1 Results

A total of 26 studies, concerned with seizure detection, were included in this review (Table number). These studies utilized a range of commercial and non-commercial devices for seizure detection. The most frequently used commercial device was the EMPATICA E4, which was tested in 5 studies [1, 2, 3, 4]. Additional studies employed either customized laboratory-developed devices or other commercial devices, such as: Shimmer [5, 6], Samsung SM-R800 watch [7], Mi- Microsoft wristband[8], Nightwatch [9], and Biovital P1 System [10, 11].

All included studies assessed various combinations of the following physiological and movement sensors and measures: Accelerometer (ACC), Blood Volume Pulse (BVP), Gyroscope (Gyro), Electromyography (EMG), Heart rate (HR), Oxygen Saturation (SPO2), Electrodermal activity (EDA), electrocardiography (ECG), temperature, and Heart rate variability (HRV).

For seizure prediction and forecasting, three studies were found eligible for this review (table number. All three papers used the Empatica E4 wristband device for wearable data acquisition, which includes sensors for electrodermal activity (EDA), blood volume pulse (BVP) via photoplethysmography, 3-axis accelerometry (ACC), and skin temperature; however, the selected modalities for training and analysis differed among the papers.

The number of patients included in data acquisition ranged from 42 [12] to 139 [13] with a mean age ranging from 9.8 [14] to 14 [12] years old, indicating a pediatric-dominant cohort. Two studies [12, 13] included a control group (patients with no seizures) besides the seizure group to differentiate seizure-related patterns from normal brain activity. All studies were conducted in an in-hospital setting.

1.1 Demographics

Reviewed studies included adults only, pediatrics only, or a mixed population as follows: Five studies [15, 16, 17, 7, 18] focused solely on adults aged 20 [17] to 64 years [15] . From which, One Study provided the mean age only [16], and one also providing the standard deviation [7]. Eleven (42.3%) studies [1, 19, 20, 21, 22, 10, 6, 23, 11, 8, 4] were conducted exclusively on pediatric patients aged between 1 month [8] and 17 years [20]. Among these, Two studies [19, 21] did not specify exact ages but indicated inclusion of children, while four provided both mean and median ages only [1, 22, 23, 4]. Ten studies [5, 24, 2, 3, 9, 25, 26, 27, 28, 29] included mixed-age populations, with ages spanning from 2 [5] to 75 years[25], including two studies that did not specify the age range [3, 28]. In addition to seizure patients, Five studies included control groups to test the performance of their seizure detection methods [5, 24, 7, 25, 29]. Sample sizes varied widely, ranging from 7 [21] to 166 participants [1] in pediatric studies, from 5 patients [18] to 36 [7] in adult studies, and from 4 patients [27] to 135 patients [2] in the mixed-age population. Of the reviewed studies, twenty (76.9%) were conducted in inpatient settings, Three studies in outpatient settings [24, 17, 26], and three were conducted in both settings [2, 3, 29]. Nine (34.62%) studies targeted generalized tonic-clonic and/or tonic-clonic seizures only [19, 24, 21, 22, 7, 25, 27, 28, 29]. Four (15.38%) studies focused solely on nocturnal seizures [5, 21, 6, 25].

Table 1: Modalities

Modality	*Sensitivity*	*FAR/24H*	*Accuracy*	*Studies*
ACC, GYR, sEMG, EDA	[81.69%–95.24%]	[0.64–1.21]	[93.16%–96.81%]	[11, 23, 27, 10, 24]
ACC, PPG	[80%–86%]	[0.2609–13.63]	—	[18, 4, 9, 1]
ACC, EDA, PPG	[93%–100%]	[0.08–2.339]	—	[28, 3]
ACC, ECG	[87%–92%]	—	—	[5, 20]
ACC, GYR	[76.84%–96%]	[0.98]	[97.28%]	[25, 26]
ACC, GYR, sEMG	[97%–100%]	—	—	[29, 6]
ACC, sEMG	[90.91%]	—	—	[19]
ACC, EDA	[93.9%–97.2%]	[0.53–1.8]	[96.7%]	[2, 22, 17]
ACC, ECG, sEMG	[90.9%]	—	—	[21]
ACC, ECG, EDA, sEMG	—	—	—	[16]
ACC, GYR, PPG	[87%]	[0.21]	[93%]	[7]
ACC, EDA, GYR, PPG	[89%]	[0.54]	—	[8]
EDA, PPG	[100%]	—	—	[15]

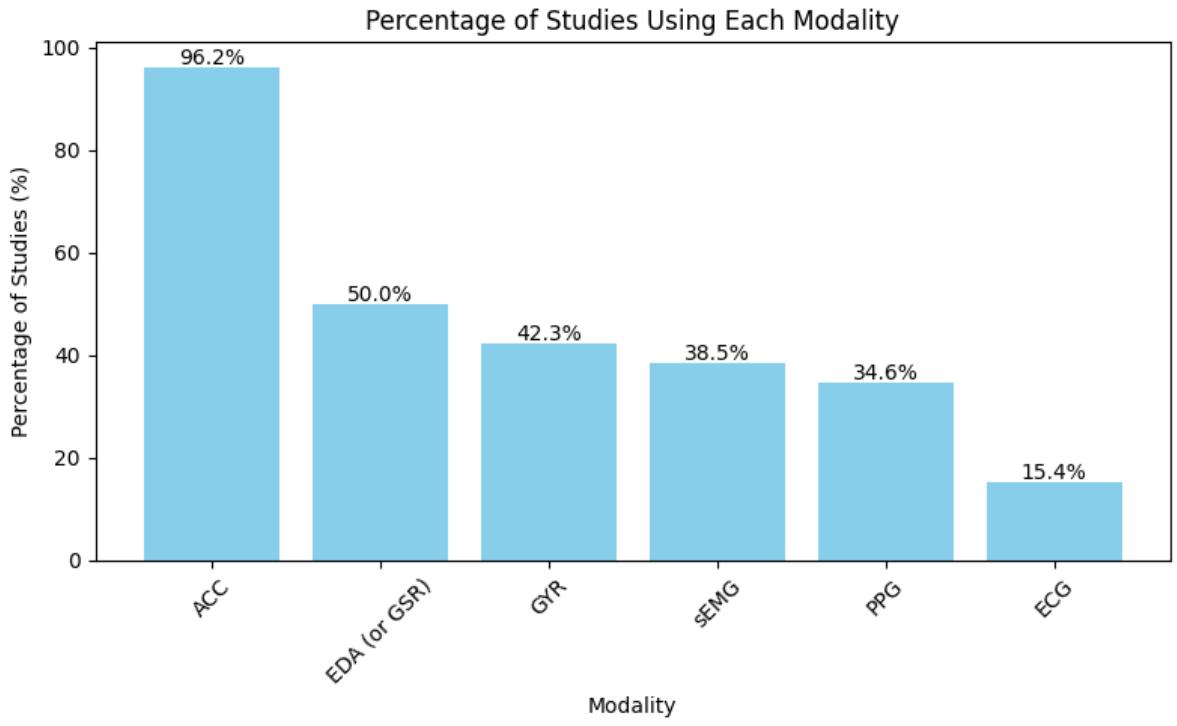

Figure 1: Percentage of Studies Using Each Modality. Please note that studies used multiple sensors, however, the combination is not shown in this graph

