IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 18 September 2024, accepted 14 November 2024, date of publication 19 November 2024, date of current version 29 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3502540

RESEARCH ARTICLE

Early-Stage Non-Severe Depression Detection Using a Novel Convolutional Neural Network Approach Based on Resting-State EEG Data

PASCAL PENAVA AND RICARDO BUETTNER, (Senior Member, IEEE)
Chair of Hybrid Intelligence, Helmut-Schmidt-University/University of the Federal Armes Forces Hamburg, 22043 Hamburg, Germany

Corresponding author: Ricardo Buettner (buettner@hsu-hh.de)

This work was supported by the Open-Access-Publication-Fund of the Helmut-Schmidt-University/University of the Federal Armed Forces Hamburg.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Medical Faculty of the University of Leipzig under Reference No. 154/13-ff.

ABSTRACT Over 300 million people worldwide are affected by depression, with symptoms that have a major impact on patients and, in the worst cases, can lead to suicide. As the severity of the disease increases over time, early detection can save a patient's life. The disease is diagnosed by professionals using questionnaires that might be influenced by biases, and of which the accuracy and reliability are not guaranteed. For this reason, an increasing number of studies are looking at physiological ways of detecting the disease, with electroencephalogram-based machine learning prediction models having been successful in recent years. However, the focus is not on the early detection of mild depression, which can be the entry point to major depression. In this work, we developed a deep learning based model using a 1D convolutional neural network to detect mild depression in resting-state electroencephalogram data. We evaluated the model using a realistic world-like dataset and were able to achieve a balanced accuracy of 69.21%. With this result, we are setting a new benchmark for resting-state-based early detection. Due to the low level of preprocessing and the associated fast computing time and low computational intensity, our innovative approach can serve as a basis for applications in the real world. This enables patients with suitable hardware to recognize the disease themselves at an early stage and thus receive timely treatment to prevent further development.

INDEX TERMS Depression detection, early-stage, EEG, resting-state, CNN.

I. INTRODUCTION

Up to 800,000 people commit suicide worldwide, with depression being one of the biggest causes [1]. With over 300 million people affected, depression is now one of the most widespread diseases [2]. In patients suffering from major depression, the illness can cause the person to have suicidal tendencies [3]. As a result, one in three patients with major depressive disorder (MDD) has attempted suicide [4]. Following the COVID-19 pandemic, the prevalence of people with psychiatric disabilities and depression has also increased [5]. At 80%, the largest proportion of psychiatric disabilities and suicides occurs in middle- and low-income countries [1], while at the same time, the availability of treatment options is lowest there [6]. Due to the serious effects of the disease and the limited treatment available, it is essential to develop a simple and accessible method for diagnosing depression.

There are usually no objective physiological tests to diagnose psychological illnesses [7]. Likewise, depression is also diagnosed using questionnaires [8], although their reliability is increasingly being questioned [9], [10]. In addition, questionnaires can generally have the disadvantage of containing biases, which might limit their objectivity [11]. Due to similar symptoms, two individuals with comparable complaints, for example, may receive the same diagnosis

The associate editor coordinating the review of this manuscript and approving it for publication was Guolong Cui.

despite having different diseases [12]. Since recognition in this way also depends heavily on the patient, they can influence it, for example, by not showing their real emotions and trying to pretend something to the doctor [13].

For this reason, more and more work is being done on the computer-aided automatic detection of MDDs in patients [14], [15], [16] and Machine Learning (ML) as well as Deep Learning (DL) methods in combination with non-invasive electroencephalogram (EEG) data in particular have recently shown potential for its recognition [17]. Several works have already achieved a classification accuracy of over 90% [18], [19], [20], using ML methods such as logistic regression [18] or support vector machines [19], as well as DL methods such as Convolutional Neural Networks (CNNs) [20]. Figure 1 compares between the traditional approach using questionnaires and the modern approach using physiological data.

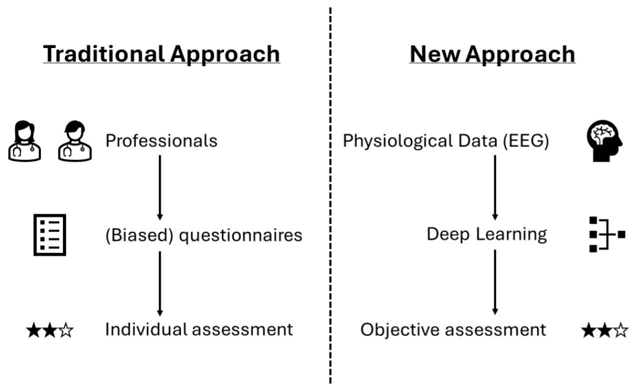

FIGURE 1. Computer aided depression detection.

One problem with depression is that it only affects a patient's behavior over time [21], and patients do not seek professional help because of the social stigma [22], [23]. Without treatment, mild depression can develop into MDD and cause the associated serious effects [24], and thus early detection of the disease may even prevent suicides [25]. However, most research focuses on MDD classification rather than early detection of mild depression, making early intervention difficult [26]. Thus, one of the main reasons for MDD disease is not getting timely treatment [27].

