, there are numerous extensions, such as Bidirectional LSTM (BiLSTM), ResNet, Seq2Seq, and Generative Adversarial Network (GAN). These novel deep-learning networks usually require a large amount of data and computing resources and have achieved great success. However, healthcare study usually has limited data, and small wearable devices have limited computing power, which brings new challenges and opportunities.

### C. Explainability

The explainability of ML has gained more and more attention. However, although ML has achieved great performance in disease diagnoses, most of the current works lack the discussion on the explainability of their models. Generally, the methods for explaining the ML models can be classified into global model-agnostic and local model-agnostic [39]. The global model-agnostic methods include Partial Dependence

Plot (PDP), Permuted Feature Importance, and Global Surrogate. The local model-agnostic methods include Individual Conditional Expectation (ICE), Local Surrogate (LIME), and Shapley Additive explanations (SHAP) [40].

The Partial Dependence Plot (PDP) can present the marginal effect of one or two features on the predicted outcomes. It can indicate the correlation between the features and outcomes. PDP is easy to implement and understand. Based on the observation of PDP, the linear and monotonic relationships can be figured out. However, it is difficult to identify more complex relationships. Also, a partial dependence function can only incorporate up to two features, and the features should be independent of each other. The time complexity of calculating the PDP for a single feature is usually between $O(n)$ and $O(n^2)$, where $n$ is the number of unique values considered for that feature. Therefore, generating PDPs for multiple features may result in higher time complexity.

Permutation Feature Importance calculates the increase in the prediction error after the permutation of a feature to measure the importance of the feature. This method is also easy to understand and efficient. The permutation importance of a feature considers both its effect on the model and its interaction with other features. It provides global insights into ML models. However, this method is sensitive to the correlation between features. The importance of one feature can be shared by its correlated features, decreasing its importance. The importance of a feature is quantified as the difference between the original performance and the performance after permutation. The time complexity depends on the complexity of the underlying ML model and the number of features.

Global Surrogate is to train an interpretable model to approximate the original black box model. Conclusions can be drawn based on the interpretation of the surrogate model. This method is straightforward. Any interpretable models can be used for the global surrogate. However, the generalization of the surrogate model is challenging. Different datasets may generate different conclusions. Also, whether an interpretation of the surrogate models can be used for the original models is still controversial. The time complexity of Global Surrogate depends on the complexity of the black-box model, the chosen surrogate model algorithm, and the size of the training dataset.

Individual Conditional Expectation (ICE) plot is a PDP for individual instances. It presents the dependence of the outcome on a feature for each instance separately. Each line in an ICE plot represents one instance. By averaging the lines in an ICE plot, the PDP can be generated. Compared with PDP, it is more intuitive and can reflect heterogeneous relationships. However, it can only clearly present the effect of one feature. Similar to PDP, it also suffers from the feature correlation problem. ICE for a specific data instance represents the expected prediction outcome as a function of a chosen feature while fixing the values of other features. The time complexity depends on the complexity of the underlying machine learning model, the number of data instances, and the number of features.

The idea of Local Surrogate (LIME) is as intuitive as Global Surrogate, but it trains local surrogate models to approximate individual predictions. When interpreting an instance, a new dataset consisting of perturbed samples and the corresponding outcomes of the original model is first generated. Then, the new samples will be assigned weights according to their proximity to the instance. A weighted interpretable model will be trained on the new dataset. LIME can make explanations that are easy to understand and works for multimodal data. However, it is difficult to define the neighborhood in LIME, leading to more parameters for tuning. Also, the reliability of the explanation is significantly influenced by the data preprocessing. The time complexity depends on the complexity of the black-box model, the chosen surrogate model algorithm, and the size of the training dataset or neighborhood.

Shapley Additive explanations (SHAP) measure the feature importance and interpret the predicted outcomes of machine learning models based on game theory [40]. It computes the contribution of each player (i.e., feature) in a collaborative game (i.e., machine learning models) [40]. SHAP can explain the model in global and local ways [41]. More specifically, we denote the $j$th feature of $x_i$ by $x_{ij}$, and the SHAP value of $x_{ij}$, denoted by $f(x_{ij})$, follows the equation below:

$$y_i = E[f(X)] + \sum_{j=1}^{M} f(x_{ij}), i = 1, 2, ..., n, \tag{7}$$

where $X$ is the training dataset; $y_i$ is the label of $x_i$; $E[f(X)]$ is usually the mean of the predicted value of all samples. The SHAP values of features reflect their impacts on the model output. It can be implemented efficiently, especially for tree-based models. However, the explanation of SHAP may be unintuitive and even misleading, especially when there are correlated features.

In general, some prerequisites, such as independent features, should be satisfied when explaining machine learning models using different approaches. Otherwise, the explanation can be misleading. In disease diagnosis applications, we should be very careful when generating conclusions by interpreting machine learning models, especially the casual insights, since there might be unmeasured confounding features and correlated input features in real-world settings [42]–[44].

## IV. Applications in Disease Diagnoses

In this section, we categorize the selected articles by their targeted diseases. The prevalent disease categories include Parkinson's Disease, Cardiovascular Diseases, Sleep Issues, Diabetes, and Respiratory and Pulmonary Diseases. Alongside these five highly focused categories, there are also works related to a diverse range of other health issues, including Epileptic Seizure, Frailty Syndrome, Sarcopenia, and Abdominal Aortic Aneurysm. However, the number of articles associated with these additional diseases is comparatively lower than those within the five most popular disease categories. Therefore, we first analyze the articles pertaining to the five most prominent categories and then review the articles associated with other diseases.

To summarize, there are 19 articles related to Parkinson's Disease, 25 articles related to cardiovascular diseases, 5 articles related to sleep issues, 6 articles related to diabetes, 15 articles related to respiratory and pulmonary diseases, and 8 papers related to other diseases. The sensor data used are

TABLE III
THE MAJOR TYPES OF SENSOR DATA USED FOR DIFFERENT HEALTH ISSUES.

| **Disease** | **Sensor Data** |
|---|---|
| Parkinson's Disease | EMG (17/19), vertical ground reaction force (1/19), IMU (1/19) |
| Cardiovascular Diseases | ECG (19/25), PPG (6/25), heart rate (2/25), IMU (1/25), Actigraphy (1/25), blood pressure (1/25) |
| Sleep Issues | PPG (2/5), Actigraphy (1/5), Polysomnography (PSG) (1/5), ECG (1/5), IMU (1/5) |
| Diabetes | Glucose (4/6), blood pressure (3/6), IMU (2/6), PPG (1/6), heart rate (1/6), skin temperature (1/6), GPS (1/6) |
| Respiratory and Pulmonary Diseases | Audio (5/15), SpO2 (4/15), IMU (3/15), PPG (3/15), skin temperature (3/15), Actigraphy (2/15), pulse rate (2/15), ratio-frequency (1/15), local trachea vibration (1/15), strain (1/15), resting heart rate (1/15), GPS (1/15), air quality (1/15) |
| Others | IMU (3/8), electromyography (EMG) (2/8), heart rate (2/8), audio (1/8), Actigraphy (1/8), PPG (1/8), ECG (1/8) |

The results are reported in the format of sensor data ($M/N$), ordered by popularity, where $M$ and $N$ are the number of articles using this data related to the health issue and the number of articles reviewed related to the health issue. Each article may use more than one sensor data.

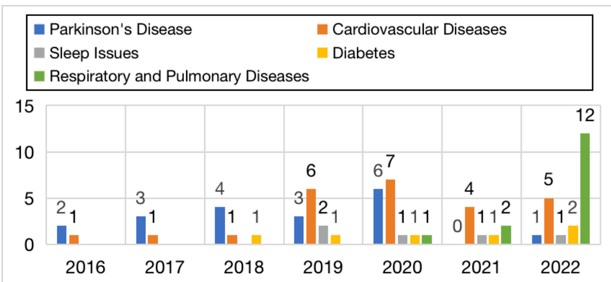


Fig. 4. The number of articles related to the five diseases across years.

summarized in TABLE III. We further look into the years of the articles with regard to the five main diseases to present the trends (the articles related to other diseases are excluded). As shown in Fig. 4, there is an explosion of articles studying respiratory and pulmonary diseases, which may be due to the worldwide outbreak and prevalence of the COVID-19 pandemic [45] since 2020. It also leads to an increase in the usage of skin temperature sensors and SpO2 sensors.

### A. Parkinson's Disease

Parkinson's disease is a progressive brain disorder. Its effects on the nervous system affect a patient's ability to move and talk, through tremors (shaking), muscle stiffness, bradykinesia (slowness), and loss of balance and coordination. Patients with Parkinson's Disease are therefore more likely to fall and injure themselves. In particular, episodes of Freezing of Gait (FoG) – a sudden, brief distortion in the ability to move – are associated with an increase in falls.

Most research articles using smart wearables on patients with Parkinson's disease focus on the detection of FoG events. Automatic detection of FoG events assists in the evaluation of the effect of medication and in the understanding of the severity and type of symptoms a particular patient has. Accurate predictions can also serve as an early warning signal, allowing for a stimulus to be sent to the patient's body to prevent FoG. There are many types of wearable devices used to take sensor measurements for detection of FoG or abnormal gait, e.g., worn on the waist [47], shank [53], foot or ankle [46], [50]–[52], or alternatively using a set of sensors simultaneously on back, hip and ankle [48] or back, arm and leg [49]. However, the technology behind each sensor is similar: in essence, all studies detecting gait make use of an inertial measurement unit (IMU) which includes a three-axis accelerometer and most often also a gyroscope. Despite the use of similar sensor technology, a wide range of machine learning techniques is applied in an effort to detect FoG events or to classify different phases in the gait, such as SVM [49], KNN [48], Layered Recurrent Networks [53], and CNN [47]. In a comparison of SVM, KNN, DT, and RF, Alam *et al.* [46] found the best performance using SVMs. Pérez-Ibarra *et al.* applied both supervised learning techniques [51] as well as unsupervised learning techniques [52], with the advantage of unsupervised learning techniques being that it removes the need for clinical tests and labeling before deployment of the devices.

Another area of research in patients with Parkinson's disease is using wearable devices to detect and classify tremor severity. With wearable devices, it is easier to follow a patient for a longer period of time, compared to a regular doctor's visit. Tremor severity is classified using measurements of an accelerometer and gyroscope on the patient's finger, which is attached with a cable to a wristwatch-type wearable device on the wrist [56]–[58]. Alternatively, a commercial off-the-shelf smartwatch with built-in accelerometer or a sensor plaster with accelerometer and gyroscope can be used [60]. The measurements from the accelerometer and gyroscope are used as input in a classification model to predict tremor severity in 5 different classes. The results of the ML methods are compared to ratings assigned by two neurologists using the Unified Parkinson's Disease Rating Scale (UPDRS) standard [56]. Jeon *et al.* [57] found decision trees to be the most accurate method for tremor severity classification, compared to SVM, DA, RF, and KNN. With an accuracy of 85% and normalized area under the curve (NAuC) of 0.98, the automatic classifier was almost as accurate as the neurologist expert rating. Another study [58] proposed the use of CNNs in which the input for training and testing consisted of a newly created 2D image of accelerometer and gyroscope signals converted into the frequency domain. This method retrieved a similar accuracy of 85% but outperformed the other methods to which it was compared, namely RF, NB, LR, DT, MLP, and SVM. CNN did particularly well in distinguishing between level 0 and level 1 tremors, i.e., differentiating between the absence of a tremor and a very slight tremor. For detection of the presence of tremor, another study used Quadratic Discrimi-

TABLE IV
OVERVIEW OF REVIEWED ARTICLES RELATED TO PARKINSON'S DISEASE.

| Article | Objective | Wearable device and sensor data | Machine learning techniques | Locations |
|---|---|---|---|---|
| Alam *et al.* 2017 [46] | Abnormal gait detection | Vertical ground reaction force sensors | SVM, KNN, Tree-based | Foot |
| Camps *et al.* 2018 [47] | FoG detection | Three-axis accelerometer, gyroscope, and magnetometer | NN | Waist |
| Demrozi *et al.* 2020 [48] | FoG prediction | Three-axis accelerometer | KNN | Back, hip, ankle |
| Ireland *et al.* 2016 [49] | Movement pattern and rhythm detection | IMU with accelerometer and gyroscope | SVM | Right leg, left arm, back |
| Mikos *et al.* 2019 [50] | FoG detection | IMU | SVM, NB, KNN, NN, Tree-based, LR | Ankle |
| Pérez-Ibarra *et al.* 2020a [51] | Gait phases and events detection | Single-IMU wearable device with three-axis velocity sensor and accelerometer | SVM | Foot |
| Pérez-Ibarra *et al.* 2020b [52] | Gait phases and events detection | Single-IMU wearable device with three-axis velocity sensor and accelerometer | HMM | Foot |
| Zia *et al.* 2016 [53] | FoG detection | Three-axis accelerometer | NN | Shank |
| Belgiovine *et al.* 2018 [54] | Dyskinesia detection | Smartwatch with three-axis accelerometer and gyroscope | SVM, Tree-based | Wrist |
| Loaiza Duque *et al.* 2019 [55] | Classification of Parkinson and Essential Tremor | Smartphone with accelerometer | DA, KNN, LR | Wrist |
| Jeon *et al.* 2017a [56] | Tremor severity classification | Finger sensor with accelerometer and gyroscope connected to wristwatch-type wearable device | SVM, DA, KNN, Tree-based | Finger, wrist |
| Jeon *et al.* 2017b [57] | Tremor severity classification | Finger sensor with accelerometer and gyroscope connected to wristwatch-type wearable device | SVM, DA, KNN, Tree-based | Finger, wrist |
| Kim *et al.* 2018 [58] | Tremor severity classification | Finger sensor with accelerometer and gyroscope connected to wristwatch-type wearable device | SVM, NB, NN, Tree-based | Finger, wrist |
| Farhani *et al.* 2022 [59] | Tremor severity classification | Surface electromyography (sEMG) data | BiLSTM | Arm |
| Shawen *et al.* 2020 [60] | Tremor and bradykinesia classification | Commercial smartwatch with accelerometer, sensor plaster with accelerometer and gyroscope | RF | Wrist |
| Mahadevan *et al.* 2020 [61] | Tremor severity classification and bradykinesia classification | Raw accelerometer data | RF | Wrist |
| Aich *et al.* 2018 [62] | Classification of Parkinson and Alzheimer | Wearable devices with 3D motion analysis system | DA, NB, Tree-based | Knee |
| Aich *et al.* 2020 [63] | Medication ON/OFF state classification | Three-axis accelerometer | SVM, NB, KNN, Tree-based | Knee |
| Khodakarami *et al.* 2019 [64] | Medication ON/OFF state classification | Smartwatch with three-axis accelerometer | SVM, Tree-based, LR | Wrist |

nant [55]. The same study also tried to distinguish between Parkinson's disease and essential tremor and found that the best performance was achieved by LR, compared to medium KNN and cubic KNN [55]. Deep learning techniques have also been leveraged to classify tremor types for people with Parkinson's Disease. In [59], a bidirectional long short-term memory (BiLSTM) algorithm was developed to identify the motion and tremor types using only surface electromyography (sEMG) data. Mahadevan *et al.* [61] considered both resting tremor and FoG. They developed a binary resting tremor classifier and a gait classifier heuristic and machine learning models in a hierarchical framework for the assessment of resting tremor and bradykinesia based on raw accelerometer data from the wrist on the most affected side.