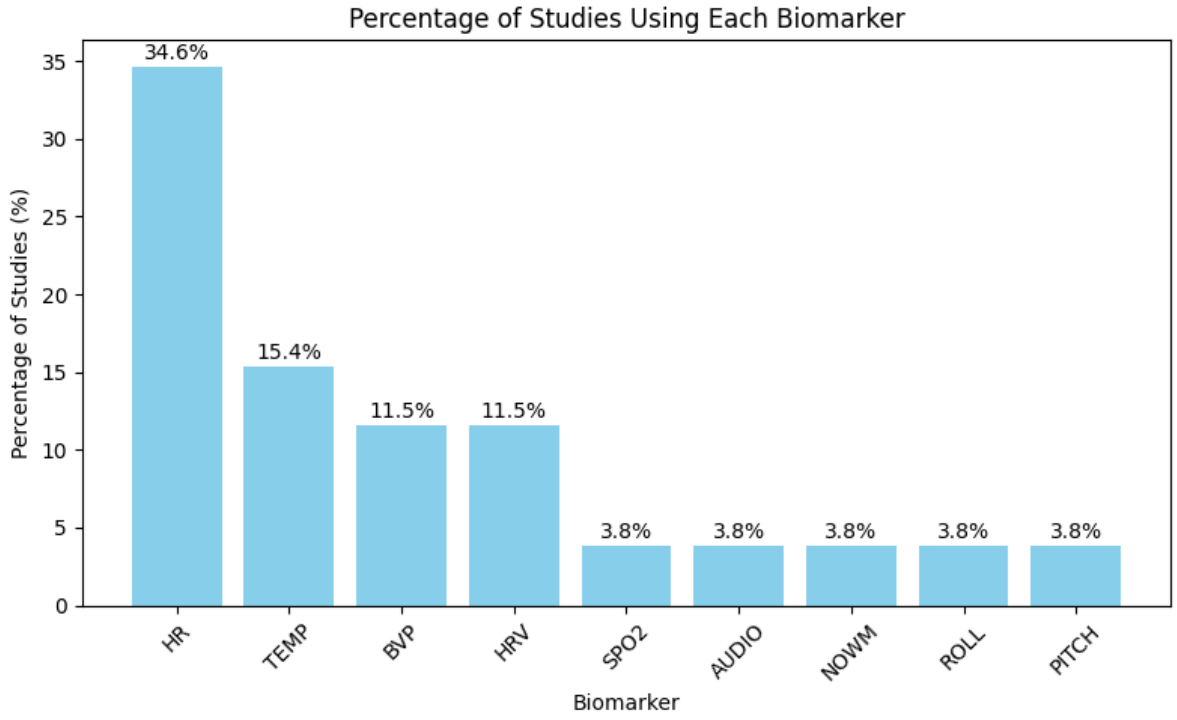

Figure 2: Percentage of Studies Using Each Biomarker

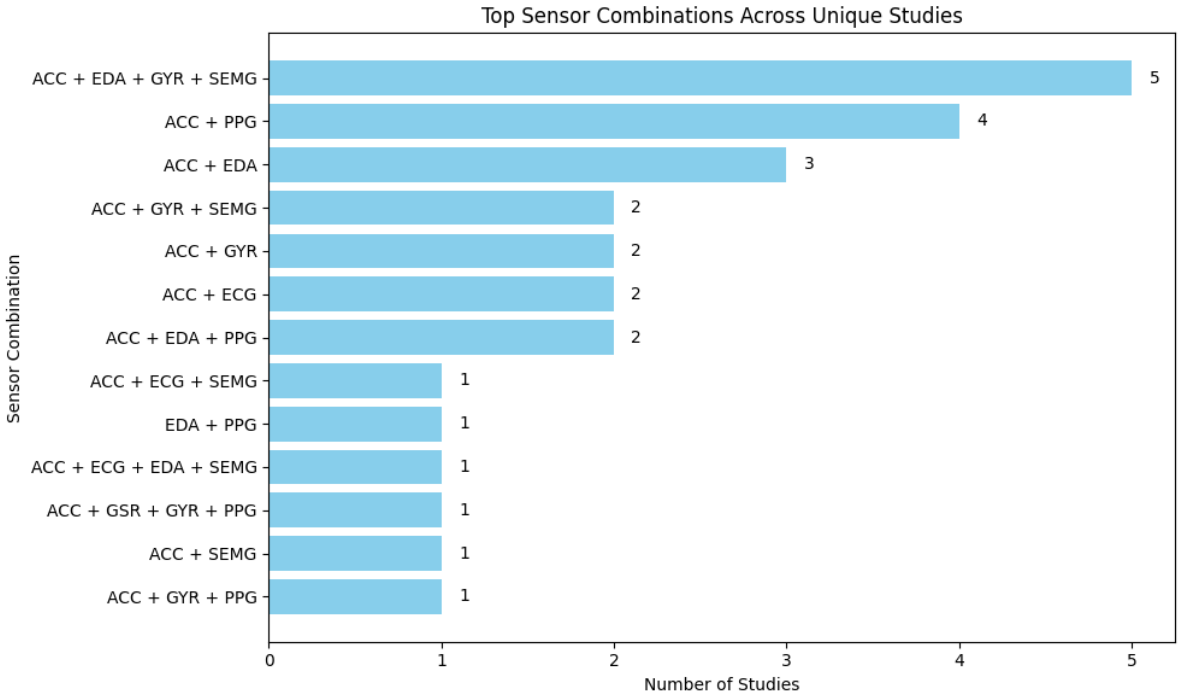

Figure 3: Frequency of Each Sensor Combination

1.2 Modalities

1.2.1 Detection

Among the 26 reviewed detection studies, ACC was by far the most frequently used modality (96.2%) to capture the convulsive motor activity associated with tonic-clonic seizures. EDA followed as the second most used modality, appearing in half of the studies, while ECG was the least commonly used (15.4%) (Figure 1).

Several studies went beyond raw sensor signals and extracted specific biomarkers such as HR, HRV, BVP, and SpO_2, which were then used as features in seizure detection models. HR was the most commonly used biomarker (34.6%), while SpO_2, audio, NOWM, and pitch and roll motion were each used in only 3.8% of the studies (Figure 2).

