Disease Prediction Using Machine Learning Approaches Considering Bio-Medical Signal Analysis: A Survey

K. Satyanarayana Murthy[1,2,*], Suribabu Korada[3]

[1]Computer Science and System Engineering, AU-TDR-HUB, Andhra University, Visakhapatnam, India
[2]Department of Information Technology, ANITS (A), Visakhapatnam, India
[3]Scientist-E, NSTL(DRDO), Visakhapatnam, India
Emails: murthy8542.mtech@gmail.com; suribabukorada2000@gmail.com

Abstract

In medical diagnosis and prognosis, symptoms provided by patients play a critical role in identifying diseases. Machine learning offers a powerful approach to analyzing and predicting illnesses based on these symptoms. In particular, classification algorithms are widely used to analyze input data and predict disease outcomes. A key factor in effective classification is the selection of relevant attributes, which directly affects the accuracy of the prediction. This research emphasizes the importance of proper feature extraction techniques in the context of disease prediction using biomedical signal analysis. Effective analysis requires both the extraction of critical features and the elimination of irrelevant data. The aim of this study is to explore existing approaches to disease prediction based on biomedical signal analysis. We focus on feature extraction from pre-processed data, which aids in distinguishing between different biomedical signals recorded by medical devices. Our objective is to identify biomedical cues that differentiate various health conditions. Examples of such signals include electroencephalogram (EEG), electrocardiogram (ECG), and electrogastrogram (EGG). Understanding how these signals differ between healthy and diseased states is crucial for accurate disease prediction. This research investigates diseases such as heart disease, kidney failure, and lung infections, considering how variations in biomedical signals can be used to predict the likelihood of severe illness. We continue to seek advancements in predicting and mitigating future health risks.

Keywords: Machine Learning; Signal Processing; Electroencephalogram (EEG)

1. Introduction

Medical diagnosis and prognosis have traditionally relied on clinical observations and patient-reported symptoms to detect and assess diseases. However, with the growing availability of biomedical data and advancements in computational techniques, machine learning has become a critical tool for automating disease prediction and improving diagnostic accuracy. Machine learning algorithms can analyze vast amounts of data, identifying patterns and correlations that may not be immediately evident through conventional methods, thereby supporting earlier detection and more accurate diagnosis of diseases. Classification algorithms, a subset of machine learning, are particularly useful in the context of disease prediction. These algorithms categorize input data into distinct groups, such as healthy and diseased states, based on specific attributes or features. In biomedical signal analysis, the selection of relevant features from complex and high-dimensional data is a critical step in ensuring the accuracy of predictive models. Proper feature extraction is essential for isolating informative characteristics from raw signals while filtering out irrelevant or noisy data. Biomedical signals, such as electroencephalograms (EEG), electrocardiograms (ECG), and electrogastrograms (EGG), provide comprehensive physiological information that can reflect the presence and progression of diseases. These signals vary according to the state of health or disease in an individual, and their analysis can reveal critical insights into conditions affecting the brain, heart, and other organs. For instance, abnormalities in ECG signals may indicate cardiovascular issues, while changes in EEG patterns can suggest neurological disorders. Effective pre-processing and feature extraction from these signals are vital for distinguishing between normal and pathological states, thus aiding in disease detection. Additionally, we review existing machine learning approaches used in pre-processing, feature extraction, and classification of

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

biomedical signals to evaluate their effectiveness in predicting diseases such as heart disease, kidney failure, and lung infections. The ultimate goal of this study is to advance the understanding of biomedical signal analysis for disease prediction and contribute to the development of models that can improve early diagnosis and patient outcomes. The paper is organized as follows: Section 2 presents the research objective. Section 3 presents the results of the systematic review. Consequently, section 4 presents the conclusions and future scope.

2. Research Objectives

The research on disease prediction started a long time ago, and the research is continuing to find the best technique for disease prediction when using biomedical signal data such as EEG, ECG and EGG. Many of the previous research papers focused on a single objective, like brain abnormalities, cerebral injuries or seizure. So this article aims to answer different questions, taking our research objectives like data sources, search strategy, addition & omission criteria, quality evaluation, data extraction, and data synthesis are all outlined in this review approach.

A. Research Questions

RQ1: Which machine learning methods are currently in use for categorizing biological signals based on published research?
RQ2: Which dataset, based on bio medical signals like electroencephalograms (EEG), is most often used to predict disease?
RQ3: Which is the effective classification technique in predicting the disease when biomedical signals data are used?
RQ4: Which publications are statically dominant in the area of biomedical signal data classification?
RQ5: Evaluation of various machine-learning algorithms' performance using validation techniques?

B. Search Procedure adopted for selection of data repository

Pubmed, National Centre for Biotechnology Information (NCBI), HealthData.gov, SEER (Surveillance, Epidemiology, and End Results), ClinicalTrails.gov, EEG BCI Competition, PhysioNet, OpenBCI, EEG-SSS, Sleep-EDFx, MNE-Python Datasets, SLEEPDATA, EMBASE, Cochrane Library, OpenTrails, BioLINCC, Online Library, and other electronic libraries are among the data sources mentioned. The search strategy includes looking for keywords related to our research. The search focused on terms from research questions as well as commonly used terms related to machine learning and biomedical signal analysis. The total papers searched were 80. After performing the addition and omission criteria on 80 papers, we were left with 36 papers, which are addressed below.