Other areas of research, in which the literature is more sparse, include classification to distinguish between Parkinson's Disease and Alzheimer [62], dyskinesia detection [54], and detection of ON/OFF state of medication [63], [64]. An overview of all reviewed articles related to Parkinson's Disease is shown in TABLE IV.

In general, smart wearable devices, in combination with ML techniques, are very well capable of detecting FoG events as well as detecting the severity of symptoms in patients with Parkinson's disease. The performance of the machine learning algorithms is very similar to the accuracy of clinical expert diagnoses while saving time and expenses. Suggestions for further studies include the exploration of other deep learning techniques capable of including the temporal nature of the sensor data (i.e., time series analysis, e.g., gated-recurrent-unit), attention mechanism, larger sample sizes, real-time evaluation, tests on more resource-constraint devices, training models on multiple subjects instead of individual subjects, and integration of multiple data modalities (e.g., skin conductance, medication cycles, and physical location) [47], [48], [53].

### *B. Cardiovascular Diseases*

Heart disease is the leading cause of death worldwide. It is a grave disease that influences the heart's functionality and gives rise to complications such as infection of the coronary artery and diminished blood vessel functions. Heart disease patients do not feel sick until the very last stage of the disease, and then it is too late because the damages have become irretrievable. Predicting heart disease is a complex task since it requires experience along with advanced knowledge. Electrocardiography monitoring devices are major tools that help physicians diagnose cardiac abnormalities. Any abnormal behavior in the electrocardiogram (ECG) signal is an indicative measure of a malfunctioning of the heart, termed an arrhythmia condition. Due to the involved complexities such as lack of human expertise and high probability of misdiagnosing, long-term monitoring based on computer-aided diagnosis (CADiag) is preferred [65]. Based on the existing works, it has been observed that machine learning methods outperform traditional methods in arrhythmia detection.

Multiple studies have used ECG signals to diagnose cardiac abnormalities. For example, in-hospital cardiac arrest is a major burden in healthcare. Although several track-and-trigger systems are used to predict cardiac arrest, they often have unsatisfactory performances. In [66], the study hypothesized that a deep-learning-based artificial intelligence algorithm (DLA) could effectively predict cardiac arrest using ECG. It developed and validated a DLA to predict cardiac arrest using ECG. The results indicate that cardiac arrest could be screened and predicted not only with a conventional 12-lead ECG, but also with a single-lead ECG using a wearable device employing the DLA. Wang *et al.* [67] studied congestive heart failure (CHF), which refers to the inadequate blood filling function of the ventricular pump and it may cause an insufficient heart discharge volume that fails to meet the needs of the body metabolism. Inspired by the inception module from GoogLeNet, it combined LSTM and an Inception module for CHF detection using EGC signals. Wu *et al.* [68] also proposed an end-to-end model for generic and personalized ECG arrhythmic heartbeat detection on ECG data from both wearable and non-wearable devices. It developed a deep learning-based model to address the problem caused by inter-patient differences in ECG signal patterns. This model achieves state-of-the-art performance for ECG heartbeat arrhythmia detection on the commonly used benchmark from the MIT-BIH Arrhythmia Database. The spatial QRS-T angle is also a promising health indicator for the early detection of dangerous cardiac events. Santos *et al.* [69] proposed to estimate the QRS-T angle using a 1-D convolutional neural network (CNN1D) based on the 12-lead ECG.

Continuous monitoring of ECG from wearable devices can enable early detection of heart diseases. Ubiquitous monitoring on wearable electronics requires a novel class of algorithms that are low-power and have low-memory requirements. Corradi *et al.* [70] proposed a wearable compatible and automatic solution for annotating ECG recordings with an LSTM. The solution can maintain high accuracy of detection even when users are carrying out daily activities such as sitting, walking, and resting. Mhamdi *et al.* [71] deployed MobileNetV2 and VGG16 algorithms to classify cardiac arrhythmia into four categories (i.e., normal, myocardial infarction, myocardial infarction, and cardiac arrhythmias) based on 12-lead ECG tracings. They found that there was a small decrease in accuracy after implementing the algorithms on Raspberry Pi. Some works improved the real-time monitoring of ECG signals in a cheaper and more portable way. Walinjkar *et al.* [72] also contributed to monitoring an individual's ECG readings using a wearable 3-lead ECG kit and performing real-time analyses to detect arrhythmia to be able to identify and predict heart risk. Meng *et al.* [73] developed a lightweight structure named LightCov Attention (LCA) to identify ventricular contraction (PVC) and supraventricular premature beat (SPB) for arrhythmia detection from dynamic ECG readings. The proposed model achieved satisfactory performance with fewer parameters than the self-attention of the Fussing Transformer. In [74], a three-layer ANN model was implemented to detect myocardial infarction based on heart rate variability (HRV) analysis of one-lead ECG signals.

In the past decade, sensor networks for healthcare IoT have advanced quickly. We observe increasing applications to incorporate instantaneous health data by linking bodies and sensors. In [75], an IoT framework is proposed to evaluate heart disease using a Modified Deep Convolutional Neural Network (MDCNN). The smartwatch and heart monitor device that is attached to the patient monitors the blood pressure and ECG. The MDCNN is utilized for classifying the received sensor data into normal and abnormal. Similarly, Sarmah [76] proposed an IoT-based monitoring system to detect heart disease using the Deep Learning Modified Neural Network (DLMNN) classifier. The system has three steps, including authentication, encryption, and classification to provide secure data transfer and reliable disease prediction. Lin *et al.* [77] also studied several common arrhythmias and built a CNN-based algorithm for cardiac disease classification.

Apart from cloud-based analysis, Akrivopoulos *et al.* [90] proposed a fog computing approach that extends the cloud computing paradigm by migrating data processing closer to the sensors, thus accelerating the system's responsiveness to events. Moreover, Scire *et al.* [78] developed a distributed solution by transferring sensor data processing and analysis tasks to the edges of the network. The resulting solution enables the analysis and interpretation of sensor data traces within the wearable device to provide actionable alerts without any dependence on cloud services. It uses a supervised-learning approach to detect heartbeats and classify arrhythmias.

TABLE V
OVERVIEW OF REVIEWED ARTICLES RELATED TO CARDIOVASCULAR DISEASES.

| Article | Objective | Wearable device and sensor data | Machine learning techniques | Locations |
|---|---|---|---|---|
| Dinakarrao *et al.* 2020 [65] | Arrhythmia detection | On-body wearables and smartphones, ECG data | Statistical methods, NN, SVM | Body |
| Kwon *et al.* 2020 [66] | Predict cardiac arrest | ECG data | A deep learning-based artificial intelligence algorithm (DLA) | Chest |
| Wang *et al.* 2019 [67] | Congestive heart failure (CHF) detection | ECG data | LSTM and an Inception module | Wrist |
| Wu *et al.* 2018 [68] | Arrhythmia detection | ECG data | CNN | Chest |
| Santos *et al.* 2022 [69] | Estimate QRS-T angle | ECG data | A 1-D CNN | Wrist |
| Corradi *et al.* 2019 [70] | Annotating ECG recordings | Fitbit Charge HR, Apple Watch Series 4, TicWatch Pro, and Empatica E4. | LSTM | Wrist |
| Mhamdi *et al.* 2022 [71] | Arrhythmia classification | ECG data | MobileNetV2 and VGG16 | Chest |
| Walinjkar *et al.* 2017 [72] | Arrhythmia detection | Wearable ECG kit (like smartwatch), ECG data | KNN | Wrist |
| Meng *et al.* 2022 [73] | Arrhythmia detection | ECG data | LightCov Attention (LCA) | Chest |
| Shahnawaz *et al.* 2021 [74] | Myocardial infarction detection | ECG data | A three-layer ANN model | Chest |
| Khan *et al.* 2020 [75] | Heart disease diagnoses and prediction | Omron HeartGuid-bp8000m, AD8232 heart monitor board, ECG data | MDCNN (a deep learning neural network with weights optimized using the adaptive elephant herd optimization algorithm) | Wrist |
| Sarmah 2020 [76] | Heart disease detection | Hungarian HD dataset | Deep Learning Modified Neural Network (DLMNN) | Chest |
| Lin *et al.* 2019 [77] | Cardiac disease detection | ECG data | CNN-based algorithm | Chest |
| Scire *et al.* 2019 [78] | Heartbeat detection and arrhythmia classification | ECG data | KNN, LSTM | Chest |
| Zhang *et al.* 2021 [79] | Arrhythmias classification | ECG data | SVM with Gaussian Radial Basis Function kernel | Chest |
| Tiwari *et al.* 2022 [80] | Atrial fibrillation detection | ECG data | An ensembled CNN architecture and LSTM architecture (ConvNet-LSTM) | Chest |
| Ramesh *et al.* 2021 [81] | Atrial fibrillation detection | ECG and PPG data | A 1-D CNN | Finger |
| Zhang *et al.* 2021 [82] | Atrial fibrillation detection | ECG data | A Transposed Projection - CNN (TP-CNN) | Chest |
| Chen *et al.* 2020 [83] | Atrial fibrillation detection | Amazfit Health Band 1S, PPG and ECG data | A deep CNN (SEResNet) | Wrist |
| Nemati *et al.* 2016 [84] | Atrial fibrillation detection | Smartwatch, Samsung Simband, PPG data | An Elastic Net logistic model | Wrist |
| Hiraoka *et al.* 2022 [85] | Atrial fibrillation detection | Pulse rate data obtained from an Apple Watch with built-in PPG sensors | GBDT | Wrist |
| Pereira *et al.* 2019 [86] | Atrial fibrillation detection | PPG data | Attention LSTM Fully Convolutional Network (ALSTM-FCL), Fully Connected Network (FCN), VVG19, ResNet18, ResNet50, and Xception | Chest |
| Maritsch *et al.* 2019 [87] | Improving heart rate variability measurements | A chest-based heart rate monitor (Firstbeat Bodyguard, consumer smartwatches equipped with an accelerometer, and an optical heart rate sensor capable of measuring inter-beat intervals | CNN | Chest |
| Choksatchawathi *et al.* 2020 [88] | Heart rate estimation | UK Biobank wearable device, actigraphy data, activity levels during sleep, chronotype-related features, and periodic features | Unsupervised sleep-wake identification algorithm based on HMM and a penalized multi-band learning approach | Chest |
| Nidigattu *et al.* 2020 [89] | Estimation of heart rate, systolic blood pressure, and diastolic blood pressure | MAX86150 FTHR EVKIT (PPG data), Dr. Morepen BP One (blood pressure), Sphygmomanometer (heart rate and blood pressure) | RF, SVM, KNN | Finger |

Atrial fibrillation (AF) is one of the most common chronic diseases, with an estimated total incidence of about 3% in adults older than 20 years. AF increases the incidence of heart failure, stroke, and dementia. Traditional diagnosis of AF depends on professional analysis of 12-lead ECG records. Many works have been proposed to detect AF directly using multi-lead or single-lead ECG signals based on machine learning. Small wearable devices are usually able to collect single-lead ECG signals. In [79], SVM with Gaussian Radial Basis Function kernel was utilized to classify Normal Sinus Rhythm (NSR), Atrial Fibrillation (AF), Other Arrhythmias, and Noise (too much noise to recognize) from single-lead ECG signals. Tiwari *et al.* [80] proposed an ensembled Convolution Neural Network architecture and LSTM architecture (ConvNet-LSTM) to detect AF based on ECG data. The proposed model can be deployed in wearable devices and achieved an accuracy of 98% on MIT-BIH atrial fibrillation database. To lower the energy of ECG acquisition and transmission for wearable devices, Zhang *et al.* [82] proposed to compress the ECG signals by a simple deterministic measurement matrix (SDMM) and obtain the approximate ECG signals in the cloud by transpose projection operation on the compressed signals. A Transposed Projection - Convolutional Neural Network (TP-CNN) was used to detect AF on the approximate ECG signals.