Altogether, 13 different multimodal sensor combinations were reported. The most common was ACC + GYR + sEMG + EDA (19.2%; Figure 3). Almost all studies that directly compared unimodal and multimodal systems [1, 19, 21, 17, 23, 29, 4, 27, 20, 22, 16, 10] found that multimodal systems outperformed unimodal ones. The only exception was [20], where a cardiac algorithm using ECG alone achieved a lower FPR (1 per day) compared to ACC + ECG (2 per day).

The most commonly used multimodal sensor combination, ACC + GYR + sEMG + EDA [11, 23, 27, 10, 24] consistently achieved high performance, with the highest reported in [10]. Among these studies, Wang et al. [11] investigated the use of derived biomarkers (PITCH and ROLL, extracted from ACC and GYR) instead of or in combination with raw ACC and GYR data. They found that substituting ACC with PITCH or ROLL improved performance across all models, with the best results obtained using SVM (Accuracy: 95.7%, Precision: 95.7%, Recall: 93.8%; for ROLL: Accuracy: 95.2%, Precision: 95.7%, Recall: 92.5%) compared to the original combination which achieved an accuracy of 93.4%, precision of 95.8%, and recall of 90.9%.

Another widely used combination was ACC + PPG [18, 4, 9, 1]. Yu et al. [1] reported that for generalized motor seizures, ACC + BVP achieved the best performance with a mean AUC-ROC of 0.805, while EDA had the lowest mean AUC-ROC of 0.513. For tonic–clonic seizures, ACC alone achieved the best performance, with an AUC-ROC of 0.973, a sensitivity of 95%, and FPR of 6.2%. Similarly, Tang et al. [4] reported that when ACC data was used alone, this model performed best for tonic-clonic seizures with an AUC-ROC of 0.995, however, for seizure type agnostic, ACC + BVP fusion performed best, with an AUC-ROC of 0.752. Another study [9], an in-home nocturnal cohort study, reported that their sensitivity was significantly high (median 85%) compared to a rhythmic movement-based bed sensor (median 21%). They further analyzed feature contributions and showed that HR was the critical modality for true positives (92%) and also for false positives, while ACC contributed only 8% of true positives and caused no false alarms.

ACC + EDA was also one of the top and high achieving combinations [2, 22, 17] with the highest sensitivity (97.2%) reported by [17] and the lowest FAR/24h (0.53) reported by [2], which validated Empatica multimodal wristbands (E4 and Embrace) as reliable tools for GTCS detection in real-world and Epilepsy Monitoring Units (EMUs) settings. Chowdhury et al. [17] reported that fusing ACC and EDA significantly improved classification accuracy (96.7%) and reduced FAR (unspecified) compared to unimodal approaches. Similarly, Poh et al. [22] reported that the overall performance was lower when only ACC features were included.

While most studies incorporate both physiological and motion-based sensors in their

sensor suite, 19.2% of the studies were purely motion-based, using sensor combinations of ACC + GYR [25, 26], ACC + GYR + sEMG [29, 6] and ACC + sEMG [19]. Among these, the ACC + sEMG + GYR configuration achieved the best performance. In addition to studying their significance in seizure detection, a few studies have tested for the optimal sensor placement [19, 21], showing that using sensors on different body locations can reduce false alarms and improve performance. [19] identified the left wrist (generally the non-dominant hand) and right ankle as optimal positions for ACC sensors, while bilateral biceps were optimal for sEMG.

For details of the remaining reported combinations, refer to the Modality Table.

1.2.2 Prediction and Forecasting

All three studies used EDA and PPG in their sensor suite. One seizure forecasting study additionally used ACC [14]. PPG was used to extract biomarkers like HR in [12, 13], HRV in [13], and BVP in [14]. TEMP was also among the biomarkers used by [13, 14], however in [13], TEMP did not differ between seizure and non-seizure groups and was therefore not included in further analysis. Vieluf et al. [12, 13] identified EDA and HR as containing sufficient seizure-predictive information, with reported performance of 62% sensitivity and an accuracy range of 68–68.89%. These findings were further supported in [13], where HRV was additionally shown to carry predictive value: patients with an impending seizure had lower HR and higher HRV compared to seizure-free patients in evening recordings. In [14], forecasting performance was highest when all modalities (EDA, BVP, TEMP, ACC) were combined achieving significant seizure forecasting (better-than-chance) in 43% of patients (30/69), where the mean sensitivity was 75.6%, the mean time in warning (TiW) was 47.2% and the mean prediction horizon was 31.6 minutes. Each modality contributed uniquely, though accelerometry sometimes reduced performance in low-performing patients.

1.3 Preprocessing

1.3.1 Signal Synchronization and Quality Control

A fundamental preprocessing step reported in the literature was the temporal alignment of data from wearable devices with a ground-truth reference, typically video-electroencephalography (video-EEG), to ensure accurate event labeling. Four studies described specific synchronization methods [1, 7, 9, 4]. The reported techniques included manual clock synchronization at the start and end of recordings to correct for time drift [1], the use of a Network Time Protocol (NTP) for automated alignment [7], a three-step timing error compensation algorithm [4], and the synchronization of sensor data with event logs from video recordings [9].

Following synchronization, quality control was performed to ensure data integrity. Ten studies detailed their quality control procedures [1, 27, 9, 8, 29, 3, 4, 18, 23, 20]. A common strategy was the removal of data segments where the device was not worn. Two studies accomplished this using temperature data, excluding periods where sensed temperature fell outside a physiological range [1, 4]. Other methods involved trimming the initial and final 15 minutes of recordings to account for sensor calibration [1] or implementing signal-specific quality checks. For instance, heart rate data were only utilized if a signal quality index exceeded an 80% threshold [9], while data from segments with poor

electrodermal activity (EDA) contact were discarded [27]. Other studies also reported using custom quality indices to clean data epochs before analysis [3, 29].

Furthermore, some protocols involved removing sliding windows with a high percentage (¿60%) of missing data [8] or conducting manual reviews to identify and exclude recordings with failures or invalid annotations [1]. General artifact removal and signal cleaning were also explicitly mentioned in several other papers [23, 20, 18].

1.3.2 Noise and Artifact Removal

Majority of the studies (16 papers) reported applying specific signal processing techniques to remove noise and artifacts from the raw sensor data [19, 21, 10, 27, 29, 16, 17, 6, 23, 28, 11, 8, 1, 26, 7, 20]. Filtering was the most common approach, with techniques tailored to the specific signal modality. For accelerometer (ACC) and gyroscope (Gyro) data, band-pass filters were frequently used, with cutoff frequencies such as 1–24 Hz [10] or 0.2–47 Hz [21], to isolate movement patterns relevant to seizures. High-pass filters were also applied to remove baseline drift or focus on significant movements [19, 27, 7, 29]. In addition to filtering, smoothing techniques like a five-point median filter [28] or a 3-second smoothing filter [15] were employed to reduce erratic signal fluctuations.