B. selection of papers after addition and omission

- Criteria for addition and omission have been developed to accurately assess the quality of the available literature. The papers were examined by the authors and discussed with an industry as well as academic expert for addition and omission decisions.
- The papers were examined in the following criteria:

Addition Criteria: The articles should be in full text format, published between the years 2015to 2024 and based on prognosis and diagnosis of diseases.
Omission Criteria: The papers not satisfying the above addition criteria. Keywords based search was also not considered

- The outcome of this stage is 36 papers

C. Quality Assessment

Experts affiliated with reputable journals and conferences chose the papers from reputable databases after being peer-reviewed. As of yet, no specified quality standards were required. The peer-reviewed studies were deemed adequate.

E. Data Extraction

Five important data characteristics are taken into consideration in the data extraction procedure that the authors developed. The five characteristics considered during the extraction process are as follows:

a. Machine Learning Methods/Approaches Discussed

b. Datasets Used

c. Year of publication

d. Performance Comparison

e. Validation Methods

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

F. Data Synthesis

The data synthesis process includes putting together the data and drawing conclusions based on the study questions. Data synthesis was carried out by analyzing the literatures using various statistical methods.

3. Results of the Systematic Review

The outcome of the review is addressed in the form of answers to the research questions. In this section the authors have answered all the research questions. These research questions will help the readers understand the quality of work done in the field of biomedical signal analysis so far, the authors of this paper tried to provide statistical analysis also wherever possible.

3.1: Categorization of the disease prediction based on the development type

Organizations have their own way to develop the project. The organizations may use different development types, environment. The procedure also varies with the development type of the project. So the prediction of disease based on the user symptoms depends on their development type is categorized into different categories. They are as follows:

RQ1: Which machine learning methods are currently in use for categorizing biological signals based on published research?

Machine learning tasks involving EEG data primarily focus on classification, where the choice of robust classification algorithms is crucial for system performance. Traditional classification algorithms commonly used in brain-computer interface (BCI) systems include Support Vector Machines (SVM), Decision Tree (DT), Random Forest, Naïve Bayesian, k-Nearest Neighbor (KNN) and Gradient Boosting. These conventional approaches have been effective but can be limited in handling high-dimensional, complex EEG data. However, a key challenge in applying machine learning to EEG data is determining the appropriate amount of data required for accurate model training and validation. The question of "how much data is sufficient?" remains a topic of ongoing debate. Most studies rely on publicly available EEG datasets, which form the foundation for training these algorithms.

Machine learning algorithms perform better and are more accurate when features in a dataset are transformed to a similar scale using the data pre-processing approach known as normalization, windowing, time domain analysis and frequency domain analysis. Feature extraction in machine learning involves transforming a set of raw, often noisy, features into a refined set of meaningful features that can be effectively utilized for analysis, prediction, or classification tasks. This process is crucial for improving the performance and accuracy of machine learning models. Various methods are employed in feature extraction, such as principal component analysis (PCA), Time-Frequency representation, Machine Learning based features, Mel Frequency Cepstral Coefficient (MFCC) and the generation of polynomial features to capture both linear and non-linear relationships.

Feature extraction plays a pivotal role in reducing the dimensionality of these complex datasets, transforming them into manageable and relevant feature sets. Additionally, biomedical signals are often affected by noise or interference from both internal (physiological) and external (environmental or device-related) sources, which can introduce inconsistencies. By applying effective feature extraction techniques, the noise can be minimized, and essential signal properties can be highlighted, leading to more accurate classification and prediction outcomes. This process ultimately ensures that machine-learning models can make reliable predictions based on the core features of the biomedical data.

RQ2: Which dataset, based on bio medical signals like electroencephalograms (EEG), is most often used to predict disease?

Pubmed, National Centre for Biotechnology Information (NCBI), HealthData.gov, SEER (Surveillance, Epidemiology, and End Results), ClinicalTrails.gov, EEG BCI Competition, PhysioNet, OpenBCI, EEG-SSS, Sleep-EDFx, MNE-Python Datasets, SLEEPDATA, EMBASE, Cochrane Library, OpenTrails, BioLINCC, Online Library, and other electronic libraries are available. The PhysioNet EEG Motor Movement/Imagery dataset is a widely used dataset for predicting diseases based on electroencephalograms (EEG). It includes EEG recordings from subjects performing motor tasks, used in brain-computer interface research. The BCI Competition datasets, particularly those from BCI Competition IV, include EEG recordings for tasks like movement intention and mental state classification. Kaggle EEG and OpenBCI datasets are popular for machine learning and signal processing research.