The study in [83] measured the sensitivity, specificity, and accuracy of a recently developed smart wristband device with both photoplethysmography (PPG) and single-channel ECG systems. The work showed that a combination of PPG, ECG, and AI algorithms is promising for facilitating AF detection. Ramesh *et al.* [81] proposed a one-dimensional deep CNN to detect AF based on both ECG and PPG data. The model was first trained on ECG data and assessed on PPG data through transfer learning. It achieved high accuracy (more than 95%) on both ECG and PPG datasets. Dagher *et al.* [91] echoed that PPG-generated pulse waveform has the potential to accurately detect episodes of atrial fibrillation and one day could replace conventional diagnostic and long-term monitoring methods. Nemati *et al.* [84] proposed a noise-resistant machine learning approach for detecting AF from noisy ambulatory PPG recorded from the wrist using a modern research watch-based wearable device (the Samsung Simband). Pulse (beat) detection was performed on the PPG waveforms, and features were extracted based on beat-to-beat variability and waveform signal quality. The experiment results showed that the described approach provides a noise-resistant accurate screening tool for AF from PPG sensors. Hiraoka *et al.* [85] proposed to detect AF from pulse rate data obtained from an Apple Watch with built-in PPG sensors based on GBDT and achieved an accuracy of 0.9416 (sensitivity: 0.909, specificity: 0.838). In addition, Pereira *et al.* [86] compared PPG quality assessment approaches based on machine learning techniques. Image representation of raw PPG data was introduced to enable the application of powerful renowned deep-learning image classification approaches. High accuracy was obtained using the ResNet18 model, outperforming the previously proposed SVM-based PPG quality assessment model. It shows that refined deep learning approaches may further benefit from additional large datasets and provide robust tools for PPG quality assessment.

Many wearable devices use PPG technology nowadays, though they are inherently less accurate than conventional electrocardiography monitoring techniques. Research work has been conducted to improve sensing and classification accuracy using machine learning methods. For example, Maritsch *et al.* [87] studied systematic error that is related to the wearer's movement. It showed that the error could be minimized by bringing into context additional available sensor information, such as accelerometer data. It further demonstrated how neural learning could minimize the error of such smartwatch HRV measurements. To improve sensing accuracy, Choksatchawathi *et al.* [88] attempted to perform a post-calibration of the heart rate (HR) estimation during the three possible states of average daily activity (resting, laying down, and intense treadmill activity states) on four popular wearable devices: Fitbit Charge HR, Apple Watch Series 4, TicWatch Pro, and Empatica E4. The experimental results demonstrated the feasibility of the proposed methods to provide HR monitoring post-calibrated with high accuracy. Nidigattu *et al.* [89] also developed a PPG-based detection system for the estimation of heart rate, systolic blood pressure, and diastolic blood pressure using machine learning methods, including RF, SVM, and KNN. Signal processing was done for noise reduction using various filtering techniques to achieve optimal quality signals. Extensive feature engineering has been conducted, and the RF algorithm achieved better performance for the heart rate and blood pressure estimation. An overview of all reviewed articles related to Cardiology is shown in TABLE V.

Despite limitations on sensing accuracy, wearable technologies are attractive in providing low-cost and non-intrusive health monitoring. Studies evaluating PPG-based wearables in conjunction with machine-learning algorithms have shown promise in cardiovascular disease detection. Research challenges of wearable technologies, including their accuracy and accessibility, and the clinical implications of wearable-detected arrhythmias remain to be further explored in the field.

### C. Sleep Issues

Sleep occupies about one-third of people's lifetime and is crucial to human health [97]. Wearable devices can realize non-intrusive and non-invasive monitoring of people's sleep status [98]. There have been numerous works on sleep monitoring using wearable devices, such as smart rings [99] and smart watches [97]. Various machine learning algorithms, such as DT, KNN, RF, SVM, and GDBT have been leveraged to measure sleep quality [97], [100], identify sleep stages [101], [102] based on accelerometer and heart rate data. Cho *et al.* [92] presented a deep learning architecture, Deep-ACTINet, to automatically detect sleep-awake based on only noise-canceled raw activity signals recorded by wrist-worn ActiGraphy during sleep. Arora *et al.* [93] proposed three sleep indicators including Daily Sleep Quality, Weekly Sleep Quality, and Sleep Consistency calculated using the data collected by commercial wearable devices, including Samsung Galaxy Smartwatch and Xiaomi MI Smartband, and evaluated with the data collected by clinical ActiGraph devices. CNN and MLP were utilized to predict sleep quality based on the proposed indicators.

Specifically, Sleep Apnea (SA) is a highly-prevalent breathing disease, which can increase the morbidity and mortality of human beings [95]. However, although many people are suffering from this disease, it is quite inconspicuous since it happens during sleep time. Traditionally, SA is usually detected by Polysomnography (PSG), which is expensive and inconvenient since it requires multiple specialized sensors that are provided in hospitals or labs. Therefore, researchers have been working on detecting SA using sensors in non-intrusive wearable devices. ECG signals are one of the most physiologically relevant to SA and can be obtained by many wearable devices [94]. Wang *et al.* [94] modified the LeNet-5 convolutional neural network with adjacent segments to detect sleep apnea using the PhysioNet Apnea-ECG dataset [103] and UCD dataset [104]. The performance of the proposed method has outperformed traditional machine learning methods including SVM, KNN, LR, and MLP.

TABLE VI
OVERVIEW OF REVIEWED ARTICLES RELATED TO SLEEP ISSUES.

| Article | Objective | Wearable device and sensor data | Machine learning techniques | Locations |
|---|---|---|---|---|
| Cho *et al.* 2019 [92] | Sleep-awake detection | Wrist-worn ActiGraphy, noise canceled raw activity signals | A deep learning architecture (Deep-ACTINet) | Wrist |
| Arora *et al.* 2020 [93] | Sleep quality evaluation | Samsung Galaxy Smartwatch or Xiaomi MI Smartband, sleep attributes | MLP, CNN | Wrist |
| Wang *et al.* 2019 [94] | Sleep apnea detection | PhysioNet Apnea-ECG dataset | Modified the LeNet-5 CNN with adjacent segments, SVM, KNN, LR, MLP | Chest |
| Ye *et al.* 2021 [95] | Sleep apnea detection | RR-interval signals | Multi-frequency dilated convolutional neural network (FENet), CNN, LSTM, and CRNN | Wrist |
| Benedetti *et al.* 2022 [96] | Sleep apnea detection | Fitbit Charge4TM, Fitbit AltaHRTM, tri-axial accelerometer data, PPG data. | MLP, RF | Wrist |

SA detection based on the interval between two consecutive pulses (RR-interval) which can be obtained from PPG pulse sensors has also achieved satisfactory performance. In [95], a novel multi-frequency dilated convolutional neural network (FENet) was proposed to detect SA based on RR-interval signals. They considered the limited battery capacity of small wearable devices and the proposed method was adaptive to down-sampled discontinuous signals. The proposed method was energy efficient and outperformed CNN, LSTM, and CRNN. Some works leveraged the sensor data and the demographics of users to detect SA. In [96], Fitbit Charge4TM and Fitbit AltaHRTM were used to collect tri-axial accelerometer data and PPG data. Energy expenditure and step counts were estimated using accelerometer data, and PPG data were used to estimate heart rates. Based on these data, six features (i.e., sleep efficiency, the ratio of total sleep time to time in bed, the total number of awakenings divided by the total sleep time, the period of wakefulness that occurs after defined sleep onset, number of awakenings after sleep onset, and the mean length of awakenings after sleep onset) were generated. Three pairs of classifiers based on MLP and RF were used to classify the severity of SA into four categories (i.e., healthy, mild, moderate, and severe) using the above-mentioned features, age, gender, and BMI. An overview of all reviewed articles related to SA is shown in TABLE VI.

Generally, most of the pervasive commercial smart wearable devices like smartwatches have provided the function of sleep monitoring, including the classification of sleep stages and estimation of sleep quality, in which machine learning algorithms play a significant role. Some more advanced devices can detect SA or provide the probability of SA. Even if the detection results are not very accurate, they can draw more attention to this prevalent but inconspicuous disease. The characteristics of small wearable devices, such as portability and non-intrusiveness, make them ideal devices for healthcare during sleep time. However, the availability of SA detection in cheaper devices is a challenge. Currently, SA detection is highly dependent on ECG or PPG data. Continuous real-time ECG or PPG monitoring during sleep time leads to the high power consumption of wearable devices, which should be taken into consideration.

### D. Diabetes

Diabetes is a globally prevalent metabolic disease [109]. It is associated with increased risks of various serious diseases, such as heart disease and vital organ failures, and has affected millions of people each year [105]. Early detection of diabetes is significant for timely treatment and can reduce the probability of serious outcomes. In tradition, diabetes can be predicted by Electronic Medical Records (EMR) consisting of prescriptions and diagnoses from doctors. Also, there are commercial blood glucose meters, which are usually used to help patients manage their diabetes by analyzing the fluctuation of blood glucose levels. A special needle is used to poke a fingertip to get a drop of blood on a test strip to estimate the blood glucose level. There are studies working on classifying diabetes based on glucose meters and ML. For example, in [105], data collected from glucose meters and multiple sensors, including motion sensors, temperature sensors, and location sensors, were used to classify diabetes patients based on using NB, RF, ZeroR, simple logistic, sequential minimal optimization (SMO), and J48. The results showed that the SMO algorithm exhibited excellent classification with the highest accuracy of 99.66%. However, glucose meters are ad-hoc and invasive, and works on diabetes classification based on glucose meters usually also requires educated specialists or other sensor data.

In recent years, the popularity and development of non-invasive wearable devices have provided us with new opportunities to facilitate the early detection of diabetes. Kaur *et al.* [106] introduced several non-invasive smart devices for continuous monitoring of glucose data, and proposed a cloud IoT-based framework for diabetes prediction based on the blood glucose data. The performances of RF, NN, NB, DT, SVM, and their ensemble models were compared. The results showed that the ensemble model of DT and NN achieved the best performance. Ramensh *et al.* [107] and Torkey *et al.* [108] both used public diabetes datasets including 8 attributes (i.e., number of pregnancies, glucose level, diastolic blood pressure, skin fold thickness, body mass index, serum insulin level, age, and a diabetes hereditary factor pedigree function), which can be collected by non-invasive glucose meters smartwatches, and smartphones. Ramensh *et al.* developed an SVM and achieved an accuracy of 83.20%. Torkey *et al.* compared RF, SVM, DT, and NB, among which RF achieved the best performance with an accuracy of 96.15%. Some studies worked on predicting

TABLE VII
OVERVIEW OF REVIEWED ARTICLES RELATED TO DIABETES.

| Article | Objective | Wearable device and sensor data | Machine learning techniques | Locations |
|---|---|---|---|---|
| Rghioui *et al.* 2020 [105] | Diabetes classification | Glucose meters, motion sensor, temperature sensor, and location sensor | NB, RF, ZeroR, simple logistic, sequential minimal optimization (SMO), and J48 | Wrist |
| Kaur *et al.* 2018 [106] | Diabetes prediction | Blood glucose data | RF, NN, NB, DT, SVM, and their ensemble models | Wrist |
| Ramensh *et al.* 2021 [107] | Diabetes classification | Public diabetes datasets | SVM | Wrist |
| Torkey *et al.* 2022 [108] | Diabetes classification | Public diabetes datasets | RF, SVM, DT, and NB | Chest |
| Hettiarachchi *et al.* 2019 [109] | Diabetes prediction | PPG data and physiological data including age, gender, weight, and height | NB, LDA, DT, RF, AdaBoost, LR, and SVM | Chest |
| Rashid *et al.* 2022 [110] | Treatment of diabetes improvement | Wristband, accelerometer, galvanic skin response, blood volume pulse, skin temperature, and heart rate data | A 1-D Convolutional RNN with LSTM layer | Wrist |

diabetes based on PPG signals. For example, Hettiarachchi *et al.* [109] proposed a method for predicting Type 2 Diabetes from short PPG signals extracted from smart devices and physiological data including age, gender, weight, and height. The NB classifier, Linear Discriminant Analysis (LDA), DT, RF, AdaBoost, LR, and SVM were employed. Wearable devices can also help the management of diabetes. Rashid *et al.* [110] proposed a 1-D Convolutional RNN with LSTM layer to classify the four states (i.e., ”Meal and Exercise”, ”no Meal but Exercise”, ”no Exercise but Meal”, ”neither Meal nor Exercise”) using the accelerometer, galvanic skin response, blood volume pulse, skin temperature, and heart rate data from the wristband and historical data to improve automated insulin delivery systems for treatment of diabetes. The summary of reviewed papers on diabetes is shown in TABLE VII.

There is an increase in works on diabetes prediction using wearable devices. However, most of them require blood glucose data, while there are only a few non-intrusive wearable devices that are able to measure blood glucose data. In addition, the accuracy of the glucose level estimation using non-intrusive wearable devices should be further improved. Diabetes detection based on other data, such as PPG signals, is also a promising direction.

### *E. Respiratory and Pulmonary Diseases*

Respiratory and pulmonary diseases include disorders or infections that affect the lungs, such as asthma, bronchiectasis, bronchiolitis, and chronic obstructive pulmonary disease (COPD). This kind of disease can cause breathing problems and reduce the quality of life. Researchers have been working on using wearable devices to identify abnormal symptoms to help early detection and non-intrusive management of respiratory and pulmonary diseases.

Most of the studies work on identifying abnormal airway symptoms such as cough, wheezing, and crackles. In [111], researchers detected three airway symptoms (i.e., cough, throat clear, and dry swallow) using audio and mechano-acoustic data, such as neck surface accelerometers. They used three datasets, including the Rainbow Passage dataset (a study of reading a standard passage scripted with airway symptom productions), the Vocal Stress dataset (a published study of vocal loading tests), and the COUGHVID dataset (a crowdsourcing COVID-19 cough sound project). They compared the performance of the ResNet architecture (ResNet18 and ResNet34), EfficientNetB0, MobileNetV2, an Encoder–Decoder–RNN, and a vanilla RNN. The CNN-based models were more accurate but slightly slower than RNN-based models. Hui *et al.* [112] presented a wearable ratio-frequency sensor to collect the local trachea vibration characteristics. The retrieved tissue vibration caused by the cough airflow burst was then analyzed by a CNN trained on the frequency-time spectra to directly identify the mild cough signal. Xue *et al.* [113] proposed a kernel-like minimum distance classifier (K-MDC) deployed in wearable devices to identify cough, breath, and wheeze from respiratory sounds recorded over the right side of the chest wall using an electret microphone.