For other modalities, filtering strategies were similarly targeted. Surface electromyography (sEMG) signals were commonly processed with a high-pass filter (e.g., at 20 Hz) to remove motion artifacts and baseline noise [19, 21, 27] or a band-pass filter (e.g., 20-90 Hz) [10]. Electrodermal activity (EDA) signals were often smoothed using a mean or median filter to reduce noise [27, 11, 1]. Furthermore, a study reported the use of wavelet transformation for denoising multiple signal types, including heart rate (HR), EDA, and ACC [8], while others used notch filters to eliminate specific sources of interference like powerline noise [19, 16]. In the context of seizure prediction and forecasting, noise removal was also a noted preprocessing step. One study focusing on seizure prediction applied low-pass filtering specifically to the EDA signal to refine the data before analysis [12]. This highlights that while the ultimate goal (detection vs. prediction) differs, the foundational need to clean and denoise raw sensor signals remains a consistent and critical practice across the field.

1.3.3 Data Segmentation and Windowing

A nearly universal preprocessing step reported in the literature was the segmentation of continuous time-series data into fixed-length windows, a necessary step for feature extraction in most machine learning models. This technique was detailed in 18 of the detection studies [19, 21, 15, 16, 24, 22, 10, 3, 7, 25, 26, 27, 28, 11, 8, 29, 20, 17] and was also a key component of all of the 3 forecasting and prediction studies [14, 13, 12].

The duration of these windows varied considerably, often tailored to the physiological signals being analyzed. For capturing the rapid motor patterns characteristic of tonic-clonic seizures, shorter window sizes ranging from 2 to 6 seconds were frequently employed [19, 21, 26, 15, 16, 10, 28, 11, 29]. A 10-second window was also a common choice [22, 3, 25, 27]. While two studies used much larger segments, ranging from 5 to 7 minutes [8, 7].

To ensure that seizure events occurring at the boundary of a window were not missed and to augment the volume of training data, studies implemented overlapping windows. The degree of overlap typically ranged from 50% [16, 26, 11, 7] to as high as 90% [24, 25]. Overlaps of 75-80% were also frequently reported [19, 21, 15, 22, 10, 8].

All reviewed forecasting and prediction studies used data preprocessing segmentation; however, segmentation methods varied across the studies. The segmentation strategies ranged from short, fixed-length epochs (30 seconds) and brief daily snapshots (15 minutes) to longer, seizure-centered windows (45 minutes). For example, one study [14] segmented continuous wristband sensor data into 30-second epochs, extracting features from each segment to train and evaluate their seizure forecasting model. Another study [13] utilized a single 15-minute autonomic recording taken in the evening (9:00–9:15 pm) as a fixed snapshot for seizure likelihood assessment, combining this with clinical data for prediction. In contrast, [12] applied segmentation of 45-minute preictal and interictal periods based on EEG seizure annotations, focusing on capturing nonlinear correlations between EDA and HR signals for unsupervised seizure prediction.

1.3.4 Class Imbalance Handling

A significant challenge inherent in seizure detection is the profound class imbalance between rare seizure events and the vast amount of non-seizure data. Seven of the reviewed detection studies explicitly reported strategies to address this issue during model training [1, 24, 10, 3, 7, 25, 4]. No such methods were detailed in the provided papers on seizure forecasting or prediction.

The most frequently cited technique was undersampling, where the number of non-seizure samples is reduced to create a more balanced dataset. Two studies employed random undersampling of non-seizure segments [1, 4]. A similar approach involved subsampling non-seizure epochs to achieve a specific, more manageable seizure-to-non-seizure ratio of 1:10 for training [7].

Conversely, other studies used oversampling or data augmentation techniques. One paper reported using oversampling for seizure events, in combination with algorithmic class weighting, to increase the influence of the minority class [25]. Another study implemented a novel approach by applying a much higher (90%) overlap in its sliding window segmentation for seizure data, which effectively increased the number of seizure-labeled samples for training achieving seizure-to-non-seizure ratio of 1:1.5 [24]. These techniques are crucial for preventing machine learning models from developing a bias towards the majority (non-seizure) class and thereby failing to detect true seizure events.

1.3.5 Features Extraction

The extraction of features from the windowed signal data is a critical step. This process transforms the raw, time-series signals into a structured format suitable for detection algorithms. Feature extraction was explicitly detailed 17 of the detection studies [19, 21, 16, 24, 22, 10, 17, 6, 23, 7, 25, 26, 27, 28, 11, 2, 29] and was central to the methodologies of all three forecasting and prediction papers [14, 13, 12].

The extracted features spanned multiple domains. The most common were statistical and time-domain features, including measures like mean, variance, interquartile range, and the number of local maxima or zero-crossings [19, 21, 26, 29]. Frequency-domain features, derived from techniques like the Fourier transform or wavelet transforms, were also widely used [22, 17, 21, 7]. The number of features extracted varied significantly, from 19 in one study [22] to as many as 594 in another [25]. This process often involved deriving features from novel signals, such as attitude angles (pitch and roll), which were shown to be effective proxies or replacements for traditional accelerometer and gyroscope data [11].

Given the high dimensionality of the feature sets, several studies employed feature selection or dimensionality reduction techniques to identify the most salient predictors and reduce computational complexity. The methods reported include Principal Component Analysis (PCA) [17], statistical tests like ANOVA [26], T-tests [23], and the Wilcoxon rank-sum test [7], as well as machine learning-based approaches like Random Forest feature selection [28] and minimum Redundancy Maximum Relevance (mRMR) [23]. These techniques were shown to optimize the feature set, for instance by reducing 91 initial features to 71 [28], and ultimately improve classification performance [23].

In the context of forecasting and prediction, feature extraction was similarly vital but sometimes focused on different aspects of the physiological data. One study calculated Heart Rate Variability (HRV) metrics, specifically the root mean square of successive differences (RMSSD), to assess impending seizure likelihood [13]. Another forecasting approach extracted features from not only the raw signals and their Fourier transforms but also incorporated non-physiological data such as signal quality indices and the time of day, creating a rich, 17-channel input for its model [14].

1.3.6 Normalization and Baseline Correction

To account for inter-patient physiological variability and to standardize the input for detection algorithms, several studies incorporated normalization or baseline correction steps. This process was detailed in seven of the detection-focused papers [3, 8, 20, 15, 23, 19, 11].

Different methodologies were employed to establish and correct for a patient's physiological baseline. A common technique was the use of a moving baseline, where incoming data from a short window (e.g., 5 seconds) were compared against a constantly updated reference to detect sudden deviations indicative of a seizure [15]. A more personalized approach involved defining subject-specific baselines, for example, by using the median of a patient's signals over a period, and then calculating residuals from this baseline for analysis [8]. Other studies mentioned baseline correction or calibration as a general step for signals like ACC and GYRO [23, 11], or used high-pass filtering as an implicit method to remove baseline drift from ACC signals [19].