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

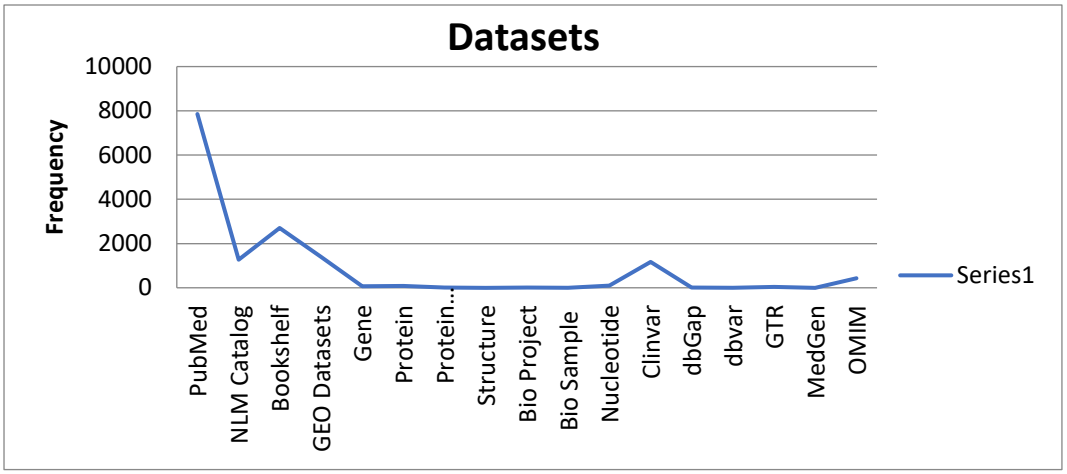

Figure 1. EEG Datasets

RQ3: Which is the effective classification technique in predicting the disease when biomedical signals data are used?

Support Vector Machine

SVM is a supervised machine-learning algorithm used for classification and regression tasks. It aims to find the optimal hyperplane to separate data points of different classes in a high-dimensional space. SVM uses labeled training data and solves an optimization problem to identify the hyperplane that maximizes the margin between classes.

- Decision Function: The decision function is based on the support vectors and is used to classify new data points. The equation of the hyperplane can be represented as:

$f(x)=w \cdot x+b$

Where w is the weight vector (normal to the hyperplane), x is the input feature vector, and b is the bias term.

- Classification Rule: A new data point is classified based on the sign of the decision function:
  - If $f(x)>0$: Class 1
  - If $f(x)<0$: Class 2

It can handle non-linear separations through the kernel trick, which transforms the input space into a higher-dimensional space. SVM includes linear, polynomial, and radial basis function kernals. A regularization parameter (C) balances the trade-off between maximizing margin and minimizing classification errors.

Random Forest:

Random Forest is an ensemble learning method used for classification and regression tasks. It builds multiple decision trees during training and merges their outputs to improve accuracy and control overfitting. Random Forest uses bagging, a technique that randomly samples the training dataset to create multiple subsets for training each decision tree. This introduces diversity among trees, making them less correlated and improving performance. For classification tasks, each tree votes for a class, and the class with the most votes is chosen as the final prediction.

The core of Random Forest is the decision tree. The splitting of nodes in a decision tree is often based on criteria such as:

- Gini Impurity (for classification):

$\text{Gini}(D)= \sum_{i=1}^{c} p\binom{2}{i}$

where D is the dataset, C is the number of classes, and p_i is the proportion of instances belonging to class i in the dataset.

- Entropy (for classification):

$\text{Entropy}(D)= -\sum_{i=1}^{c} pi \log_2(pi)$

This measures the impurity of a dataset D and is used to decide how to split nodes in the tree.

- Mean Squared Error (MSE) (for regression):

$\text{MSE}=1/n \sum_{j=1}^{n} (y_j - \hat{y}_j)^2$

wherey_j is the actual value, $\hat{y}_j$ is the predicted value, and n is the number of observations.

Random Forest uses bagging (Bootstrap Aggregating), which involves creation of m datasets by randomly sampling with replacement from the original dataset. The size of each sample is usually the same as the original dataset.When splitting nodes, Random Forest selects a random subset of features F from the total features M. The number of features to consider is often specified by:

$k= \sqrt{M}$ (for classification) or $k=M/3$ (for regression)

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

For classification, the final prediction P for a new instance x is made based on the majority vote from all n trees:
P(x)=mode(T1(x),T2(x),…,Tn(x))
where Ti(x) is the prediction from the i^{th} tree.
For regression, the final prediction is the average of all tree predictions:
$P(x)=1/n \sum_{i=1}^{n} Ti(x)$

Gradient Boosting:

A potent machine learning method, gradient boosting is mostly used to classification and regression problems. In order to produce a powerful predictive model, it combines weak learners—usually decision trees—in a systematic manner. It begins with a basic model, which is often the target variable's mean. For a certain number of repeats, determine the residuals, or the difference between the predicted and actual values. Fit a new model to these residuals in order to precisely describe the inaccuracy of the existing ensemble. To update the projections, add the new model's predictions, scaled by the learning rate. The final model is the sum of all the models produced throughout the each iteration.

Let $\{(x_i, y_i)^n\}_{i=1}$ be a dataset with n samples, where y_i is the target variable and x_i is the feature vector. Following m iterations, the ensemble model's prediction may be expressed as follows:

$$F(x) = \sum_{j=1}^{m} fj(x)$$

Where fj(x) is the ^{jth} weak learner's (decision tree, for example) prediction. Define a loss function L(y, F(x)) that quantifies the discrepancy between the expected value F(x) and the actual goal y. Increasing Procedures

Step 1: Make your first forecast, which is usually the mean: For i=1 to n, $F_0(x)$= arg min$\sum L(y_i,\gamma)$
Step 2: For every m iteration between 1 and M: Determine the residuals: $r_i^{(m)}=-\partial L(y_i,F_{m-1}(x_i))/ \partial F|F=F_{m-1}(x_i)$. Fit the residuals $r_i(m)$ to a new weak learner $f_m(x)$.Revise the model: $F_m(x)=F_{m-1}(x)+\eta f_m(x)$, where η represents the rate of learning.
Step 3: For m=1 to M, the final model is $F(x)=F_0(x)+\sum \eta f_m(x)$ after M iterations. Regularization terms, including restricting the depth of trees or incorporating penalties in the loss function, may be included to avoid overfitting. This may be expressed as follows: $F_m(x)=F_{m-1}(x)+\eta(f_m(x)-\lambda R(f_m))$, where λ regulates the intensity of regularization and $R(f_m)$ is a regularization term.