However, the above-mentioned methods rely on throat-fixed flexible sensors, which are less comfortable and more obvious. In [114], a deep CNN-RNN model was proposed to classify respiratory sounds based on Melspectrograms using the public respiratory sound database containing samples recorded with different equipment from hospitals. The proposed method can be deployed in wearable devices with microphones for detecting abnormal respiratory sounds such as crackles and wheezes. Yang *et al.* [115] proposed a chest-laminated electronic skin for cough identification. The e-skin has mixed dumbbell-like networks and through-holes and is sensitive to stretching force and sweat permeation. The ’inflates’ and ’deflates’ of e-skin lead to sensor strain and generate corresponding electrical signals, based on which LSTM was utilized to identify cough. Zhang *et al.* [116] collected the accelerometer amplitude and microphone audio using wearable devices to detect cough. Accelerometer amplitude data were first used to identify suspicious cough cases through a simple threshold filter. When the accelerometer amplitude passes through the filter, the corresponding multimodal data is processed by auto-encoders. The attention mechanism was utilized to assign different weights to different modalities. The representation of multimodal data was aggregated for cough detection.

Furthermore, the worldwide outbreak and prevalence of the COVID-19 pandemic have attracted global attention [45]. The epidemic makes the hospitals overburdened. Hospitals around

TABLE VIII
OVERVIEW OF REVIEWED ARTICLES RELATED TO RESPIRATORY AND PULMONARY DISEASES.

| Article | Objective | Wearable device and sensor data | Machine learning techniques | Locations |
|---|---|---|---|---|
| Groh *et al.* 2022 [111] | Three airway symptoms (i.e., cough, throat clear, and dry swallow) detection | Neck surface accelerometers (NSAs), audio data and mechano-acoustic data | ResNet architecture (ResNet18 and ResNet34), EfficientNetB0, MobileNetV2, an Encoder-Decoder–RNN, and a vanilla RNN | Neck |
| Hui *et al.* 2021 [112] | Mild cough signal identification | A wearable ratio-frequency sensor, local trachea vibration characteristics | CNN | Neck |
| Xue *et al.* 2022 [113] | Identification of cough, breath, and wheeze | Electret microphone, respiratory sounds | A kernel-like minimum distance classifier (K-MDC) | Chest |
| Acharya *et al.* 2020 [114] | Detection of abnormal respiratory sounds (e.g., crackles and wheeze) | Respiratory sounds | A deep CNN-RNN model | Chest |
| Yang *et al.* 2022 [115] | Cough identification | E-skin, strain | LSTM | Chest |
| Zhang *et al.* 2021 [116] | Cough detection | Accelerometer amplitude and microphone audio | Auto-encoders and attention mechanism | Throat and chest |
| Soltanian *et al.* 2022 [117] | COVID-19 detection | Cough sounds | Separable quadratic Convolution layers, RF, LR, CNN, SVM, GRU, and ordinary convolution. | Chest |
| Abir *et al.* 2022 [118] | COVID-19 detection | Resting Heart Rate (RHR) | A LSTM Variational Autoencoder (LSTM-VAE) | Wrist |
| Nayan *et al.* 2022 [119] | COVID-19 detection | A low-cost pulse oximeter, PPG data | Discriminant analysis, KNN, DT, SVM, and ANN | Finger |
| Hirten *et al.* 2022 [120] | COVID-19 detection | Apple Watch Series 4 or higher, PPG data | GBM, elastic-net, partial least squares, SVM, and RF | Wrist |
| Mankodiya *et al.* 2022 [121] | COVID-19 detection | SpO2, temperature, and pulse rate | LR, Bernoulli naive Bayes (BNB), SVM, and DT | Wrist |
| Mason *et al.* 2022 [122] | COVID-19 detection | Oura Ring, dermal temperature, PPG data, and accelerometer data | RF-based machine learning algorithm | Finger |
| Jadav *et al.* 2022 [123] | COVID-19 detection | Blood oxygen level, pulse rate, and body temperature | KNN and LR | Wrist |
| Wu *et al.* 2022 [124] | Chronic obstructive pulmonary disease prediction | A location-based smartphone APP, wearable devices, air quality sensing devices, and open environmental data API to collect various human lifestyle data (i.e., physical activities, heart rate, SpO2, and sleep patterns) and environmental data | RF, DT, LDA, AdaBoost, and DNN | Wrist |
| Skibinska *et al.* 2022 [125] | Distinguishing COVID-19 and two types of Influenza | Fitbit smartwatches, including the sleep records, step counts, and heart rates | XGBoost, KNN, SVM, RF, DT, and LR | Wrist |

the world have faced a significant shortage of manpower, and people's healthcare resources are quite limited, especially in developing countries. The early detection of COVID-19 is critical to slow down or even cut off its spread and is vital to relieve the heavy burden on medical systems [126]. Upon this basis, numerous works have been proposed to identify COVID-19 using wearable devices. Soltanian *et al.* [117] proposed separable quadratic Convolution layers to identify COVID-19 from cough sounds. The proposed network outperformed RF, LR, CNN, SVM, GRU, and ordinary convolution. The network is lightweight to be deployed in wearable devices. Many works leveraged the heart rate data collected by PPG to detect COVID-19 infection. For example, in [118], a Long Short-term Memory Variational Autoencoder (LSTM-VAE)-based anomaly detection framework (PCovNet) was proposed to detect COVID-19 infection in the presymptomatic stage using Resting Heart Rate (RHR) derived from the wearable devices, i.e., a smartwatch or fitness tracker. Nayan *et al.* [119] compared the performance of DA, KNN, DT, SVM, and ANN on COVID-19 identification using PPG features extracted from a low-cost pulse oximeter. The ANN showed the best performance with 95.45% accuracy, 100% sensitivity, and 90.91% specificity by using six input features. In [120], PPG data were collected from 407 participants using Apple Watch Series 4 or higher. Heart rate variability (HRV) and RHR were calculated, based on which gradient-boosting machines (GBM), elastic-net, partial least squares, SVM, and RF were implemented to predict COVID-19. GBM achieved the best performance with an average area under the receiver operating characteristic (auROC) = 86.4% (confidence interval [CI] 84–89%).

Some works leveraged various sensor data collected by multiple sensors to improve detection accuracy. In [121], SpO2, temperature, and pulse rate obtained from wearable devices were used, and four machine learning algorithms, including LR, Bernoulli naive Bayes (BNB), SVM, and DT were compared. The SVM model achieved the best performance with an F1-score of 96.64% and an accuracy score of 96.57%.

Mason *et al.* [122] collected dermal temperature, PPG data, and accelerometer data using Oura Ring. PPG data were used to generate heart rate, heart rate variability, and respiratory rate, and accelerometer data were used to estimate physical activity. Based on these data, the authors developed an RF-based machine learning algorithm to identify COVID-19 onset and achieved high sensitivity (82%) and moderate specificity (63%). In [123], a blockchain-based framework, BaRCODe, was proposed to detect COVID and provide prompt medical treatment. Smartwatches were used to collect the blood oxygen level, pulse rate, and body temperature of users. KNN and LR were incorporated into the framework to identify COVID-19 patients and malevolent wearable devices, respectively. Wu *et al.* [124] further incorporated the environmental data. They used a location-based smartphone APP, wearable devices, air quality sensing devices, and open environmental data API to collect various human lifestyle data (i.e., physical activities, heart rate, SpO2, and sleep patterns) and environmental data from 1,667 patients for 24 months. Based on these data, RF, DT, LDA, AdaBoost, and DNN were implemented to predict chronic disease including chronic obstructive pulmonary disease, obesity, and panic disorder.

Besides, Skibinska *et al.* [125] studied the usage of ML algorithms to distinguish COVID-19 and two types of Influenza. The data were obtained by Fitbit smartwatches, including sleep records, step counts, and heart rates. They compared the performance of XGBoost, KNN, SVM, RF, DT, and LR in the middle stage of the pandemic. The evaluation results show that KNN achieved the highest accuracy (73%). Also, there are some works on recognizing people's high-risk behaviors of virus transmission, such as touching face [127] and washing hands [128], which are not the focus of our review. To summarize, the reviewed papers are listed in TABLE VIII.

Generally, before the outbreak of COVID-19, works on respiratory and pulmonary disease detection using wearable devices mainly focused on identifying abnormal airway symptoms through analyzing the audio data or vibration collected by wearable devices. There are still many challenges and opportunities. For example, the neck wearable devices for collecting acoustic signals are too obvious, and those e-skin-based sensors are too ad-hoc. For those wrist-worn devices, some of them require the users to use a specific pose to collect the audio data. The success rate of data collection and accuracy of detection need to be further improved. Future work can consider incorporating multi-modal data to expand the respiratory and pulmonary diseases that can be monitored by wearable devices. On the other hand, the worldwide prevalence of COVID-19 has attracted great interest in exploring the potential of wearable devices for pandemic detection and control. Heterogeneous data collected by multiple sensors in wearable devices can benefit pandemic detection and control from different perspectives. For example, the location sensor can help track potential patients and predict infection risk. Accelerometer and gyroscope data can be used to detect critical behaviors. In this work, we mainly reviewed works that directly identify the disease using machine learning algorithms and sensor data from wearable devices. According to the above-reviewed papers, distinguishing COVID-19 and other influenza types is a great challenge since the superficial characteristics of COVID-19 are quite similar to influenza. The key difference is that COVID-19 is more likely to damage the lungs and cause shortness of breath, which is challenging for wearable devices to identify.

### F. Others

Apart from the health issues and diseases discussed in the above sections, there are numerous works related to a wide variety of other health issues. Since there are not as many articles associated with each of these diseases as the above-mentioned diseases, we briefly review and summarize these articles in this section.

*1) Epileptic Seizure:* Epilepsy is one of the most common chronic neurological diseases. Epilepsy Seizure can lead to accidents and sudden unexpected death [129]. It is important to monitor epilepsy patients in real time. By detecting epileptic seizures and notifying the caregivers, the patients can get timely help to reduce the risk of seizure-related accidents. There are works on detecting epileptic seizures based on ECG data. In [129], a method for distributing machine learning computations between the edge, fog, and cloud considering the trade-offs in terms of energy consumption, latency, and performance was proposed for epileptic seizure detection. The detection system was implemented in the Smart-Cardia INYU wearable sensor [136] using the EPILEPSIA dataset [137], which consists of ECG data from 30 patients with 277 seizures recorded in 4603 hours. Kok *et al.* [130] first extract Mel-frequency Cepstral Coefficients (MFCCs) from acoustic signals collected from the neck wearable devices, and then used a random under-sampling and boosting (RUSBoost) classification algorithm to identify the ictal and non-ictal acoustic segments. A simple post-processing stage was applied to the classification results to identify seizure episodes. In [131], multiple sensor data, including the heart rates, sleep, and step counts collected by wearable smartwatches, were used to estimate the seizure risk either daily or hourly based on an ensemble of an LSTM, an RF regressor, and an LR classifier.

*2) Frailty Syndrome:* Frailty is a common clinical syndrome related to aging [138]. It increases the risk of adverse health outcomes for the elderly. Garcia *et al.* [132] collected tri-axial accelerometer, gyroscope, and heart rate data using Samsung Gear S3, and implemented four machine learning algorithms, KNN, SVM, RF, NB, to assess frailty status for the elderly. The results showed that KNN achieved the best performance. Wearable devices can help family members or caregivers keep track of the frailty status of elders, and take timely measures for frailty prevention and treatment.

*3) Sarcopenia:* Sarcopenia is defined as an involuntary loss of skeletal muscle mass and strength due to aging or immobility [139]. It can be identified by analyzing the motion or electromyography patterns. Kim *et al.* [38] collected the tri-axial acceleration and angular velocity signals of feet to derive spatiotemporal and descriptive statistics. Shapley Additive explanations [40] were utilized to select important parameters. SVM, RF, MLP, CNN, and BiLSTM were used to identify sarcopenia based on the collected data. The SVM

TABLE IX
OVERVIEW OF REVIEWED ARTICLES RELATED TO OTHER HEALTH ISSUES.

| Article | Objective | Wearable device and sensor data | Machine learning techniques | Locations |
|---|---|---|---|---|
| Forooghifar *et al.* 2019 [129] | Epileptic seizure detection | ECG data | A method for distributing machine learning computations | Wrist |
| Kok *et al.* 2022 [130] | Epileptic seizure detection | Neck wearable devices, acoustic signals | A random under-sampling and boosting (RUSBoost) classification algorithm | Chest |
| Stirling *et al.* 2021 [131] | Epileptic seizure detection | Smartwatches, heart rates, sleep and step counts | An ensemble of an LSTM, an RF regressor, and an LR classifier | Wrist |
| Garcia *et al.* 2020 [132] | Frailty status assessment | Samsung Gear S3, tri-axial accelerometer, gyroscope, and heart rate data | KNN, SVM, RF, NB | Wrist |
| Kim *et al.* 2021 [38] | Sarcopenia detection | Tri-axial acceleration and angular velocity | SHAP, SVM, RF, MLP, CNN, and BiLSTM | Foot |
| Leone *et al.* 2022 [133] | Sarcopenia detection | Small electrodes, raw electromyography data | SVM, RF, LR, NB, DT, KNN, XGB, MLP | Chest |
| Chen *et al.* 2022 [134] | Sarcopenia detection | Electromyography signals and tri-axial accelerometer data | LCNet | Thigh, calves, waist, back |
| Wang *et al.* 2021 [135] | Abdominal Aortic Aneurysm detection | Digital artery, Photoplethysmogram arterial pulse wave (PW) signal | BRNN with LSTM model | Wrist, finger |

model with 20 descriptive statistical parameters achieved the best performance. In [133], an sEMG-based platform for sarcopenia evaluation was designed and implemented. Raw electromyography data were collected through small electrodes to generate sarcopenia-related features. Then, machine learning algorithms, including SVM, RF, LR, NB, DT, KNN, XGB, and MLP, were leveraged to classify the sarcopenia levels, among which the SVM classifier achieved the best performance. Chen *et al.* [134] collected signals of electromyography and tri-axial accelerometer data using wearable devices. After data augmentation by the DTW Barycenter Average algorithm, various indicators were calculated by the Bodi algorithm. Then, LCNet was implemented to classify the risk of sarcopenia into two classes, i.e., high risk and low risk.