The automated detection of mild depression based on EEG data is significantly more difficult than that of MDD [28]. Although, a few works have nevertheless achieved good results based on EEG data generated with stimuli, using CNNs [26], [29], and ML methods [30]. Using resting-state EEG data, there is only one study that succeeded in detecting mild depression with an accuracy of 68.87% [13].

The state-of-the-art research on computer-based automated depression detection with EEG data reveals a research gap that calls for a solution that reliably detects mild depression using resting-state EEG data. Since many studies are not subject-independent due to the small datasets and the results are therefore not usable for applications in the real world [17], a solution is needed that guarantees that no data leakage occurs during the evaluation of the approach.

In this work, we aim to close this research gap by developing a CNN based on a dataset that well represents real-world conditions, with which we can reliably classify mild depression patients and healthy controls using resting-state EEG data. Thereby, we want to answer the question whether DL methods are suitable for predicting mild depression with resting-state EEG data.

Since it is in the nature of people suffering from depression to seek little help [23], our solution should be designed in such a way that it works without intensive computing power and does not require complex preprocessing, and is therefore also suitable for potential private use.

With regard to the research gap, the main contributions of this work are as follows:

1) With a balanced accuracy of 68.57%, we set a new benchmark for resting-state EEG-based detection of mild depression based on ML methods.

2) We show an innovative 1D-CNN (SIEPTNet) architecture that can be used in practice for early MDD detection without the need for complex and computationally intensive preprocessing of the data.

This work is structured as follows: First, we give an overview of the disease and show symptoms and diagnostic options, before presenting automated options for detecting it. We then show our approach in the methodology section, with the structure of the architecture and the training as well as the evaluation process. In the results section, we then show the outcomes and discuss them in the following section. There we also briefly discuss the practical implications to then complete the work with a summarizing conclusion and potential future work.

II. RELATED WORK

As outlined in the introduction, depression is a serious and far-reaching psychiatric disability that affects over 300 million people worldwide [2]. The symptoms of the disease can be both psychological and physical. For example, sadness and suicidal thoughts are psychological symptoms, while headaches or back pain can be physical symptoms [31]. Even though there are some medical and psychological treatment options for mild to severe depression, there are too few services available, especially in low-income countries [6]. When we talk about depression, we are usually referring to MDD [8]. With industrialization, MDD has become one of the most common diseases in industrialized countries [32]. Historically, depression is diagnosed via questionnaires, which must be conducted by a trained medical professional or psychologist [8].

A. HAMILTON DEPRESSION SCALE

The Hamilton Depression Scale is one of the most widely used methods/questionnaires for diagnosing depression and its severity [17]. The scale is a questionnaire, which originally consists of 21 questions, each of which is rated with scores of 0-4 or 0-2, depending on the item. The individual scores of the items are then added together to form the Hamilton Depression Score, which is used to diagnose depression and determine its severity [33].

There are four levels for the severity of depression. An overall score of 0-6 corresponds to no depression. Scores of 7-17 correspond to mild depression, while a score of 18-24 equates to moderate depression. Scores greater than 25 again indicate that the patient has severe depression [34].

B. DEPRESSION DETECTION IN EEG DATA

Depression is a neuronal disorder that is usually diagnosed using scales such as the Hamilton Depression Scale [33]. However, there has been increasing criticism of such questionnaires for detecting depression for some time now and their validity has been questioned [9], [10]. For this reason, more and more studies are investigating ways of detecting the disease using other measures. With the rapid development of algorithms, ML methods combined with (resting-state) EEG data have emerged as a promising method for diagnosing MDD [17].

EEG data is a cost-effective and easy-to-use way of recording physiological signals and therefore is suitable for real-world scenarios [35]. Many studies have found correlations and biomarkers for depression in EEG data. With the hippocampus, the prefrontal cortex, and the amygdala, depression mainly affects three areas of the brain. These three areas are located in the temporal lobe, frontal lobe, and frontal section of the temporal lobe of the brain, in the order mentioned [8]. Especially in the right area of the brain and in the frontal and parietal-occipital cortex, many studies have shown connections with depression [36]. A study has found that patients with MDD may have functional dysfunction in the left hemisphere [37]. Frontal asymmetry, which describes alpha band differences in the anterior right and left brain regions, was also found to be a marker for depression risk in the EEG data, with less activity in the left area indicating depressive symptoms [38]. Last but not least, it also seems that in subjects with depressive illness, the brain structure appears more random [39].

Even if the markers are not always consistent between the different research studies, they suggest that there must be a connection between depression and the EEG signals. Based on this information in the data, it should be possible to recognize the differences between depressed and healthy subjects. Since it is very difficult for a human being to detect these differences visually, automated systems and algorithms are needed to perform this task [26]. DL and ML methods have proven to be suitable for this task in recent years [40].

1) MAJOR DEPRESSIVE DISORDER

Therefore, many studies have used these bio-markers to predict MDD in subjects based on EEG data. Most studies proceed in such a way that they make a binary classification by taking a healthy control group and a second group with depressed subjects (regardless of their severity).