3.2 Classification and data analysis Approaches

The various classification and data analysis methods used by the authors in prediction of disease under EEG are shown in below table. Mainly it includes Artificial Neural Networks, Support Vector Machines, Random Forest, Statistical Analysis, Correlation Analysis, Regression Trees, Gradient Boosting, Retrospective and observational approaches, according to the statistical results. Each Approach has its own set of conditions and features that must be met before it can be employed.

Table 1: Summary of Machine learning algorithms for disease prediction

S.No	Authors name	Methodology	Data description	Performance/Significance of work
1	Hitesh Yadav, Surita Maini (2023)	SVM, LDA	Electroencephalogram signal data	Comparison of various EEG MI-BCI applications using machine-learning techniques.
2	Cataldo, Aet al. (2023)	Multi-Scale Entropy and correlation analyis	COVID-19 and Neurodegenerative disease	To investigate the possible common EEG entropic features between COVID-19 and neurodegenerative diseases.
3	Korkmaz, O.Eet al. (2021)	Artificial Neural Networks (ANN) model	Artificial Neural Networks (ANN) model was used to detect the P300 signal.	Namely, all subjects average of Oz channel CA rates before and after COVID-19 for 15 repetitions are 95.58% with a standard deviation of 2.54.
4	Tantillo GBet al. (2022)	Demographics, VEEG indications	Demographics, comorbidities, neuroimaging, VEEG	Most COVID-19 patients who underwent VEEG monitoring had severe COVID-19 and over one-

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

			indications and findings, treatment, and outcomes were collected.	third had acute cerebral injury (e.g., stroke, anoxia)
5	Yao Yet al. (2023)	entropy, energy spectrum	electroencephalography signal data	Results preliminarily demonstrate that COVID-19 patients exhibit certain brain abnormalities during rest,
6	Luíza Alves CORAZZAet al. (2021)	Retrospective, observational, and non-interventional study	Data were collected during the period from Mar 1 to Jun 30, 2020, either confirmed (positive RT-PCR) or probable cases (CO-RADS 4/5) who had performed EEG during hospitalization.	75% of 21 patients had severe ARDS-related conditions requiring mechanical ventilation, while 56.1% developed adjunct sepsis during hospitalization, but no specific COVID-19 pattern was identified in EEG.
7	Sáez-Landete Iet al. (2022)	Retrospective, observational, and Classification methods	They made a retrospective analysis of 29 EEGs recorded in 15 patients with COVID-19 and neurological symptoms.	They have not found a statistically significant association between voltage of acute EEG and non neurological clinical status
8	Yang, Yet al. (2024)	Data analytics and the correlation analysis	125 pediatric patients infected with SARS-CoV2 and showing neurological symptoms, and their continuous EEG was recorded	The proportion of patients diagnosed with febrile seizure was higher in the normal EEG group than in the abnormal EEG group ($P = 0.002$),
9	Gogia Bet al. (2021)	observational, and Classification methods	COVID-19 positive patients (1468) records	The study found that patients with encephalopathy and focal lesions are at a higher risk of developing new-onset status epilepticus.
10	Karadas, Oet al. (2022)	Statistical and correlation analysis	Recordings of 87 intensive care patients who were diagnosed with COVID-19.	Abnormal EEG findings were detected in 93.1% ($n = 81$) of the patients, which were found to increase significantly with age ($p < 0.001$).
11	A.R. Antony et al. (2020)	Statistical and correlation analysis	COVID-19 infected 617 patients with EEG findings reported in 84 studies.	In studies that utilized continuous EEG, 96.8 % (n = 243) of the 251 patients were reported to have abnormalities compared to 85.0 % (n = 311) patients who did not undergo continuous EEG monitoring ($\chi^2 = 22.8$, $p =< 0.001$).
12	Marinelli, Let al. (2022)	Retrospective and Correlation analysis	The EEGs recorded during the first year of the COVID-19 pandemic.	Patients with attenuated EEG and non-survivors showed lower PaO_2/FiO_2 values. Neuroradiological findings were very heterogeneous with a prevalence of lesions suggestive of a microangiopathic substrate.
13	Carolyn Tsaiet al. (2022)	Data analytics	COVID-19 Data set	facilitate earlier detection and treatment of seizures
14	W. Guanet al. (2019)	Statistical analysis	Authors extracted data regarding 1099 patients with	In 157 of 877 patients (17.9%) with non severe disease and in 5 of 173 patients (2.9%) with severe