In addition to baseline correction, signal normalization or standardization was also performed. Normalization helped make signals more comparable between different people or recording sessions. Some studies used z-score normalization, which adjusts signals to have a mean of zero and a standard deviation of one [3, 20]. These procedures ensure that the scale of different signals and features does not unduly influence the model's performance and help in adapting the algorithm to individual physiological characteristics.

1.4 Algorithms

1.4.1 Deep Learning Methods

Studies have leveraged deep learning architectures to automatically learn hierarchical features and model complex temporal dependencies in multimodal sensor data. Five of the reviewed detection papers [1, 3, 25, 11, 4] and two prediction papers [12, 14] employed deep learning methods. The most prominent architectures included Convolutional Neural Networks (CNNs) [1, 4], Long Short-Term Memory (LSTM) networks [14, 1, 11], and hybrid models combining both [1].

For the seizure detection task, hybrid CNN-LSTM models were shown to be particularly effective. One study identified a CNN-LSTM fusion model using accelerometer (ACC) and blood volume pulse (BVP) data as the best overall algorithm, achieving 83.9% sensitivity and a detection delay of 28 seconds across 28 seizure types [1]. For generalized tonic-clonic (GTC) seizures specifically, this model reached 95% sensitivity [1]. Standalone LSTM networks were also successfully applied, particularly for their ability to capture time-series dynamics. One such study demonstrated that an LSTM model with transfer learning significantly outperformed traditional learning, achieving 93% sensitivity for in-hospital motor seizures with a false alarm rate (FAR) of 2.3 per day [3]. Another study utilized an LSTM with attitude angle signals, among others, to achieve an accuracy of 83.4% [11].

Other neural network architectures were also explored. A CNN-based model was found to be feasible for detecting a broad variety of seizure types using ACC and BVP signals [4]. Simpler Artificial Neural Networks (ANNs) also proved effective, with one study reporting 100% sensitivity in detecting nocturnal tonic seizures in an independent test set [25]. In the context of personalization, an autoencoder model achieved 100% sensitivity for GTC seizures in select patients, although it was noted to be less scalable [1].

For the forecasting and prediction tasks, one study [14] employed supervised long short-term memory (LSTM) neural networks trained using leave-one-subject-out cross-validation on continuously collected retrospective data. Another study [12] selected Deep Canonically Correlated Autoencoders (DCCAE) for training and validation, testing three different architectures: Fully Connected DCCAE (FC-DCCAE), Convolutional Neural Network DCCAE (CNN-DCCAE), and Gated Recurrent Unit DCCAE (GRU-DCCAE). Among these, GRU-DCCAE yielded the best clustering accuracy of 68.89%. Both studies lacked real-time implementation, relying instead on retrospective or offline data analysis.

1.4.2 Ensemble Methods

Ensemble learning, which combines multiple machine learning models to improve predictive performance, was a prominent and effective strategy in several detection studies. Six detection studies reported the use of ensemble classifiers, including bagging, boosting, and more complex stacked architectures [24, 17, 7, 26, 8, 10], and one forecasting study reported the use of random forest classifier [13].

For the seizure prediction task, bagging-based methods, such as Random Forest and Bagged Decision Trees, were particularly common [17, 24, 10, 7]. One study demonstrated that a Random Forest algorithm achieved 90% sensitivity with a low false alarm rate of 1.21 per 24 hours for tonic-clonic seizures in a daily setting [24]. Another study found that a Bagged Decision Tree classifier, when applied to fused accelerometer (ACC) and electrodermal activity (EDA) data, yielded a seizure detection recall of 97.2% and an overall accuracy of 96.7% [17].

Boosting algorithms, including AdaBoost and XGBoost, were also evaluated, with one study comparing their performance alongside other classifiers like Support Vector Machines (SVM) [28]. A more advanced approach was taken in a study that developed a Two-Layer Ensemble Method (TLEM), which stacked multiple base learners including Random Forest (RF), Extra Trees (ET), Gradient Boosting Decision Tree (GBDT), and AdaBoost (ADB) [26]. This stacked model was shown to outperform all of its single-layer components, achieving a particularly high sensitivity of 94.57% and a false alarm rate of

0.46 per 24 hours for nocturnal seizures [26]. This body of work indicates that ensemble methods are a robust choice for aggregating information from multimodal sensors and improving the reliability of seizure detection.

For the seizure forecasting task, [13] evaluated the performance of seven supervised learning algorithms with random forest achieving the higher accuracy of 68% and sensitivity of 62%.

1.4.3 Traditional Machine Learning

Eight detection studies reported the successful application of models such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Linear Discriminant Analysis (LDA) [19, 16, 22, 23, 27, 28, 29, 21].

Support Vector Machines were one of the most frequently and successfully implemented classifiers [19, 21, 22, 23, 27, 28, 29]. One study found that an SVM delivered the best trade-off between accuracy and false alarms, achieving a 100% accurate recognition rate with just 0.08 false alarms per day [28]. Another study concluded that a linear SVM (SVM-L) provided the optimal sensitivity and overall performance among several tested models [29]. The effectiveness of SVMs was also noted when using attitude angle signals, where they yielded the highest overall accuracy compared to decision trees and LDA [11].

Other traditional classifiers also showed strong performance. KNN, particularly when using a cosine distance metric on features from four combined modalities, achieved a high sensitivity of 88.16% [23]. A non-patient-specific classifier achieved 88% sensitivity with a low false alarm rate of one per 24 hours [22]. The collective results indicate that when paired with robust feature engineering, traditional machine learning models remain a viable and powerful option for developing seizure detection systems.

1.4.4 Rule-based and Threshold-based Methods

These methods rely on predefined physiological patterns or thresholds to trigger a seizure detection. Five of the reviewed detection papers used this approach [15, 18, 20, 6, 9].

One study developed a system based on a multi-biosignal pattern, defining a seizure event as a sequence of heart rate increase, followed by a decrease in blood oxygen saturation, and a subsequent rise in electrodermal activity [15]. This pattern-based method successfully detected all seizures from 6 out of 10 patients in the study [15]. Another study implemented a system with decision rules based on three factors: shaking, from an accelerometer, heart rate, and temperature [18]. By classifying risk into discrete levels, the system achieved a sensitivity of 85% [18].

A threshold-based algorithm using cardiac features was also reported, detecting 92% of seizures with tonic/clonic movements [20]. Another approach applied a Shewhart control chart with exponentially weighted moving averages to inertial and muscular activity, achieving a 97% detection rate with a 4% false alarm rate [6]. A further study employed a combined accelerometry and heart rate threshold algorithm in a residential care setting, reaching a median sensitivity of 86% with a positive predictive value of 49% [9].