*4) Abdominal Aortic Aneurysm:* Abdominal Aortic Aneurysm (AAA) is difficult to detect since it often grows without noticeable symptoms. Wang *et al.* [135] proposed to utilize BRNN with the LSTM model to facilitate early detection of AAA based on the photoplethysmogram arterial pulse wave (PW) signal measured in the digital artery with wearable devices. The early detection of aneurysms can prevent exacerbation of the disease.

The above-reviewed papers are summarized in TABLE IX. Many diseases influence human physiological signals, such as heart rates, mobility, and sleep quality. Various sensors in small wearable devices can collect these physiological data. By analyzing the abnormal patterns in these data, various diseases can be detected. The reviewed papers show the power and potential of wearable devices for disease diagnoses.

## V. CHALLENGES AND FUTURE DIRECTIONS

Although current research has achieved great success in disease diagnoses based on wearable devices and machine learning, there are still many challenges. In this section, we discuss these challenges and attempt to present some new research directions.

### A. Privacy Concerns

Most reviewed works rely on patients sharing their sensitive health data. Many works collected data from participants for model training. Although some models can be trained by desensitized public datasets, privacy concerns arise when vital signals of users are required for inference. This issue can be addressed from several perspectives. First, differential privacy [140] can be applied to databases, such as adding noise to datasets without influencing the results of data analysis. However, it cannot always guarantee privacy. The datasets may be denoised, and users still have to upload their personal data. Also, there may be trade-offs between accuracy and privacy. The generalizability of this method has also been questioned.

Federated learning (FL) scheme [141] is another promising solution. Users download a global model, based on which local models are trained using local datasets, and the trained parameters can be uploaded to improve the global model. In this case, the private data are kept on devices without uploading to servers or sharing with others. However, due to the limitations of datasets, the FL scheme may have slightly lower performance than the original scheme. Also, it usually requires more communication between edge devices and servers. For disease diagnoses in wearable devices, an FL scheme can address the problem of insufficient data at the early stage. The communication scheme can be designed to be more flexible. For example, since the application is quite personal, the local models can be updated based on only local data when they have already achieved satisfactory performance. The global model can be updated by parts of local updates and transfer the knowledge to local models when needed. In this way, the federated scheme can be more effective and efficient for disease diagnoses.

Furthermore, although the FL framework addresses the privacy concern to some extent, there still remains a risk of privacy breaches and communication bottlenecks during parameter transmission. Another novel computing paradigm, over-the-air computing has emerged as a promising approach.

It involves performing certain types of computations directly over the wireless medium, thus mitigating the need for extensive data transmission, reducing communication overhead, and enhancing privacy protection [142], [143]. Integrating over-the-air computing with FL offers a promising solution in the healthcare domain [144]. It can be particularly beneficial in tasks like patient data analysis from wearable or remote monitoring systems, offering enhanced privacy and communication efficiency [145], [146]. However, challenges such as handling noise and interference in wireless channels and ensuring secure computations over-the-air need to be addressed.

### *B. Security Concerns*

The proliferation of wearable devices has raised new security concerns in healthcare [147]. While these small, portable devices offer great convenience and potential for disease diagnosis, they are susceptible to security vulnerabilities [148]. With limited computing power, small wearable devices often have to connect to networks or other devices for analysis results, making them become attractive targets for malicious actors seeking unauthorized access to sensitive personal health information [149]. The security vulnerabilities in the software or communication protocols can lead to data breaches, identity theft, or unauthorized tampering with health data, hindering users' trustworthiness of wearable devices in disease diagnosis and monitoring applications.

To mitigate security concerns, it is crucial to implement robust security measures throughout the entire ecosystem of wearable devices. This includes secure data transmission protocols, encryption techniques to protect data at rest and in transit, and strong authentication mechanisms to ensure that only authorized individuals can access the device or its data [150]. Regular software updates and patches should be provided to address any vulnerabilities that may arise over time. Additionally, manufacturers and developers should conduct thorough security assessments and testing to identify and address potential weaknesses in the devices and their associated software. User education and awareness regarding security best practices can also play a significant role in minimizing risks.

### *C. Transmission Latency and Reliability*

Since many wearable devices require connection to a network for data storage and analysis, transmission latency and reliability are critical factors that need to be addressed when utilizing wearable devices for disease diagnosis. Data generated by wearable devices need to be transmitted in real-time for timely and accurate diagnosis [151]. However, the reliance on wireless networks and the potential for network congestion can introduce latency, causing delays in data transmission, and hindering time-sensitive diagnoses and real-time monitoring. Moreover, the reliability of data transmission is crucial to ensure the integrity of the collected data. Transmission errors or data loss can lead to incorrect diagnoses, compromising the effectiveness of the healthcare monitoring process.

To reduce latency, a potential solution is implementing efficient data compression techniques [152]. Edge computing, where data processing occurs closer to the source, can also minimize the dependency on distant servers, improving response times [153]. Furthermore, state-of-the-art transmission techniques based on machine learning have significant potential to reduce latency and increase reliability for disease diagnoses based on wearable devices [154], particularly when it comes to data processing with ultra-reliable and low latency requirements. They can help optimize future networks [155] and assist in the compression, real-time analysis, and interpretation of data collected from wearable devices.

Specifically, different smart wearable systems for different diseases may have different requirements for transmission speed and reliability. For example, the timely detection of epileptic seizures requires low latency and high reliability, while monitoring sleep issues is not as compromised with slightly higher latency. It is essential to strike a balance between transmission speed and reliability and cost depending on the monitored disease and associated risks.

### *D. Energy Consumption*

Most of the current works require continuous monitoring of vital signs, which consumes a lot of electrical power, while small wearable devices usually have small batteries and limited electrical power. Although many commercial wearable devices have already supported ECG tracking, opening this function would significantly reduce the battery life. Also, currently, most commercial wearable devices identify diseases by uploading user information to the server and obtaining results from the server. To reduce energy consumption, researchers can consider designing new sampling methods to reduce the sampling rate of vital signs while keeping the performance, applying interpolation methods to impute the low sampling data, or designing new model architectures [156] that work well on data with low sampling rates. Also, some stages can be built into the devices to reduce the communication between the server and the devices.

On the other hand, to address the privacy concerns discussed above, wearable devices are expected to train and run machine learning models locally, which definitely requires more electrical power. To address this issue, both the hardware and software should be improved. As for the hardware part, it is important to design small batteries with more electricity storage and reduce the energy consumption of the communication module. As for the software part, besides designing the federated scheme that is more effective and efficient for disease diagnoses, more lightweight machine learning models that require less computing power should also be developed.

However, it is a significant challenge to strike a balance between creating lightweight models and maintaining the necessary complexity to accommodate the variety and sophistication of tasks. Besides, as ML models grow increasingly complex, their sizes inevitably increase [157]. This results in an increasing need for more computational resources, thus increasing energy consumption. To address this issue, the following directions can be considered. Firstly, the communication of parameters across different layers consumes significant energy, especially when dealing with large datasets or complex models. Quantization and model compression techniques can be used to reduce communication costs [158], [159]. However, these methods may increase the model complexity from another perspective and degrade the performance. Secondly, the inference phase of complex models can also be energy-consuming, which is often overlooked but important in the context of IoT. A promising solution is Knowledge Distillation [160], which involves training a smaller, simpler model (student) to mimic the behavior of a larger, complex model (teacher) [161]. The distilled model can achieve comparable performance to the original model, while requiring fewer computational resources for inference, leading to energy savings. In addition, there is currently a lack of standard benchmarks and tools to measure the energy consumption of ML models accurately. Building such benchmarks will be significantly helpful to compare different models or strategies and to track improvements in energy efficiency.

### E. Multi-Modality, Multi-Sensors, and Multi-Devices

Based on the literature review, we can find that the data used in many works are quite common. Most works used accelerometers, ECG, and PPG data. Besides the sensor data, some works [108], [109] also utilized the personal information of patients, such as gender, age, and weight to help the diagnoses. Since many diseases, such as diabetes and cardiovascular diseases, are significantly related to personal information, it is reasonable to take this information into consideration. Besides, the environmental contexts are also important factors, such as weather conditions, air pollution, and living conditions, which are related to people's health status. Multi-modal data, including text, audio, image, and video can be collected to represent various personal and environmental contexts [162]. These data can be combined with frequently used vital signs to improve the disease diagnosis of wearable devices.

To increase the variety of data collected, apart from the popular sensors, such as accelerometer, gyroscope, and PPG sensor, various sensors should be developed. For example, Yang *et al.* [115] developed an e-skin to identify cough. These novel sensors facilitate the collection of multi-modal data, increasing the coverage of diseases that can be detected by wearable devices and improving performance.

Furthermore, new wearable devices are also worth studying. Currently, wristbands are the most prevalent wearable devices in both research and the market, since it is convenient and acceptable for participants to wear wristbands continuously. However, the signals collected from wristbands can be very localized and not accurate enough for disease diagnoses. On the one hand, this disadvantage can be overcome by improving the sensors and algorithms. On the other hand, new forms of wearable devices can be developed. For example, Kim *et al.* [163] proposed a sensory face mask. Corresponding to various diseases, various locations on the human body can be taken into consideration to obtain more accurate physiological data. When designing the device, many important factors, such as comfort, size, looks, convenience, charging frequency, and weight, should be taken seriously, since these factors will influence people's willingness to wear continuously and the quality of data collected.

However, the increasing heterogeneity of data collected by various wearable devices also introduces new challenges, impacting both the efficiency and performance of models. To address these challenges, various techniques can be utilized. Firstly, data preprocessing, such as scaling and normalization, can be used to standardize diverse data types and minimize discrepancies [164]. Secondly, feature selection can help identify the most relevant features, reducing data dimensionality and model complexity [165]. Thirdly, multimodal learning has aroused great attention recently, which deals with models trained on data coming from multiple different input types or modalities [166]. Developing robust multimodal learning models that can handle different types of data simultaneously and effectively is a promising solution to handle data heterogeneity.

### F. Evaluation Metrics

When evaluating the performance of the proposed methods, current works usually report the accuracy, AUC, sensitivity, and specificity. Sensitivity represents the ability of the model to identify patients with a disease correctly, while specificity represents the ability of the model to identify healthy people. Usually, a higher sensitivity correlates with a lower specificity. In the literature reviewed, when reporting the results, sensitivity and specificity are based on the threshold selected according to the ROC curve to achieve a balance between them. However, sensitivity and specificity may have different weights in different application scenarios. For some high-risk diseases, sensitivity should be prioritized, while for some low-risk diseases, specificity should be prioritized to avoid fake alarms. Therefore, it is meaningful to design evaluation metrics for different purposes and apply a more flexible alarm mechanism. For example, in practice, rather than using a fixed threshold, the threshold can be adjusted by users. Feedback mechanisms can be added to the system, and thresholds can be selected based on expert experience or crowd intelligence.

### G. Explainability

Explaining machine learning models is of vital importance in healthcare. However, from the literature reviewed, few works evaluate the explainability of their models. Among the works discussed in this survey, only Kim *et al.* [38] leveraged SHAP to select important parameters, indicating that the explainability of machine learning in healthcare using wearable devices has not been fully studied. Therefore, we encourage researchers to pay more attention to explaining their models and discussing why their models can achieve good performance. By making the decision-making process more transparent, the results can be more reliable and persuasive. By explaining the machine learning model, important features can be identified. Thereby, unnecessary and less important features can be pruned to remove unnecessary sensors and make the models more lightweight, which is important for edge devices with limited computing and electricity power. Also, by analyzing the correlation between input features and outcomes, risk factors to people's health status can be identified. In this way, a system that can make suggestions for reducing the probability of diseases can be developed.

The model explanation can provide many insights. However, as discussed in Section III-C, we should be very careful to avoid over-claims when making suggestions. Generally, there are many opportunities and challenges in this direction.

### *H. Generalization and Personalization*

The emergence of large-scale machine learning models has attracted great attention, which also has a significant impact on disease diagnoses due to their immense capacity for learning complex patterns from vast amounts of data [167]. However, they also bring new challenges in the context of wearable devices. Firstly, the computational requirements of these models usually exceed the capabilities of resource-constrained wearable devices. Besides using central servers to deploy the large models, potential solutions include distilling the knowledge from large models to build lightweight models [168] and developing efficient model compression and optimization techniques [169]. Secondly, the large generic models, trained on diverse datasets offering broad diagnostic capabilities, might lack the specificity required for individuals' unique health profiles. A growing trend is to consider personal characteristics and adapt the models to individual variations for diagnostic insights, while the development of personalized models poses challenges in terms of data availability, privacy and security concerns, and explainability as discussed above.

Generally, research efforts could focus on striking a balance between the benefits of large-scale generic models and the need for personalized models, developing techniques leveraging the advantages of both approaches. This includes exploring federated learning, transfer learning, and hybrid models that combine generic knowledge with patient-specific adaptation.