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

			laboratory-confirmed Covid-19 from 552 hospitals	disease. Lymphocytopenia was present in 83.2% of the patients on admission.
15	L. Maoet al. (2020)	Retrospective and observational analysis	Data were collected from January 16, 2020, to February 19, 2020, at 3 designated special care centres for COVID-19	Of 214 patients (mean [SD] age, 52.7 [15.5] years; 87 men [40.7%]) with COVID-19, 126 patients (58.9%) had non severe infection and 88 patients (41.1%) had severe infection according to their respiratory status.
16	A. Varatharajet al. (2020)	Statistical analysis	COVID-19 data from UK Government public health bodies.	77 (62%) of 125 patients presented with a cerebrovascular event, of whom 57 (74%) had an ischaemic stroke, nine (12%) an intracerebral haemorrhage, and one (1%) CNS vasculitis.
17	M.U. Ahmedet al. (2020)	Retrospective and observational analysis	COVID-19 using PUBMED and subsequent proceedings. A total of 118 articles were thoroughly reviewed	The neurological manifestations associated with COVID-19 such as Encephalitis, Meningitis, acute cerebrovascular disease, and GuillainBarré Syndrome (GBS) are of great concern.
18	M.A. Ellulet al. (2020)	Data analytics	214 hospitalized patients with COVID-19 in Wuhan, China, and 40 (69%) of 58 patients in intensive care with COVID-19 in France.	Careful clinical, diagnostic, and epidemiological studies are needed to help define the manifestations and burden of neurological disease caused by SARS-CoV-2.
19	Galanopoulou ASet al. (2020)	Retrospective analysis	COVID-19 (30-83 years old) patient records.	Sporadic epileptiform discharges (EDs) were present in 40.9% of COVID-19-positive and 16.7% of COVID-19-negative patients.
20	Pasini Eet al. (2020)	Observational and correlation analysis	All EEG recordings were analyzed by two experienced neurologists (RM, EP) trained in EEG interpretation.	Repeat EEGs were available in 3 patients and showed marked improvement of posterior activity, in accord with the improvement of neurological symptoms.
21	Pastor Jet al. (2020)	Retrospective and observational analysis	Analyzed scalp EEGs performed in 20 ICU patients (17 men, 3 women) diagnosed with COVID-19.	We objectively identified some numerical EEG features in severely ill COVID patients that can allow positive diagnosis of this encephalopathy.
22	Hirsch LJ, Laroche SM, Gaspard N, et al.	Statistical Analysis	Quantification and Categorization of Sporadic (Non-rhythmic and Non-periodic) Epileptiform Discharges	Significant improvement, with almost perfect inter-rater agreement for main terms and substantial agreement for amplitude and frequency modifiers.
23	Lau J, Ioannidis JPA, Schmid CH	Meta-regression analyses	continuous and binary data	evaluate the effect of individual variables on the magnitude of an observed effect and thus may sometimes explain why study results differ

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

24	Ayub N, Cohen J, Jing J, Jain A, Tesh R, Mukerji SS, et al	Statistical Analysis	37 patients with SARS-CoV-2 who underwent EEG	Patients with epileptiform findings were more likely to have preexisting brain pathology (6/14, 43%) than patients without epileptiform findings (2/23, 9%; p=0.042).
25	Louis S, Dhawan A, Newey C, Nair D, Jehi L, Hantus S, Punia V.	Data Analytics	22 COVID-19 patients, 19 underwent cEEGs, and 3 underwent routine EEGs (<1 h).	There were no acute neuroimaging findings. Periodic discharges were noted in one-third of patients and encephalopathic EEG findings
26	Canham LJW, Staniaszek LE, Mortimer AM, Nouri LF, et al	Statistical Analysis	Mar 1st to Jun 30th, 2020, COVID-19 confirmed cases	Electroencephalogram (EEG) figures as a useful tool to differentiate encephalopathy from nonconvulsive epilepticus status.
27	Pellinen J, Carroll E, et al	Data Analytics	COVID-19 (111 records)	COVID-19 can affect the central nervous system
28	Pilato MS, Urban A, Alkawadri R, et al	Classification Algorithm	Eight COVID-19 positive patients who underwent EEG monitoring	COVID-19 patients with epilepsy may have an increased risk of neurological manifestations and abnormal EEG.
29	Delorme C, Paccoud O et al	Data Analytics	COVID-19-related encephalopathy	All patients improved after immunotherapy
30	Hepburn M, Mullaguri N, George P, Hantus S, Punia V,et al	Data Analytics	critically ill patients with COVID-19	Report two cases of acute symptomatic seizures, in non-epileptic patients, associated with severe COVID-19 disease
31	Mohammadi S, Moosaie F, Aarabi MH.	Classification Algorithm	COVID-19 Data	Development of COVID-19 neurological complications, namely Alzheimer's disease, Parkinson's disease, stroke, multiple sclerosis.
32	Assenza G, Lanzone J et al.	Data Analysis	COVID-19 Data	Patients executed hyperventilation only for real clinical needs, but often (56%) with a mask.
33	Emami A, Fadakar N, et al.	Classification Algorithm	COVID-19 (6147 records)	Critically ill patients should be treated as acute symptomatic seizures.
34	El-Kafrawy, N. M., Hegazy, D.,	SVM	EEG 4 classes on BCI Competition IV 2a	Classification Accuracy 100%.
35	Liu, T., and Yang, D. (2021).	Deep Learning Algorithm(CNN)	EEG Signal Data	Pre-processing performances increased by adopting different techniques
36	Vallabhaneni, R. B., Sharma, P., Kumar, V., et al.	CNN, RNN, DBN	EEG Signal Data	Improved the performance of decoding difficulties caused by changes in signal distribution

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

In the studies reviewed, 96.1% of participants exhibited abnormal background activity, potentially indicating encephalopathy. EEG is a critical tool for monitoring changes in mental status, as encephalopathy can manifest through a variety of symptoms, including fever, hypoxia, and altered consciousness.