1.4.5 Methods with Personalization

Recognizing the high degree of inter-patient variability in seizure manifestation, many studies investigated or recommended personalized algorithms. Eight detection papers

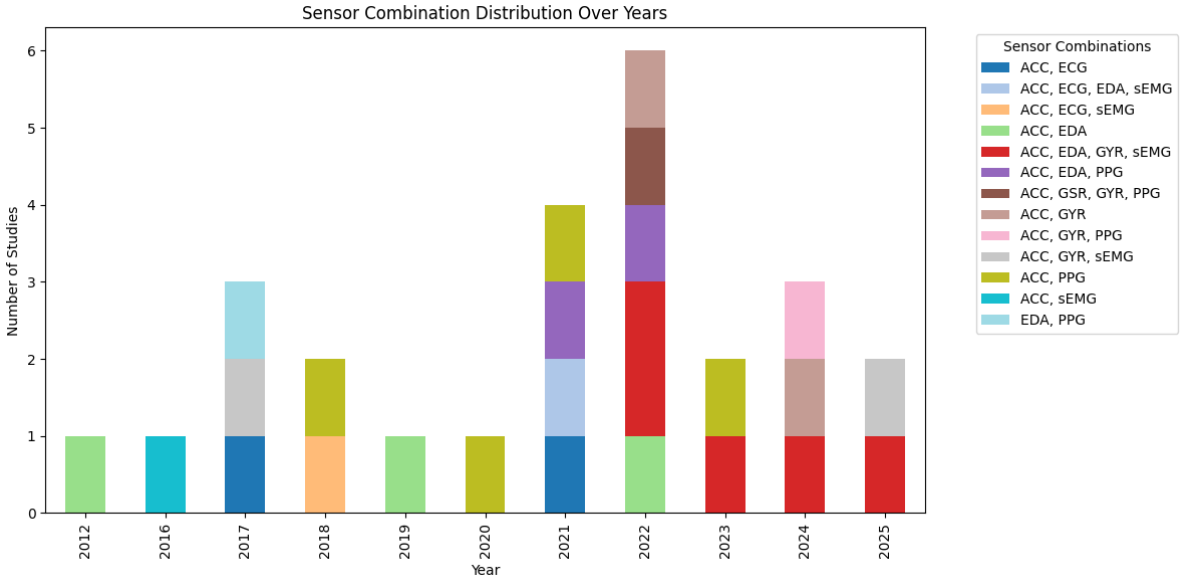

Figure 4: Sensor combination distribution over the years.

highlighted the benefits of tailoring models to individual patients [1, 22, 3, 19, 16, 8, 20, 11].

One study showed that a "semi-patient-specific" approach, which included prior seizure examples from the test patient in the training data, improved sensitivity from 88% to 94% compared to a generic model [22]. Similarly, a personalized autoencoder was able to achieve 100% sensitivity for GTC seizures in certain individuals, a level of performance not reached by the generalized model [1]. A particularly effective technique was transfer learning, where a pre-trained general model was fine-tuned on patient-specific data. This method significantly improved performance, reducing the false alarm rate from 11.3 per day with traditional learning to 2.33 per day [3].

Other studies underscored the need for personalization by observing that false alarms and key predictive features varied significantly between individuals [19, 16]. Some methodologies were inherently personalized, such as a system that tracked individual "physiomes" by establishing personal physiological baselines to detect seizure-related deviations [8]. The consensus across these studies is that while generalized models can provide a strong out-of-the-box solution, personalization is a critical step for maximizing detection accuracy and minimizing false alarms in real-world applications.

2 Discussion

2.1 Modalities

2.1.1 Detection

After reviewing all the included studies, it is apparent that non-EEG multimodal sensor configurations for detecting motor seizures are not only feasible but also show great promise compared to conventional detection and management methods. Several studies have assessed the performance of different combinations of sensor modalities and biomarkers, highlighting the importance of identifying the optimal configuration. However, to

determine the best-performing combination of modalities and biomarkers, it is important to note that direct comparison across studies is challenging due to methodological heterogeneity, differences in study populations, and variability in training datasets.

Nevertheless, there has been a recent trend toward using the ACC + GYR + EDA + sEMG sensor combination (Figure 4). This combination performance is consistently high across studies (table no) despite different study setups. ACC + EDA is also one of the top performing combinations. This may be because these setups combine the reliability of motion-based detection through ACC, GYR, and sEMG with the stabilizing contribution of physiological signals such as EDA, which helps reduce false alarms and enhance overall performance. While this highlights the significance of EDA data in seizure detection, some studies have reported poor performance of EDA as a standalone modality and even when combined with other sensors such as PPG or ACC [1, 4]. This variability across studies stems perhaps from the age-related variability in EDA signals reported by the pediatric study [23], highlighting the need for algorithms trained on EDA to count for such variability.

Other physiological sensors such as PPG have also been widely used to extract biomarkers including HR [15, 3, 7, 28, 9, 8], HRV [7, 8], SpO_2 [15], and BVP [3, 4, 1]. Among these, BVP in particular has shown strong potential, consistently achieving high performance across seizure types.

ECG has also been employed to derive HR [5, 20, 21] and HRV [20], with some studies reporting that unimodal ECG systems can achieve performance comparable to multimodal approaches [21, 20]. However, ECG was used less frequently in the reviewed studies, largely due to its susceptibility to motion artifacts [5] and the high inter-patient variability observed in cardiac-derived signals [21, 5].

Because of these limitations, ECG has mostly been used in nocturnal studies [5, 21]. This is relevant since SUDEP is more likely to happen at night [30], and HR and HRV are under investigation as potential SUDEP biomarkers [31]. Thus, ECG may serve a dual role in seizure detection and monitoring SUDEP-related autonomic changes.

Some studies have also investigated the use of non-traditional biomarkers for seizure detection, and interestingly, several of these have demonstrated detection capabilities comparable to, or even exceeding, those of more established biomarkers. [16] investigated the incorporation of audio features , derived from a MIC, in the detection system and reported that audio signals proved to be among the top ten features establishing separability between seizure and non-seizure data in four or five patients. [11] investigated the incorporation of attitude angle signals like PITCH or ROLL, which describe the orientation of the body/limb in the 3D space, in their detection system and reported that in multimodal combinations, adding PITCH or ROLL alongside or replacing ACC, GYR outperformed combinations that excluded them across all models. Instead of using raw ACC data, [28] used Number of Wrist Movements (NOWM), a derived feature that summarizes movement frequency over time windows, to distinguish seizure-like activity from normal daily movements.