### *I. Social Influence and Human Factors*

Current health-related research based on wearable devices and machine learning mainly aims to increase the performance and generalization of models, lacking the consideration of social influence and human factors. Although many researchers from Computer Science, Electrical and Electronic Engineering, Medicine, Public Health, etc. have been working together in this area, there are few researchers from psychology and sociology. For example, patients' willingness to wear devices that could potentially diagnose their diseases should be evaluated. It will be interesting to study whether this kind of devices may increase patients' anxiety and adversely influence their comfort and mental health. Also, if patients visit the doctor less often with the help of wearable devices, it may lead to less social contact and medical control. In addition, there are still many other factors, such as placebo effect [170] and human-computer interaction (HCI) [171], remaining to be explored. Therefore, we encourage more interdisciplinary collaborations to conduct social-physiological analyses and HCI studies, and consider the well-being and benefits of both individuals and the whole society when designing the systems.

## VI. Conclusion

Nowadays, smart wearable devices, such as smart watches, are becoming more and more popular and have changed the way we live. Various small sensors and the development of machine learning make it feasible to diagnose various diseases with wearable devices. In this survey, we comprehensively reviewed articles related to specific diseases or health issues based on small wearable devices and machine learning. We classify these articles according to their corresponding disease types and summarize their objective, sensor data, machine learning techniques used, and wearing locations. Based on the literature review, we discuss the challenges and propose future directions, aiming to inspire researchers in this field.

## VII. Acknowledgments

The authors would like to thank the reviewers and editors for their constructive suggestions. This research was supported by the Hong Kong General Research Fund grant numbers 17203320 and 17209822, and the start-up grant from The University of Hong Kong.

## References

[1] S. Iqbal, I. Mahgoub, E. Du, M. A. Leavitt, and W. Asghar, "Advances in healthcare wearable devices," *NPJ Flexible Electronics*, vol. 5, no. 1, pp. 1–14, 2021.

[2] M. Richards and K. DeBonis, "Psychiatric training during a global pandemic: how covid-19 has affected clinical care, teaching, and trainee well-being," *Psychiatric Services*, vol. 71, no. 12, pp. 1300–1302, 2020.

[3] E. J. Emanuel, G. Persad, R. Upshur, B. Thome, M. Parker, A. Glickman, C. Zhang, C. Boyle, M. Smith, and J. P. Phillips, "Fair allocation of scarce medical resources in the time of covid-19," pp. 2049–2055, 2020.

[4] X. Liang, R. Ma, Y. Cheng, Q. Gong, J. Li, and B. Zhu, "The right to accessible covid-19 testing in the post-epidemic period under the urban–rural integration: Haishu district, ningbo city, china," *Sustainability*, vol. 14, no. 3, p. 1636, 2022.

[5] Y. Koumpouros and T. Kafazis, "Wearables and mobile technologies in autism spectrum disorder interventions: A systematic literature review," *Research in Autism Spectrum Disorders*, vol. 66, p. 101405, 2019.

[6] A. Channa, N. Popescu *et al.*, "Managing covid-19 global pandemic with high-tech consumer wearables: A comprehensive review," in *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE, 2020, pp. 222–228.

[7] T. Pereira, N. Tran, K. Gadhoumi, M. M. Pelter, D. H. Do, R. J. Lee, R. Colorado, K. Meisel, and X. Hu, "Photoplethysmography based atrial fibrillation detection: a review," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–12, 2020.

[8] J. Andrés, G. Bailador, C. Gibbons, A. Olmo, D. P. Consortium *et al.*, "Designing and testing healthtracker for activity recognition and energy expenditure estimation within the daphne platform," *Procedia Computer Science*, vol. 98, pp. 348–355, 2016.

[9] M. V. C. Caya, A. N. Yumang, J. V. Arai, J. D. A. Niñofranco, and K. A. S. Yap, "Human activity recognition based on accelerometer vibrations using artificial neural network," in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. IEEE, 2019, pp. 1–5.

[10] J. Manjarres, P. Narvaez, K. Gasser, W. Percybrooks, and M. Pardo, "Physical workload tracking using human activity recognition with wearable devices," *Sensors*, vol. 20, no. 1, p. 39, 2019.

[11] K. Li, R. Habre, H. Deng, R. Urman, J. Morrison, F. D. Gilliland, J. L. Ambite, D. Stripelis, Y.-Y. Chiang, Y. Lin *et al.*, "Applying multivariate segmentation methods to human activity recognition from wearable sensors' data," *JMIR mHealth and uHealth*, vol. 7, no. 2, p. e11201, 2019.

[12] A. L. S. Varma, V. K. Sinha, and D. Das, "Health assessment for office workers by tracking the wrist motion centric activity," in *2019 IEEE 16th India Council International Conference (INDICON)*. IEEE, 2019, pp. 1–4.

[13] N. Varshney, B. Bakariya, A. K. S. Kushwaha, and M. Khare, "Human activity recognition by combining external features with accelerometer sensor data using deep learning network model," *Multimedia Tools and Applications*, pp. 1–20, 2022.

[14] W. L. Haskell, I.-M. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson, and A. Bauman, "Physical activity and public health: updated recommendation for adults from the american college of sports medicine and the american heart association," *Circulation*, vol. 116, no. 9, p. 1081, 2007.

[15] S. Asthana, A. Megahed, and R. Strong, "A recommendation system for proactive health monitoring using iot and wearable technologies," in *2017 IEEE International Conference on AI & Mobile Services (AIMS)*. IEEE, 2017, pp. 14–21.

[16] M. Gołosz and D. Mrozek, "Detection of dangers in human health with iot devices in the cloud and on the edge," in *International Conference: Beyond Databases, Architectures and Structures*. Springer, 2019, pp. 40–53.

[17] J. Guo, L. Yang, R. Bie, J. Yu, Y. Gao, Y. Shen, and A. Kos, "An xgboost-based physical fitness evaluation model using advanced feature selection and bayesian hyper-parameter optimization for wearable running monitoring," *Computer Networks*, vol. 151, pp. 166–180, 2019.

[18] S. H. Chae, Y. Kim, K.-S. Lee, and H.-S. Park, "Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: prospective comparative study," *JMIR mHealth and uHealth*, vol. 8, no. 7, p. e17216, 2020.

[19] A. Johnson, F. Yang, S. Gollarahalli, T. Banerjee, D. Abrams, J. Jonassaint, C. Jonassaint, N. Shah *et al.*, "Use of mobile health apps and wearable technology to assess changes and predict pain during treatment of acute pain in sickle cell disease: Feasibility study," *JMIR mHealth and uHealth*, vol. 7, no. 12, p. e13671, 2019.

[20] U. M. Al-Saggaf, S. F. Naqvi, M. Moinuddin, S. A. Alfakeh, and S. S. A. Ali, "Performance evaluation of eeg based mental stress assessment approaches for wearable devices," *Frontiers in Neurorobotics*, vol. 15, p. 819448, 2022.

[21] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study," *Sensors*, vol. 19, no. 8, p. 1849, 2019.

[22] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, "Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches," *IEEE Access*, vol. 8, pp. 38 146–38 163, 2020.

[23] D.-K. Nguyen, C.-L. Chan, A.-H. A. Li, D.-V. Phan, and C.-H. Lan, "Decision support system for the differentiation of schizophrenia and mood disorders using multiple deep learning models on wearable devices data," *Health Informatics Journal*, vol. 28, no. 4, p. 14604582221137537, 2022.

[24] A. Apicella, P. Arpaia, F. Isgrò, G. Mastrati, and N. Moccaldi, "A survey on eeg-based solutions for emotion recognition with a low number of channels," *IEEE Access*, vol. 10, pp. 117 411–117 428, 2022.

[25] A. G. Horwitz, S. D. Kentopp, J. Cleary, K. Ross, Z. Wu, S. Sen, and E. K. Czyz, "Using machine learning with intensive longitudinal data to predict depression and suicidal ideation among medical interns over time," *Psychological Medicine*, pp. 1–8, 2022.

[26] A. Ahmed, J. Ramesh, S. Ganguly, R. Aburukba, A. Sagahyroon, and F. Aloul, "Investigating the feasibility of assessing depression severity and valence-arousal with wearable sensors using discrete wavelet transforms and machine learning," *Information*, vol. 13, no. 9, p. 406, 2022.

[27] T. Mullick, A. Radovic, S. Shaaban, A. Doryab *et al.*, "Predicting depression in adolescents using mobile and wearable sensors: Multimodal machine learning–based exploratory study," *JMIR Formative Research*, vol. 6, no. 6, p. e35807, 2022.

[28] S. Fukuda, Y. Matsuda, Y. Tani, Y. Arakawa, and K. Yasumoto, "Predicting depression and anxiety mood by wrist-worn sleep sensor," in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2020, pp. 1–6.

[29] M. Zanetti, T. Mizumoto, L. Faes, A. Fornaser, M. De Cecco, L. Maule, M. Valente, and G. Nollo, "Multilevel assessment of mental stress via network physiology paradigm using consumer wearable devices," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 4, pp. 4409–4418, 2021.

[30] J. Han, Z. Zhang, C. Mascolo, E. André, J. Tao, Z. Zhao, and B. W. Schuller, "Deep learning for mobile mental health: Challenges and recent advances," *IEEE Signal Processing Magazine*, vol. 38, no. 6, pp. 96–105, 2021.

[31] M. Kaushik, S. Pathak, P. Bhatt, S. Singh, A. Tripathi, and P. Pandey, "Artificial intelligence (ai)," *Intelligent system algorithms and applications in science and Technology*, pp. 119–133, 2022.

[32] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, "Artificial intelligence in radiology," *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

[33] K. El Bouchefry and R. S. de Souza, "Learning in big data: Introduction to machine learning," in *Knowledge discovery in big data from astronomy and earth observation*. Elsevier, 2020, pp. 225–249.

[34] J. M. Banda, M. Seneviratne, T. Hernandez-Boussard, and N. H. Shah, "Advances in electronic phenotyping: from rule-based definitions to machine learning models," *Annual review of biomedical data science*, vol. 1, pp. 53–68, 2018.

[35] X.-S. Yang, *Introduction to algorithms for data mining and machine learning*. Academic press, 2019.

[36] N. Pitropakis, E. Panaousis, T. Giannetsos, E. Anastasiadis, and G. Loukas, "A taxonomy and survey of attacks against machine learning," *Computer Science Review*, vol. 34, p. 100199, 2019.

[37] M. A. Myszczynska, P. N. Ojamies, A. M. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook, and L. Ferraiuolo, "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Reviews Neurology*, vol. 16, no. 8, pp. 440–456, 2020.

[38] J.-K. Kim, M.-N. Bae, K. B. Lee, and S. G. Hong, "Identification of patients with sarcopenia using gait parameters based on inertial sensors," *Sensors*, vol. 21, no. 5, p. 1786, 2021.

[39] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020.

[40] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 4765–4774.

[41] S. Mangalathu, S.-H. Hwang, and J.-S. Jeon, "Failure mode and effects analysis of rc members based on machine-learning-based shapley additive explanations (shap) approach," *Engineering Structures*, vol. 219, p. 110927, 2020.

[42] E. Dillon, J. LaRiviere, S. Lundberg, J. Roth, and V. Syrgkanis, "Be careful when interpreting predictive models in search of causal insights," 2022, https://shap.readthedocs.io/en/latest/overviews.html.

[43] M. Prosperi, Y. Guo, M. Sperrin, J. S. Koopman, J. S. Min, X. He, S. Rich, M. Wang, I. E. Buchan, and J. Bian, "Causal inference and counterfactual prediction in machine learning for actionable healthcare," *Nature Machine Intelligence*, vol. 2, no. 7, pp. 369–375, 2020.

[44] Z. Jiang, L. Lin, X. Zhang, J. Luan, R. Zhao, L. Chen, J. Lam, K.-M. Yip, H.-K. So, W. H. Wong *et al.*, "A data-driven context-aware health inference system for children during school closures," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, pp. 1–26, 2023.

[45] G. WHO, "Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-ncov)," *World Health Organization*, 2020.

[46] M. N. Alam, A. Garg, T. T. K. Munia, R. Fazel-Rezai, and K. Tavakolian, "Vertical ground reaction force marker for parkinson's disease," *PloS one*, vol. 12, no. 5, p. e0175951, 2017.

[47] J. Camps, A. Sama, M. Martin, D. Rodriguez-Martin, C. Perez-Lopez, J. M. M. Arostegui, J. Cabestany, A. Catala, S. Alcaine, B. Mestre *et al.*, "Deep learning for freezing of gait detection in parkinson's disease patients in their homes using a waist-worn inertial measurement unit," *Knowledge-Based Systems*, vol. 139, pp. 119–131, 2018.

[48] F. Demrozi, R. Bacchin, S. Tamburin, M. Cristani, and G. Pravadelli, "Toward a wearable system for predicting freezing of gait in people affected by parkinson's disease," *IEEE journal of biomedical and health informatics*, vol. 24, no. 9, pp. 2444–2451, 2019.

[49] D. Ireland, Z. Wang, R. Lamont, and J. Liddle, "Classification of movement of people with parkinsons disease using wearable inertial movement units and machine learning," in *Digital Health Innovation for Consumers, Clinicians, Connectivity and Community*. IOS Press, 2016, pp. 61–66.

[50] V. Mikos, C.-H. Heng, A. Tay, S.-C. Yen, N. S. Y. Chia, K. M. L. Koh, D. M. L. Tan, and W. L. Au, "A wearable, patient-adaptive freezing of gait detection system for biofeedback cueing in parkinson's disease," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 3, pp. 503–515, 2019.

[51] J. C. Perez-Ibarra, A. A. Siqueira, and H. I. Krebs, "Identification of gait events in healthy and parkinson's disease subjects using inertial sensors: A supervised learning approach," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14 984–14 993, 2020.