RQ4: Which publications are statically dominant in the area of biomedical signal data classification?
IEEE Xplore, Springer, Inderscience, Elsevier, and Wiley were all taken into account in our research. The majority of the research has been published on Springers and Sciencedirect, individually with 28% and 19% of the total being published. In addition, 17% studies have been published in the Inderscience. Wiley published 15% studies, whereas IEEE Xplore published 9% studies. Journal of Systems and Software has published 2% studies in bio medical signal classification under eeg, Expert Systems with Applications(ESA) has published 2% studies, Information and Software Technology (IST) has published 2%, Empirical Software Engineering (ESE) published 1% and Applied Soft Computing (AS) with 2% studies published in the domain of biomedical signal analysis in disease prediction.

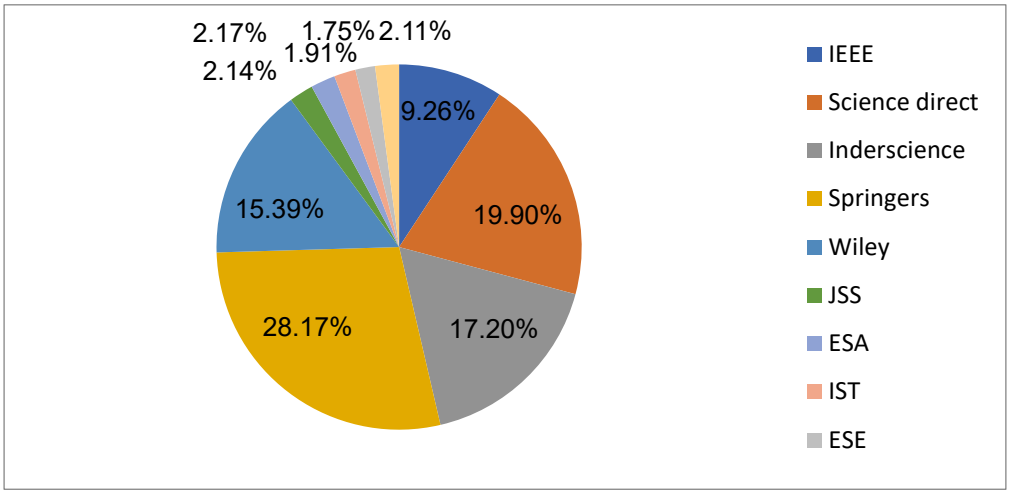

Figure 2. List of publication on EEG findings

Table 2: Year wise publication details

Electronic Sources/ Databases	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
IEEE Xplore	5	11	7	6	10	12	26	35	43	43	71
Springer	46	41	38	46	52	56	59	108	94	104	121
Inderscience	23	19	14	31	39	44	46	61	66	72	83
Wiley	4	6	14	10	6	26	49	109	91	77	70
Science Direct	5	10	13	7	27	40	30	59	115	113	171
Expert Systems with Applications (ESA)	3	1	11	11	1	2	11	3	7	1	2
Information and Software Technology (IST)	3	3	3	3	6	4	1	7	12	4	6
Empirical Software Engineering (ESE)	2	2	6	8	4	1	3	6	8	3	5

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

Journal of Systems and Software(JSS)	4	6	5	5	2	4	2	7	9	10	7
Applied Soft Computing (AS)	5	3	4	3	3	0	0	9	15	10	6
Total	100	102	115	130	150	189	227	404	460	437	542

RQ5: Evaluation of various machine-learning algorithms' performance using validation techniques?
Conducting a performance analysis of different machine learning algorithms for disease prediction using biomedical signals involves several steps, including selecting appropriate algorithms, validation methods, and metrics for evaluation. In this analysis, validation methods are K-Fold cross-validation, Leave-one-out cross-validation; Train-Test split and stratified sampling methods. Some of the most widely used performance metrics methods are accuracy, precision, recall, F1 score and ROC curve.

Table 3: Performance of various machine-learning methods on different datasets

Disease prediction by EEG findings	Algorithms	Performance metrics (%)				
		Accuracy	Precision	Recall	F1Score	ROC curve
Epilipsey	SVM	84	79	74	76	0.81
	RF	87	82	70	80	0.85
	DT	83	78	72	75	0.8
	LG	85	80	75	77.5	0.82
	GB	86	83	79	81	0.87
	XGB	88	84	80	82	0.9
	Bagging	85	81	76	78	0.82
Alzheimers Disease	SVM	85	80	78	79	0.86
	RF	87	83	80	81	0.88
	DT	80	75	72	73	0.78
	LG	82	76	74	75	0.83
	GB	84	79	77	78	0.85
	XGB	86	82	79	80	0.9
	Bagging	83	78	75	76	0.82
Multiple Sclerosis	SVM	84	78	76	77	0.85
	RF	86	83	81	80	0.86
	DT	81	78	71	76	0.8
	LG	81	75	73	74	0.82
	GB	85	81	80	79	0.83