Thus, future studies should validate the incorporation of these biomarkers into detection systems, while also assessing optimal sensor placement. Evidence from [19, 21] shows that different sensor placements produce different outcomes, highlighting the importance of placement as a design consideration. In addition, studies should account for the variability exhibited by physiological signals across different populations by increasing and diversifying their cohorts, and by developing personalized models for highly patient-dependent modalities such as HR and HRV.

2.1.2 Prediction and Forecasting

Although the tasks and methodological scopes of the three studies varied, together they provide strong evidence for the feasibility of using wearable data from the Empatica E4 wristband in combination with machine learning or deep learning algorithms for seizure prediction and forecasting. Vieluf et al. [12, 13] demonstrated that the purely autonomic sensor set of EDA and PPG (from which HR and HRV are derived) could successfully discriminate pre-ictal periods in a substantial proportion of patients in their cohorts. However, the highest prediction performance was reported by Meisel et al. [14], where all available sensor modalities on the E4 device (EDA, PPG/BVP, ACC, and skin temperature) were incorporated into the forecasting models.

2.2 Preprocessing

The preprocessing pipeline across the reviewed studies reveals a consistent, multi-stage approach to preparing wearable sensor data for analysis. The initial steps of synchronization and quality control are fundamental for ensuring data integrity, yet there is a notable lack of a standardized protocol. While methods ranging from manual clock alignment [1] to automated NTP [7] were reported, the reliance on manual or semi-automated processes can be labor-intensive and represents a barrier to the seamless, large-scale deployment of these systems. Future work should focus on developing fully automated and robust algorithms for real-time signal quality assessment and data synchronization.

Following quality control, the application of filtering techniques was nearly universal, with methods appropriately tailored to specific signal modalities, such as band-pass filtering for motion sensors [10, 21] and high-pass filtering for sEMG [19]. The primary challenge in this stage is the trade-off between noise reduction and the preservation of subtle, seizure-relevant physiological signatures. Similarly, the segmentation of data into windows is a critical design choice that directly influences model performance. The literature showed a clear distinction, with shorter windows (2-10s) used for capturing the rapid dynamics of motor seizures [19, 25] and longer windows (>30s) used for forecasting or analyzing slower autonomic changes [14, 8]. The optimal window size and overlap remain an open question and may require patient-specific or context-aware adaptation.

The pervasive issue of class imbalance was addressed with various strategies, primarily under- and over-sampling [1, 4, 25]. While effective, simple random undersampling risks discarding valuable information from non-seizure periods, which could be crucial for building robust models that can distinguish seizures from vigorous daily activities. More advanced methods, such as synthetic data generation or cost-sensitive learning algorithms, were not widely reported and represent a significant area for future research. Finally, the use of feature engineering and subsequent selection [23, 28] versus the end-to-end learning approach of deep learning models marks a key divergence in methodology. While hand-crafted features offer interpretability, the trend towards deep learning suggests a move to reduce reliance on domain-specific feature design. Normalization and baseline correction [8, 3] were identified as crucial steps for personalization, enabling models to adapt to the significant physiological variability among individuals.

2.3 Algorithms

The reviewed literature demonstrates wide methodological diversity in algorithms for GTCS detection, prediction, and forecasting, ranging from interpretable rule-based sys-

tems to advanced deep learning architectures. For detection, Artificial Neural Networks (ANNs) have shown particular promise, with Larsen et al. [25] achieving 100% sensitivity and a false alarm rate of 0.16 per night in nocturnal seizure detection. While CNN and hybrid CNN-LSTM models have been explored, their higher false positive rates remain a barrier to clinical translation [1, 4]. Ensemble methods such as the Multi-Level Dynamic Time Pile (MLDTP) framework [10] and traditional classifiers like Support Vector Machines (SVMs) [22] also demonstrate strong performance, particularly when validated against real-world non-ictal data. Simpler threshold- and rule-based algorithms provide interpretable, real-time solutions, exemplified by the cardiac-based detector of Hegarty-Craver et al. [20], though their limited adaptability reduces applicability across diverse patients.

In forecasting, supervised recurrent neural networks, especially Long Short-Term Memory (LSTM) models, are the dominant approach. Meisel et al. [14] reported significant forecasting in 43.5% of patients without requiring patient-specific data or dependence on seizure type, suggesting broader generalizability than previously assumed. In contrast, Nasseri et al. [32] achieved forecasting in 83% of focal epilepsy patients, highlighting both the potential of LSTMs and the need to validate generalization across seizure types. Beyond supervised learning, unsupervised methods such as Deep Canonically Correlated Autoencoders (DCCAE) [12] demonstrate the ability to extract predictive features from multimodal physiological signals without labeled seizure events, offering a path forward for wearable-based forecasting where annotation is limited.

A consistent challenge across detection and forecasting is inter-patient variability. While some architectures appear capable of patient-independent generalization, many models fail to provide reliable performance across all individuals. Personalization strategies, such as transfer learning [3] or integration of clinical information [13], have shown promise in improving model robustness, most notably through reductions in false alarm rates and better adaptation to individual physiological profiles.

Overall, while advanced deep learning and ensemble methods have achieved encouraging results, no single algorithm yet combines the sensitivity, low false alarm rate, generalizability, and real-time capability required for clinical adoption. Future progress will depend on balancing generalizable architectures with personalized adaptation, and on validating algorithms in real-world ambulatory settings to ensure reliability and safety.

2.4 Limitations - Prediction and forecasting Only

Despite the noted progress in the field of seizure prediction or forecasting, several limitations remain. The reviewed studies primarily involved pediatric cohorts. Given physiological and behavioral differences between pediatric and adult populations, these findings may not directly translate to adults without further validation. Larger, more diverse datasets are necessary to validate and extend these results. Additionally, all studies were conducted in inpatient settings, where patients' daily activities and stress levels differ significantly from outpatient environments. While inpatient monitoring allows for gold-standard seizure characterization via video EEG, the transferability of these models to outpatient settings is limited. It is worth noting that although outpatient data collection still presents challenges such as low data quality and lack of continuous clinical supervision, it is critical for practical seizure forecasting applications.

Another key limitation is that none of the studies performed real-time data analysis. Real-time seizure forecasting is crucial for timely intervention and practical clinical use.

Future research should therefore focus on developing algorithms that can process data in real time.

Finally, the optimal timing and duration for data recording in seizure prediction remain unclear. Although a time interval of 9:00 to 9:15 pm was hypothesized to be ideal for predicting seizures that could occur during nighttime or early morning hours [13], which are especially important, particularly for pediatric patients, as nighttime supervision can help reduce the risk of sudden unexpected death in epilepsy (SUDEP) [33], it is necessary to identify the shortest data windows that still enable reliable prediction across different patient groups to improve model accuracy and reliability.

In summary, while the current evidence supports the potential of wearable multimodal sensor data combined with machine learning or deep learning for seizure forecasting, addressing these limitations such as expanding cohorts, validating in outpatient settings, implementing real-time analysis, and optimizing data acquisition will be essential for moving toward clinical application.