[52] ——, "Identification of gait events in healthy subjects and with parkinson's disease using inertial sensors: An adaptive unsupervised learning approach," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 12, pp. 2933–2943, 2020.

[53] J. Zia, A. Tadayon, T. McDaniel, and S. Panchanathan, "Utilizing neural networks to predict freezing of gait in parkinson's patients," in *Proceedings of the 18th international ACM SIGACCESS conference on computers and accessibility*, 2016, pp. 333–334.

[54] G. Belgiovine, M. Capecci, L. Ciabattoni, M. Fiorentino, A. Montcriù, L. Pepa, and L. Romeo, "Upper limbs dyskinesia detection and classification for patients with parkinson's disease based on consumer electronics devices," in *2018 Zooming Innovation in Consumer Technologies Conference (ZINC)*. IEEE, 2018, pp. 156–157.

[55] J. D. Loaiza Duque, A. M. González-Vargas, A. J. Sánchez Egea, and H. A. González Rojas, "Using machine learning and accelerometry data for differential diagnosis of parkinson's disease and essential tremor," in *Workshop on Engineering Applications*. Springer, 2019, pp. 368–378.

[56] H. Jeon, W. Lee, H. Park, H. J. Lee, S. K. Kim, H. B. Kim, B. Jeon, and K. S. Park, "High-accuracy automatic classification of parkinsonian tremor severity using machine learning method," *Physiological measurement*, vol. 38, no. 11, p. 1980, 2017.

[57] ——, "Automatic classification of tremor severity in parkinson's disease using a wearable device," *Sensors*, vol. 17, no. 9, p. 2067, 2017.

[58] H. B. Kim, W. W. Lee, A. Kim, H. J. Lee, H. Y. Park, H. S. Jeon, S. K. Kim, B. Jeon, and K. S. Park, "Wrist sensor-based tremor severity quantification in parkinson's disease using convolutional neural network," *Computers in biology and medicine*, vol. 95, pp. 140–146, 2018.

[59] G. Farhani, Y. Zhou, M. E. Jenkins, M. D. Naish, and A. L. Trejos, "Using deep learning for task and tremor type classification in people with parkinson's disease," *Sensors*, vol. 22, no. 19, p. 7322, 2022.

[60] N. Shawen, M. K. O'Brien, S. Venkatesan, L. Lonini, T. Simuni, J. L. Hamilton, R. Ghaffari, J. A. Rogers, and A. Jayaraman, "Role of data measurement characteristics in the accurate detection of parkinson's disease symptoms using wearable sensors," *Journal of neuroengineering and rehabilitation*, vol. 17, no. 1, pp. 1–14, 2020.

[61] N. Mahadevan, C. Demanuele, H. Zhang, D. Volfson, B. Ho, M. K. Erb, and S. Patel, "Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–12, 2020.

[62] S. Aich, K.-W. Choi, P. M. Pradhan, J. Park, and H.-C. Kim, "A performance comparison based on machine learning approaches to distinguish parkinson's disease from alzheimer disease using spatiotemporal gait signals," *Advanced Science Letters*, vol. 24, no. 3, pp. 2058–2062, 2018.

[63] S. Aich, J. Youn, S. Chakraborty, P. M. Pradhan, J.-h. Park, S. Park, and J. Park, "A supervised machine learning approach to detect the on/off state in parkinson's disease using wearable based gait signals," *Diagnostics*, vol. 10, no. 6, p. 421, 2020.

[64] H. Khodakarami, L. Ricciardi, M. F. Contarino, R. Pahwa, K. E. Lyons, V. J. Geraedts, F. Morgante, A. Leake, D. Paviour, A. De Angelis *et al.*, "Prediction of the levodopa challenge test in parkinson's disease using data from a wrist-worn sensor," *Sensors*, vol. 19, no. 23, p. 5153, 2019.

[65] S. M. P. Dinakarrao, A. Jantsch, and M. Shafique, "Computer-aided arrhythmia diagnosis with bio-signal processing: A survey of trends and techniques," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–37, 2019.

[66] J.-m. Kwon, K.-H. Kim, K.-H. Jeon, S. Y. Lee, J. Park, and B.-H. Oh, "Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography," *Scandinavian journal of trauma, resuscitation and emergency medicine*, vol. 28, no. 1, pp. 1–10, 2020.

[67] L. Wang and X. Zhou, "Detection of congestive heart failure based on lstm-based deep network via short-term rr intervals," *Sensors*, vol. 19, no. 7, p. 1502, 2019.

[68] M.-H. Wu, E. J. Chang, and T.-H. Chu, "Personalizing a generic ecg heartbeat classification for arrhythmia detection: a deep learning approach," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2018, pp. 92–99.

[69] A. Santos Rodrigues, R. Augustauskas, M. Lukoševičius, P. Laguna, and V. Marozas, "Deep-learning-based estimation of the spatial qrs-t angle from reduced-lead ecgs," *Sensors*, vol. 22, no. 14, p. 5414, 2022.

[70] F. Corradi, J. Buil, H. De Cannière, W. Groenendaal, and P. Vandervoort, "Real time electrocardiogram annotation with a long short term memory neural network," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2019, pp. 1–4.

[71] L. Mhamdi, O. Dammak, F. Cottin, and I. B. Dhaou, "Artificial intelligence for cardiac diseases diagnosis and prediction using ecg images on embedded systems," *Biomedicines*, vol. 10, no. 8, p. 2013, 2022.

[72] A. Walinjkar and J. Woods, "Ecg classification and prognostic approach towards personalized healthcare," in *2017 International Conference On Social Media, Wearable And Web Analytics (Social Media)*. IEEE, 2017, pp. 1–8.

[73] L. Meng, W. Tan, J. Ma, R. Wang, X. Yin, and Y. Zhang, "Enhancing dynamic ecg heartbeat classification with lightweight transformer model," *Artificial Intelligence in Medicine*, vol. 124, p. 102236, 2022.

[74] M. B. Shahnawaz and H. Dawood, "An effective deep learning model for automated detection of myocardial infarction based on ultrashort-term heart rate variability analysis," *Mathematical Problems in Engineering*, vol. 2021, 2021.

[75] M. A. Khan, "An iot framework for heart disease prediction based on mdcnn classifier," *IEEE Access*, vol. 8, pp. 34 717–34 727, 2020.

[76] S. S. Sarmah, "An efficient iot-based patient monitoring and heart disease prediction system using deep learning modified neural network," *Ieee access*, vol. 8, pp. 135 784–135 797, 2020.

[77] Y.-J. Lin, C.-W. Chuang, C.-Y. Yen, S.-H. Huang, P.-W. Huang, J.-Y. Chen, and S.-Y. Lee, "Artificial intelligence of things wearable system for cardiac disease detection," in *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. IEEE, 2019, pp. 67–70.

[78] A. Scirè, F. Tropeano, A. Anagnostopoulos, and I. Chatzigiannakis, "Fog-computing-based heartbeat detection and arrhythmia classification using machine learning," *Algorithms*, vol. 12, no. 2, p. 32, 2019.

[79] X. Zhang, M. Jiang, W. Wu, and V. H. C. de Albuquerque, "Hybrid feature fusion for classification optimization of short ecg segment in iot based intelligent healthcare system," *Neural Computing and Applications*, pp. 1–15, 2021.

[80] S. Tiwari, A. Jain, V. Sapra, D. Koundal, F. Alenezi, K. Polat, A. Alhudhaif, and M. Nour, "A smart decision support system to diagnose arrhythymia using ensembled convnet and convnet-lstm model," *Expert Systems with Applications*, vol. 213, p. 118933, 2023.

[81] J. Ramesh, Z. Solatidehkordi, R. Aburukba, and A. Sagahyroon, "Atrial fibrillation classification with smart wearables using short-term heart rate variability and deep convolutional neural networks," *Sensors*, vol. 21, no. 21, p. 7233, 2021.

[82] H. Zhang, Z. Dong, M. Sun, H. Gu, and Z. Wang, "Tp-cnn: A detection method for atrial fibrillation based on transposed projection signals with compressed sensed ecg," *Computer Methods and Programs in Biomedicine*, vol. 210, p. 106358, 2021.

[83] E. Chen, J. Jiang, R. Su, M. Gao, S. Zhu, J. Zhou, and Y. Huo, "A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation," *Heart rhythm*, vol. 17, no. 5, pp. 847–853, 2020.

[84] S. Nemati, M. M. Ghassemi, V. Ambai, N. Isakadze, O. Levantsevych, A. Shah, and G. D. Clifford, "Monitoring and detecting atrial fibrillation using wearable technology," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 3394–3397.

[85] D. Hiraoka, T. Inui, E. Kawakami, M. Oya, A. Tsuji, K. Honma, Y. Kawasaki, Y. Ozawa, Y. Shiko, H. Ueda *et al.*, "Diagnosis of atrial fibrillation using machine learning with wearable devices after cardiac surgery: Algorithm development study," *JMIR Formative Research*, vol. 6, no. 8, p. e35396, 2022.

[86] T. Pereira, C. Ding, K. Gadhoumi, N. Tran, R. A. Colorado, K. Meisel, and X. Hu, "Deep learning approaches for plethysmography signal quality assessment in the presence of atrial fibrillation," *Physiological measurement*, vol. 40, no. 12, p. 125002, 2019.

[87] M. Maritsch, C. Bérubé, M. Kraus, V. Lehmann, T. Züger, S. Feuerriegel, T. Kowatsch, and F. Wortmann, "Improving heart rate variability measurements from consumer smartwatches with machine learning," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 934–938.

[88] T. Choksatchawathi, P. Ponglertnapakorn, A. Ditthapron, P. Leelaarporn, T. Wisutthisen, M. Piriyajitakonkij, and T. Wilaiprasitporn, "Improving heart rate estimation on consumer grade wrist-worn device using post-calibration approach," *IEEE Sensors Journal*, vol. 20, no. 13, pp. 7433–7446, 2020.

[89] G. R. Nidigattu, G. Mattela, and S. Jana, "Non-invasive modeling of heart rate and blood pressure from a photoplethysmography by using machine learning techniques," in *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2020, pp. 7–12.

[90] O. Akrivopoulos, D. Amaxilatis, I. Mavrommati, and I. Chatzigiannakis, "Utilising fog computing for developing a person-centric heart

monitoring system," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 3, pp. 237–259, 2019.

[91] L. Dagher, H. Shi, Y. Zhao, and N. F. Marrouche, "Wearables in cardiology: Here to stay," *Heart Rhythm*, vol. 17, no. 5, pp. 889–895, 2020.

[92] T. Cho, U. Sunarya, M. Yeo, B. Hwang, Y. S. Koo, and C. Park, "Deep-actinet: End-to-end deep learning architecture for automatic sleep-wake detection using wrist actigraphy," *Electronics*, vol. 8, no. 12, p. 1461, 2019.

[93] A. Arora, P. Chakraborty, and M. Bhatia, "Analysis of data from wearable sensors for sleep quality estimation and prediction using deep learning," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10 793–10 812, 2020.

[94] T. Wang, C. Lu, G. Shen, and F. Hong, "Sleep apnea detection from a single-lead ecg signal with automatic feature-extraction through a modified lenet-5 convolutional neural network," *PeerJ*, vol. 7, p. e7731, 2019.

[95] G. Ye, H. Yin, T. Chen, H. Chen, L. Cui, and X. Zhang, "Fenet: a frequency extraction network for obstructive sleep apnea detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2848–2856, 2021.

[96] D. Benedetti, U. Olcese, S. Bruno, M. Barsotti, M. M. Tassoni, E. Bonanni, G. Siciliano, and U. Faraguna, "Obstructive sleep apnoea syndrome screening through wrist-worn smartbands: A machine-learning approach," *Nature and Science of Sleep*, vol. 14, p. 941, 2022.

[97] B. Camcı, C. Ersoy, and H. Kaynak, "Abnormal respiratory event detection in sleep: a prescreening system with smart wearables," *Journal of biomedical informatics*, vol. 95, p. 103218, 2019.

[98] B. Jiao, "Anti-motion interference wearable device for monitoring blood oxygen saturation based on sliding window algorithm," *IEEE Access*, vol. 8, pp. 124 675–124 687, 2020.

[99] M. A. Grandner, Z. Bromberg, A. Hadley, Z. Morrell, A. Graf, S. Hutchison, and D. Freckleton, "Performance of a multisensor smart ring to evaluate sleep: in-lab and home-based evaluation of generalized and personalized algorithms," *Sleep*, 2022.

[100] W. Hidayat, T. D. Tambunan, and R. Budiawan, "Empowering wearable sensor generated data to predict changes in individual's sleep quality," in *2018 6th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 2018, pp. 447–452.

[101] X. Li and H. Zhao, "Automated feature extraction from population wearable device data identified novel loci associated with sleep and circadian rhythms," *PLoS genetics*, vol. 16, no. 10, p. e1009089, 2020.

[102] D. M. Roberts, M. M. Schade, G. M. Mathew, D. Gartenberg, and O. M. Buxton, "Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography," *Sleep*, vol. 43, no. 7, p. zsaa045, 2020.

[103] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[104] C. Heneghan, "St. vincent's university hospital/university college dublin sleep apnea database," 2011.

[105] A. Rghioui, J. Lloret, S. Sendra, and A. Oumnad, "A smart architecture for diabetic patient monitoring using machine learning algorithms," in *Healthcare*, vol. 8, no. 3. MDPI, 2020, p. 348.

[106] P. Kaur, N. Sharma, A. Singh, and B. Gill, "Ci-dpf: a cloud iot based framework for diabetes prediction," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2018, pp. 654–660.

[107] J. Ramesh, R. Aburukba, and A. Sagahyroon, "A remote healthcare monitoring framework for diabetes prediction using machine learning," *Healthcare Technology Letters*, vol. 8, no. 3, pp. 45–57, 2021.