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

	XGB	87	81	79	78	0.89
	Bagging	83	82	76	77	0.81
Traumatic Brain Injury	SVM	81	77	75	76	0.81
	RF	86	79	77	76	0.87
	DT	80	75	71	78	0.79
	LG	82	77	79	77	0.8
	GB	84	83	77	81	0.82
	XGB	87	83	79	80	0.87
	Bagging	84	79	76	77	0.81
Brain Tumors	SVM	83	76	74	80	0.81
	RF	88	82	80	81	0.89
	DT	81	79	74	76	0.8
	LG	85	80	76	78	0.82
	GB	86	84	80	82	0.87
	XGB	88	83	80	81	0.89
	Bagging	83	80	77	78	0.82

Table 4: Performance of various machine-learning algorithms based on validation methods under EEG

Accuracy(%) of validation methods on EEG findings	SVM	RF	DT	LG	GB	XGB	Bagging
K-Fold Cross-Validation	88	88	83	82	88	89	84
Leave-One-Subject-Out Cross-Validation	88	88	81	81	88	90	86
Train-Test Split	87	91	78	85	88	90	83
Stratified Sampling	88	90	82	85	89	92	88

Table 4enumerates the validation processes carried out to confirm the outcomes of different strategies. The validation techniques such as k-fold cross validation, leave-one-subject-out cross-validation, and train test split and stratified sampling methods are almost equal in accuracy rate (%) on prediction of disease by considering bio medical signals data like EEG. Specifically among all the validation methods, extreme gradient boosting algorithm shows some high accuracy rate compared with other machine learning algorithms by considering EEG findings. This meta-analysis and comprehensive review focused on evaluating EEG outcomes across various patient groups. During the pandemic, EEG monitoring was employed in critically ill patients to manage neurological complications and optimize medical facility resources. Among the patients, 23% experienced epileptiform discharges (EDs), despite not being proven beyond a reasonable doubt whether these discharges were more

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

prevalent in those with pre-existing epilepsy. The study reported that 0.5% of patients had clinical seizure episodes, while 0.03% experienced status epileptics. The low frequency of such episodes provides valuable insights for healthcare providers, particularly in resource-constrained environments, aiding in the prioritization of care.

4. Conclusion and Future Scope

This research improves our comprehension of the EEG and clinical features in patients of different ages, such as young people and senior citizens, and both males and females, who presented with neurological symptoms. The findings can help inform EEG triage policies in hospitals, especially during the treatment of diseases affecting the nervous system. Further research in EEG findings and their clinical states, short- and long-term prognoses, may assist clinicians in determining which patients would gain the most from EEG surveillance, ultimately leading to improved clinical outcomes. Future perspectives for research on EEG signal processing in individuals with neurological and biological problems are provided by this review. Despite recent advancements, significant challenges remain in EEG-based BCI systems and signal processing due to our limited understanding of the brain's complexities. One major issue is the low cross-subject accuracy, highlighting the limited generalization of current models. The goal of future studies should be to create more resilient systems with enhanced generalization capabilities, while reducing the computational time required, thereby addressing these open questions in the field.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] H. Yadav and S. Maini, "Electroencephalogram based brain-computer interface: Applications, challenges, and opportunities," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 82, pp. 47003-47047, 2023.
[2] A. Cataldo, S. Criscuolo, E. De Benedetto, A. Masciullo, M. Pesola, and R. Schiavoni, "Uncovering the correlation between COVID-19 and neurodegenerative processes: Toward a new approach based on EEG entropic analysis," *Bioengineering*, vol. 10, no. 4, p. 435, 2023.
[3] O. E. Korkmaz, O. Aydemir, E. A. Oral, and I. Y. Ozbek, "Investigating the effect of COVID-19 infection on P300 based BCI application performance," *Traitement du Signal*, vol. 38, no. 6, pp. 1767-1773, 2021.
[4] G. B. Tantillo et al., "Electroencephalography at the height of a pandemic: EEG findings in patients with COVID-19," *Clin. Neurophysiol.*, vol. 137, pp. 102-112, May 2022.
[5] Y. Yao et al., "Study on brain damage patterns of COVID-19 patients based on EEG signals," *Front. Hum. Neurosci.*, vol. 17, p. 1280362, 2023.
[6] L. A. Corazza et al., "Electroencephalographic findings among inpatients with COVID-19 in a tertiary hospital from a middle-income country," *Arq. Neuropsiquiatr.*, vol. 79, no. 4, pp. 315-320, 2021.
[7] I. Sáez-Landete et al., "Retrospective analysis of EEG in patients with COVID-19: EEG recording in acute and follow-up phases," *Clin. EEG Neurosci.*, vol. 53, no. 3, pp. 215-228, May 2022.
[8] Y. Yang, T. Yu, and J. Yang, "Clinical manifestations and EEG findings in children infected with COVID-19 and exhibiting neurological symptoms," *BMC Pediatr.*, vol. 24, p. 49, 2024.
[9] B. Gogia et al., "EEG characteristics in COVID-19 survivors and non-survivors with seizures and encephalopathy," *Cureus*, vol. 13, no. 10, p. e18476, 2021.
[10] O. Karadas, B. Ozturk, and A. R. Sonkaya, "EEG changes in intensive care patients diagnosed with COVID-19: A prospective clinical study," *Neurol. Sci.*, vol. 43, pp. 2277–2283, 2022.
[11] A. R. Antony and Z. Haneef, "Systematic review of EEG findings in 617 patients diagnosed with COVID-19," *Seizure: Eur. J. Epilepsy*, vol. 83, pp. 234–241, 2020.
[12] L. Marinelli et al., "The value of EEG attenuation in the prediction of outcome in COVID-19 patients," *Neurol. Sci.*, vol. 43, pp. 6159–6166, 2022.
[13] C. Tsai, S. E. Wilson, and C. Rubinos, "SARS-CoV-2 infection and seizures: The perfect storm," *J. Integr. Neurosci.*, vol. 21, no. 4, p. 115, 2022.
[14] W. Guan et al., "Clinical characteristics of coronavirus disease 2019 in China," *N. Engl. J. Med.*, vol. 382, pp. 1708-1720, 2020.
[15] L. Mao et al., "Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China," *JAMA Neurol.*, 2020.
[16] A. Varatharaj et al., "Neurological and neuropsychiatric complications of COVID-19 in 153 patients: A UK-wide surveillance study," *Lancet Psychiatry*, 2020.
[17] M. U. Ahmed et al., "Neurological manifestations of COVID-19 (SARS-CoV-2): A review," *Front. Neurol.*, vol. 11, p. 518, 2020.
[18] M. A. Ellul et al., "Neurological associations of COVID-19," *Lancet Neurol.*, vol. 4422, pp. 2-3, 2020.
[19] A. S. Galanopoulou et al., "EEG findings in acutely ill patients investigated for SARS-CoV-2/COVID-19: A small case series preliminary report," *Epilepsia Open*, vol. 5, pp. 314–324, 2020.