References

[1] S. Yu et al. "Artificial intelligence-enhanced epileptic seizure detection by wearables". en. In: *Epilepsia* 64.12 (Dec. 2023), pp. 3213–3226.

[2] G. Regalia et al. "Multimodal wrist-worn devices for seizure detection and advancing research: Focus on the Empatica wristbands". en. In: *Epilepsy Res.* 153 (July 2019), pp. 79–82.

[3] M. Nasseri et al. "Non-invasive wearable seizure detection using long-short-term memory networks with transfer learning". en. In: *J. Neural Eng.* 18.5 (Apr. 2021), p. 056017.

[4] J. Tang et al. "Seizure detection using wearable sensors and machine learning: Setting a benchmark". en. In: *Epilepsia* 62.8 (Aug. 2021), pp. 1807–1819.

[5] J. van Andel et al. "Multimodal, automated detection of nocturnal motor seizures at home: Is a reliable seizure detector feasible?" en. In: *Epilepsia Open* 2.4 (Dec. 2017), pp. 424–431.

[6] M. Gheryani, O. Salem, and A. Mehaoua. "Detection of nocturnal epileptic seizures from wireless inertial measurements and muscular activity". In: *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. Singapore: IEEE, Dec. 2017.

[7] Y. S. Vakilna et al. "Reliable detection of generalized convulsive seizures using an off-the-shelf digital watch: A multisite phase 2 study". en. In: *Epilepsia* 65.7 (July 2024), pp. 2054–2068.

[8] P. Jiang et al. "Longitudinally tracking personal physiomes for precision management of childhood epilepsy". en. In: *PLOS Digit. Health* 1.12 (Dec. 2022), e0000161.

[9] J. Arends et al. "Multimodal nocturnal seizure detection in a residential care setting: A long-term prospective trial". en. In: *Neurology* 91.21 (Nov. 2018), e2010–e2019.

[10] D. Wu et al. "A novel seizure detection method based on the feature fusion of multimodal physiological signals". In: *IEEE Internet Things J.* 11.16 (Aug. 2024), pp. 27545–27556.

[11] J. Wang et al. “Epileptic seizure detection based on attitude angle signal of wearable device”. In: *IEEE Trans. Instrum. Meas.* 74 (2025), pp. 1–10.

[12] S. Vieluf et al. “Developing a deep canonical correlation-based technique for seizure prediction”. en. In: *Expert Syst. Appl.* 234.120986 (Dec. 2023), p. 120986.

[13] S. Vieluf et al. “Development of a multivariable seizure likelihood assessment based on clinical information and short autonomic activity recordings for children with epilepsy”. en. In: *Pediatr. Neurol.* 148 (Nov. 2023), pp. 118–127.

[14] C. Meisel et al. “Machine learning from wristband sensor data for wearable, non-invasive seizure forecasting”. en. In: *Epilepsia* 61.12 (Dec. 2020), pp. 2653–2666.

[15] D. Cogan et al. “Multi-biosignal analysis for epileptic seizure monitoring”. en. In: *Int. J. Neural Syst.* 27.1 (Feb. 2017), p. 1650031.

[16] A. Hamlin et al. “Assessing the feasibility of detecting epileptic seizures using non-cerebral sensor data”. en. In: *Comput. Biol. Med.* 130.104232 (Mar. 2021), p. 104232.

[17] M. E. H. Chowdhury et al. “Wearable real-time epileptic seizure detection and warning system”. In: *Biomedical Signals Based Computer-Aided Diagnosis for Neurological Disorders*. Cham: Springer International Publishing, 2022, pp. 233–265.

[18] S. N. Ali and M. J. Alam. “Development of a wearable 3-risk factor accumulated epileptic seizure detection system with IoT based warning alarm”. In: *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*. Greater Noida, India: IEEE, Oct. 2020.

[19] M. Milosevic et al. “Automated detection of tonic–clonic seizures using 3-D accelerometry and surface electromyography in pediatric patients”. In: *IEEE J. Biomed. Health Inform.* 20.5 (Sept. 2016), pp. 1333–1341.

[20] M. Hegarty-Craver et al. “Cardiac-based detection of seizures in children with epilepsy”. en. In: *Epilepsy Behav.* 122.108129 (Sept. 2021), p. 108129.

[21] T. De Cooman et al. “Comparison and combination of electrocardiogram, electromyogram and accelerometry for tonic-clonic seizure detection in children”. In: *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. Las Vegas, NV, USA: IEEE, Mar. 2018.

[22] M.-Z. Poh et al. “Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor”. en. In: *Epilepsia* 53.5 (May 2012), e93–7.

[23] Y. Ge et al. “Epilepsy analysis with portable device based multi-modal physiological signals”. In: *2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. Hefei, China: IEEE, Aug. 2023.

[24] Q. Wang et al. “Daily epileptic seizure detection algorithm based on multi-modal physiological signals”. In: *2022 5th International Conference on Communication Engineering and Technology (ICCET)*. Shanghai, China: IEEE, Feb. 2022.

[25] S. A. Larsen, D. H. Johansen, and S. Beniczky. “Automated detection of tonic seizures using wearable movement sensor and artificial neural network”. en. In: *Epilepsia* 65.9 (Sept. 2024), e170–e174.

[26] C. Dong et al. “A two-layer ensemble method for detecting epileptic seizures using a self-annotation bracelet with motor sensors”. In: *IEEE Trans. Instrum. Meas.* 71 (2022), pp. 1–13.

[27] W. Li et al. “Generalized tonic-clonic seizure detection through the use of physiological signals”. In: *2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML)*. Chengdu, China: IEEE, July 2022.

[28] G. Xu et al. “Total tonic clonic seizure recognition of wrist signals”. In: *2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*. Guangzhou, China: IEEE, July 2022.

[29] G. Wang et al. “Seizure detection using the wristband accelerometer, gyroscope, and surface electromyogram signals based on in-hospital and out-of-hospital dataset”. en. In: *Seizure* 127 (Apr. 2025), pp. 127–134.

[30] D. Friedman. “Sudden unexpected death in epilepsy”. en. In: *Curr. Opin. Neurol.* 35.2 (Apr. 2022), pp. 181–188.

[31] N. Barot and M. Nei. “Autonomic aspects of sudden unexpected death in epilepsy (SUDEP)”. en. In: *Clin. Auton. Res.* 29.2 (Apr. 2019), pp. 151–160.

[32] M. Nasseri et al. “Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning”. en. In: *Sci. Rep.* 11.1 (Nov. 2021), p. 21935.

[33] M. Trivisano et al. “Risk of SUDEP during infancy”. en. In: *Epilepsy Behav.* 131.Pt B (June 2022), p. 107896.