[108] H. Torkey, E. Ibrahim, E. Hemdan, A. El-Sayed, and M. A. Shouman, "Diabetes classification application with efficient missing and outliers data handling algorithms," *Complex & Intelligent Systems*, vol. 8, no. 1, pp. 237–253, 2022.

[109] C. Hettiarachchi and C. Chitraranjan, "A machine learning approach to predict diabetes using short recorded photoplethysmography and physiological characteristics," in *Conference on Artificial Intelligence in Medicine in Europe*. Springer, 2019, pp. 322–327.

[110] M. M. Rashid, M. R. Askari, C. Chen, Y. Liang, K. Shu, and A. Cinar, "Artificial intelligence algorithms for treatment of diabetes," *Algorithms*, vol. 15, no. 9, p. 299, 2022.

[111] R. Groh, Z. Lei, L. Martignetti, N. Y. Li-Jessen, and A. M. Kist, "Efficient and explainable deep neural networks for airway symptom detection in support of wearable health technology," *Advanced Intelligent Systems*, vol. 4, no. 7, p. 2100284, 2022.

[112] X. Hui, J. Zhou, P. Sharma, T. B. Conroy, Z. Zhang, and E. C. Kan, "Wearable rf near-field cough monitoring by frequency-time deep learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 4, pp. 756–764, 2021.

[113] B. Xue, W. Shi, S. H. Chotirmall, V. C. A. Koh, Y. Y. Ang, R. X. Tan, and W. Ser, "Distance-based detection of cough, wheeze, and breath sounds on wearable devices," *Sensors*, vol. 22, no. 6, p. 2167, 2022.

[114] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE transactions on biomedical circuits and systems*, vol. 14, no. 3, pp. 535–544, 2020.

[115] X. Yang, M. Zhang, S. Gong, M. Sun, M. Xie, P. Niu, and W. Pang, "Chest-laminated and sweat-permeable e-skin for speech and motion artifact-insensitive cough detection," *Advanced Materials Technologies*, p. 2201043, 2022.

[116] W. Zhang, J. Li, Y. Wen, and Y. Luo, "Toward a wearable crowdsource system to monitor respiratory symptoms for pandemic early warning," *Ieee Network*, vol. 35, no. 3, pp. 56–63, 2021.

[117] M. Soltanian and K. Borna, "Covid-19 recognition from cough sounds using lightweight separable-quadratic convolutional network," *Biomedical Signal Processing and Control*, vol. 72, p. 103333, 2022.

[118] F. F. Abir, K. Alyafei, M. E. Chowdhury, A. Khandakar, R. Ahmed, M. M. Hossain, S. Mahmud, A. Rahman, T. O. Abbas, S. M. Zughaier *et al.*, "Pcovnet: A presymptomatic covid-19 detection framework using deep learning model using wearables data," *Computers in biology and medicine*, vol. 147, p. 105682, 2022.

[119] N. A. Nayan, C. J. Yi, M. Z. Suboh, N.-F. Mazlan, P. Periyasamy, M. Y. Z. A. Rahim, and S. A. Shah, "Covid-19 prediction with machine learning technique from extracted features of photoplethysmogram morphology," *Frontiers in public health*, vol. 10, 2022.

[120] R. P. Hirten, L. Tomalin, M. Danieletto, E. Golden, M. Zweig, S. Kaur, D. Helmus, A. Biello, R. Pyzik, E. P. Bottinger *et al.*, "Evaluation of a machine learning approach utilizing wearable data for prediction of sars-cov-2 infection in healthcare workers," *JAMIA open*, vol. 5, no. 2, p. ooac041, 2022.

[121] H. Mankodiya, P. Palkhiwala, R. Gupta, N. K. Jadav, S. Tanwar, B.-C. Neagu, G. Grigoras, F. Alqahtani, and A. M. Shehata, "A real-time crowdsensing framework for potential covid-19 carrier detection using wearable sensors," *Mathematics*, vol. 10, no. 16, p. 2927, 2022.

[122] A. E. Mason, F. M. Hecht, S. K. Davis, J. L. Natale, W. Hartogensis, N. Damaso, K. T. Claypool, S. Dilchert, S. Dasgupta, S. Purawat *et al.*, "Detection of covid-19 using multimodal data from a wearable device: results from the first tempredict study," *Scientific reports*, vol. 12, no. 1, pp. 1–15, 2022.

[123] D. Jadav, D. Patel, R. Gupta, N. K. Jadav, and S. Tanwar, "Barcode: A blockchain-based framework for remote covid detection for healthcare 5.0," in *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2022, pp. 782–787.

[124] C.-T. Wu, S.-M. Wang, Y.-E. Su, T.-T. Hsieh, P.-C. Chen, Y.-C. Cheng, T.-W. Tseng, W.-S. Chang, C.-S. Su, L.-C. Kuo *et al.*, "A precision health service for chronic diseases: Development and cohort study using wearable device, machine learning, and deep learning," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–14, 2022.

[125] J. Skibinska and R. Burget, "Is it possible to distinguish covid-19 cases and influenza with wearable devices? analysis with machine learning," *Journal of Advances in Information Technology*, vol. 13, no. 3, 2022.

[126] I. F. Miller, A. D. Becker, B. T. Grenfell, and C. J. E. Metcalf, "Disease and healthcare burden of covid-19 in the united states," *Nature Medicine*, vol. 26, no. 8, pp. 1212–1217, 2020.

[127] H. Rizk, T. Amano, H. Yamaguchi, and M. Youssef, "Smartwatch-based face-touch prediction using deep representational learning," in *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer, 2021, pp. 493–499.

[128] E. Lattanzi, L. Calisti, and V. Freschi, "Unstructured handwashing recognition using smartwatch to reduce contact transmission of pathogens," *IEEE Access*, vol. 10, pp. 83 111–83 124, 2022.

[129] F. Forooghifar, A. Aminifar, and D. Atienza, "Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 6, pp. 1338–1350, 2019.

[130] X. H. Kok, S. A. Imtiaz, and E. Rodriguez-Villegas, "Assessing the feasibility of acoustic based seizure detection," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 7, pp. 2379–2389, 2022.

[131] R. E. Stirling, D. B. Grayden, W. D'Souza, M. J. Cook, E. Nurse, D. R. Freestone, D. E. Payne, B. H. Brinkmann, T. Pal Attia, P. F. Viana *et al.*, "Forecasting seizure likelihood with wearable technology," *Frontiers in neurology*, p. 1170, 2021.

[132] F. M. Garcia-Moreno, M. Bermudez-Edo, J. L. Garrido, E. Rodríguez-García, J. M. Pérez-Mármol, and M. J. Rodríguez-Fórtiz, "A microservices e-health system for ecological frailty assessment using wearables," *Sensors*, vol. 20, no. 12, p. 3427, 2020.

[133] A. Leone, G. Rescio, A. Manni, P. Siciliano, and A. Caroppo, "Comparative analysis of supervised classifiers for the evaluation of sarcopenia using a semg-based platform," *Sensors*, vol. 22, no. 7, p. 2721, 2022.

[134] I. Chen, P.-Y. Yeh, T.-C. Chang, Y.-C. Hsieh, and C.-L. Chin, "Sarcopenia recognition system combined with electromyography and gait obtained by the multiple sensor module and deep learning algorithm," *Sensors and Materials*, vol. 34, no. 6, pp. 2403–2425, 2022.

[135] T. Wang, W. Jin, F. Liang, and J. Alastruey, "Machine learning-based pulse wave analysis for early detection of abdominal aortic aneurysms using in silico pulse waves," *Symmetry*, vol. 13, no. 5, p. 804, 2021.

[136] S. Murali, F. Rincon, and D. Atienza, "A wearable device for physical and emotional health monitoring," in *2015 Computing in Cardiology Conference (CinC)*. IEEE, 2015, pp. 121–124.

[137] M. Ihle, H. Feldwisch-Drentrup, C. A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage, "Epilepsiae–a european epilepsy database," *Computer methods and programs in biomedicine*, vol. 106, no. 3, pp. 127–138, 2012.

[138] Q.-L. Xue, "The frailty syndrome: definition and natural history," *Clinics in geriatric medicine*, vol. 27, no. 1, pp. 1–15, 2011.

[139] X. Chen, G. Mao, and S. X. Leng, "Frailty syndrome: an overview," *Clinical interventions in aging*, vol. 9, p. 433, 2014.

[140] C. Dwork, "Differential privacy: A survey of results," in *International conference on theory and applications of models of computation*. Springer, 2008, pp. 1–19.

[141] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[142] L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Over-the-air computation for iot networks: Computing multiple functions with antenna arrays," *IEEE Internet of Things journal*, vol. 5, no. 6, pp. 5296–5306, 2018.

[143] K. Arakadakis, P. Charalampidis, A. Makrogiannakis, and A. Fragkiadakis, "Firmware over-the-air programming techniques for iot networks-a survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 9, pp. 1–36, 2021.

[144] N. Yan, K. Wang, C. Pan, and K. K. Chai, "Private federated learning with misaligned power allocation via over-the-air computation," *IEEE Communications Letters*, vol. 26, no. 9, pp. 1994–1998, 2022.

[145] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE transactions on wireless communications*, vol. 19, no. 3, pp. 2022–2035, 2020.

[146] T. Sery, N. Shlezinger, K. Cohen, and Y. C. Eldar, "Over-the-air federated learning from heterogeneous data," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3796–3811, 2021.

[147] L. Cilliers, "Wearable devices in healthcare: Privacy and information security issues," *Health information management journal*, vol. 49, no. 2-3, pp. 150–156, 2020.

[148] K. W. Ching and M. M. Singh, "Wearable technology devices security and privacy vulnerability analysis," *International Journal of Network Security & Its Applications*, vol. 8, no. 3, pp. 19–30, 2016.

[149] M. Guerar, L. Verderame, A. Merlo, F. Palmieri, M. Migliardi, and L. Vallerini, "Circlepin: a novel authentication mechanism for smartwatches to prevent unauthorized access to iot devices," *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 3, pp. 1–19, 2020.

[150] D. Bhushan and R. Agrawal, "Security challenges for designing wearable and iot solutions," *A handbook of internet of things in biomedical and cyber physical system*, pp. 109–138, 2020.

[151] M. Haghi, S. Danyali, S. Ayasseh, J. Wang, R. Aazami, and T. M. Deserno, "Wearable devices in health monitoring from the environmental towards multiple domains: A survey," *Sensors*, vol. 21, no. 6, p. 2130, 2021.

[152] W. Lalouani, M. Younis, I. White-Gittens, R. N. Emokpae Jr, and L. E. Emokpae, "Energy-efficient collection of wearable sensor data through predictive sampling," *Smart Health*, vol. 21, p. 100208, 2021.

[153] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE communications surveys & tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[154] V. P. Rekkas, S. Sotiroudis, P. Sarigiannidis, S. Wan, G. K. Karagiannidis, and S. K. Goudos, "Machine learning in beyond 5g/6g networks—state-of-the-art and future trends," *Electronics*, vol. 10, no. 22, p. 2786, 2021.

[155] S. Zhang and D. Zhu, "Towards artificial intelligence enabled 6g: State of the art, challenges, and opportunities," *Computer Networks*, vol. 183, p. 107556, 2020.

[156] W. Deng, E. C. Ngai, and V. Van Zoest, "Energy-efficient monitoring of potential side effects from covid-19 vaccines," in *2022 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*. IEEE, 2022, pp. 222–227.

[157] J. L. Guo, M. Januszyk, and M. T. Longaker, "Machine learning in tissue engineering," *Tissue Engineering Part A*, vol. 29, no. 1-2, pp. 2–19, 2023.

[158] A. Polino, R. Pascanu, and D. Alistarh, "Model compression via distillation and quantization," in *International Conference on Learning Representations*, 2018.

[159] M. Haroush, I. Hubara, E. Hoffer, and D. Soudry, "The knowledge within: Methods for data-free model compression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8494–8502.

[160] J. H. Cho and B. Hariharan, "On the efficacy of knowledge distillation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 4794–4802.

[161] L. Wang and K.-J. Yoon, "Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 6, pp. 3048–3068, 2021.

[162] Z. Jiang, H. Chen, R. Zhou, J. Deng, X. Zhang, R. Zhao, C. Xie, Y. Wang, and E. C. Ngai, "Healthprism: A visual analytics system for exploring children's physical and mental health profiles with multimodal data," *arXiv preprint arXiv:2307.12242*, 2023.

[163] J.-H. Kim, C. Marcus, R. Ono, D. Sadat, A. Mirzazadeh, M. Jens, S. Fernandez, S. Zheng, T. Durak, and C. Dagdeviren, "A conformable sensory face mask for decoding biological and environmental signals," *Nature Electronics*, vol. 5, no. 11, pp. 794–807, 2022.

[164] R. Bagutdinov and M. Stepanov, "Methods of integration, reduction of sizes and normalization of processing of heterogeneous and multi-scale data," *International Journal of Open Information Technologies*, vol. 9, no. 2, pp. 39–44, 2021.

[165] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: homogeneous and heterogeneous approaches," *Knowledge-Based Systems*, vol. 118, pp. 124–139, 2017.

[166] Q. Cai, H. Wang, Z. Li, and X. Liu, "A survey on multimodal data-driven smart healthcare systems: approaches and applications," *IEEE Access*, vol. 7, pp. 133 583–133 599, 2019.

[167] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[168] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision*, vol. 129, pp. 1789–1819, 2021.

[169] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A comprehensive survey on model compression and acceleration," *Artificial Intelligence Review*, vol. 53, pp. 5113–5155, 2020.

[170] K. T. Hall, J. Loscalzo, and T. J. Kaptchuk, "Genetics and the placebo effect: the placebome," *Trends in molecular medicine*, vol. 21, no. 5, pp. 285–294, 2015.

[171] M. G. Helander, *Handbook of human-computer interaction*. Elsevier, 2014.