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025

[20] E. Pasini et al., "EEG findings in COVID-19 related encephalopathy," *Clin. Neurophysiol.*, vol. 131, pp. 2265–2277, 2020.
[21] J. Pastor, L. Vega-Zelaya, and E. Martín Abad, "Specific EEG encephalopathy pattern in SARS-CoV-2 patients," *J. Clin. Med.*, 2020.
[22] L. J. Hirsch et al., "American clinical neurophysiology society's standardized critical care EEG terminology: 2012 version," *J. Clin. Neurophysiol.*, vol. 30, pp. 1–27, 2013.
[23] J. Lau, J. P. A. Ioannidis, and C. H. Schmid, "Quantitative synthesis in systematic reviews," *Ann. Intern. Med.*, vol. 127, pp. 820–826, 1997.
[24] N. Ayub et al., "Clinical electroencephalography findings and considerations in hospitalized patients with coronavirus SARS-CoV-2," *MedRxiv*, 2020.
[25] S. Louis et al., "Continuous electroencephalography characteristics and acute symptomatic seizures in COVID-19 patients," *Clin. Neurophysiol.*, vol. 131, pp. 2651–2661, 2020.
[26] L. J. W. Canham et al., "Electroencephalographic (EEG) features of encephalopathy in the setting of COVID-19: A case series," *Clin. Neurophysiol. Pract.*, 2020.
[27] J. Pellinen et al., "Continuous EEG findings in patients with COVID-19 infection admitted to a New York academic hospital system," *Epilepsia*, 2020.
[28] M. S. Pilato et al., "EEG findings in coronavirus disease," *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.*, 2020.
[29] C. Delorme et al., "COVID-19-related encephalopathy: A case series with brain FDG-PET/CT findings," *Eur. J. Neurol.*, 2020.
[30] M. Hepburn et al., "Acute symptomatic seizures in critically ill patients with COVID-19: Is there an association?," *Neurocrit. Care*, 2020.
[31] S. Mohammadi, F. Moosaie, and M. H. Aarabi, "Understanding the immunologic characteristics of neurologic manifestations of SARS-CoV-2 and potential immunological mechanisms," *Mol. Neurobiol.*, vol. 57, 2020.
[32] G. Assenza et al., "Electroencephalography at the time of COVID-19 pandemic in Italy," *Neurol. Sci.*, vol. 41, pp. 1999–2004, 2020.
[33] A. Emami et al., "Seizure in patients with COVID-19," *Neurol. Sci.*, vol. 41, pp. 3057–3061, 2020.
[34] N. M. El-Kafrawy, D. Hegazy, and M. F. Tolba, "Features extraction and classification of EEG signals using empirical mode decomposition and support vector machine," in *Adv. Mach. Learn. Technol. Appl.*, Springer, 2014, pp. 189–198.
[35] T. Liu and D. Yang, "A three-branch 3D convolutional neural network for EEG-based different hand movement stages classification," *Sci. Rep.*, vol. 11, p. 10758, 2021.
[36] R. B. Vallabhaneni et al., "Deep learning algorithms in EEG signal decoding application: A review," *IEEE Access*, vol. 9, pp. 125778–125786, 2021.

DOI: https://doi.org/10.54216/FPA.190223
Received: January 02, 2025 Revised: February 03, 2025 Accepted: March 02, 2025