Cai et al. [41], for example, measured the EEG data in combination with sound impulses and in the resting-state using three electrodes and were able to predict MDD with 79% and 76% using the k-nearest neighbor algorithm [41].

With 90% and 98% respectively, Hosseinifard et al. [18] and Khan et al. [20] achieved far better results. While Hosseinifard et al. [18] use logistic regression classifiers, Khan et al. [20] use 2D CNNs for the prediction of MDD.

Sharma et al. [19] found an even better accuracy of 99.58% for the classification of MDD. They used a least square support vector machine for the prediction.

Kaushik et al. [42] compared in their work whether MDD can be better predicted with resting-state or task/stimuli-related EEG data. They used several DL methods to predict the vulnerability of MDD. They found that CNNs work best and that resting-state data is better suited for prediction [42].

Wu et al. [17] have addressed the problem of many detection studies that they work with a small dataset, which means that most studies are not subject-independent or data leakage occurs. Subject-independence means that no information about people from the training dataset ends up in the test dataset. For this reason, Wu et al. [17] used a large dataset and achieved a balanced accuracy of 84% in the detection of MDD subjects using support vector machines by means of strict cross-validation [17].

In addition to these examples, many other studies have detected MDDs using various ML methods with a maximum accuracy of over 90% [14], [15], [16]. This study selection shows that ML and DL methods are well-suited for detecting MDD in subjects.

2) EARLY STAGE DETECTION

Contrary to the fact that the majority of the research is concerned with the detection of MDDs, most of the people suffering from the disease only have mild depression [43]. Mild depression develops in severity over time [24] and can develop into MDD without treatment. However, there is no focus on early detection of the disease, which prevents early intervention [26]. Cognitive impairment increases with the severity of the depression, which means that earlier stages of the disease are more difficult to detect without testing than more advanced stages [28]. Even if the focus is not on recognizing mild depression, there are few works that have addressed the topic.

Li et al. [26], for example, predicted mild depression with an accuracy of 85.62% using CNNs. They used images of facial expressions to emotions as stimuli. Here they used transfer learning with CNNs and a combination of the EEG data, their RGB images once with and once without location information with leave one subject out validation [26].

In an earlier paper, Li et al. [30] also developed a mild depression detection algorithm and aimed to find biomarkers in the EEG data that are related to mild depression. Here, too, facial expression emotion images served as stimuli, and several ML algorithms were used to make the predictions.

With 92% to 98%, the authors achieved very good results and were also able to find that the left parietotemporal lobe in the beta EEG frequency band has a greater effect on the detection [30].

Using the same stimuli, Li et al. [29] again attempted to detect mild depression using EEG data in a more recent study with an innovative CNN method. Here, however, they use a more complex approach by using functional connectivity matrices of the EEG data as an image and taking these as input for a 2D CNN. With this approach, the authors achieve an accuracy of 80.74% [29].

A less complex approach that requires fewer preprocessing steps was implemented by Thulasi et al. [13]. The authors used resting-state EEG data recorded with the consumer EEG headset Neurosky Mindwave Mobile 2. Using several ML algorithms, they tried to detect mild depression in the data and achieved an accuracy of 68.87% with a support vector machine [13].

In addition to the few studies that deal with the classification of subjects as healthy or with mild depression, many studies have carried out further severity detection in addition to the classification into healthy vs MDD after diagnosis with MDD. For example, Mahato et al. [44] predicted the severity based on the resting state EEG data with 79% after the detection of depression with 96% using support vector machines. Li et al. [45] first detected depression using a CNN with an accuracy of 66.4% and then detected the severity of the depression with an accuracy of 66.93%. The authors note that the results are inferior in comparison, but this may be due to the minimal preprocessing and cleaning of the data in their work, as artifacts can have a large impact on predictive power [45].

TABLE 1. Related work for EEG based mild depression detection.

Author	Year	Method	Stimuli	Accuracy	Subjects
Thulasi et al. [13]	2023	Support Vector Machine	Resting-State	68.87%	55
Li et al. [29]	2020	CNN (functional connectivity matrices)	Emotional Facial Expressions	80.74%	51
Li et al. [26]	2019	Transfer Learning with CNN	Emotional Facial Expressions	85.62%	51
Li et al. [30]	2016	Several ML algorithms	Emotional Facial Expressions	92% to 98%	37

3) OUR WORK

From the presented state-of-the-art research in the field of EEG-based depression detection, it quickly becomes apparent that the distinction between a healthy control group and mild/non-severe depression is a little-studied phenomenon, which can also be seen in table 1. Many of the studies deal with MDDs and use specially prepared data sets, although early detection of the disease is also very important, as shown.

Of the few studies that deal with this topic, only one works with resting-state EEG data since it is particularly difficult to predict non-severe depressions, as the characteristics of the effect are only very weakly pronounced in these [46].

Last but not least, some of the studies use datasets in which the subjects with MDD are already being treated with medication, which can falsify the classification results, as these can have an influence on the EEG data [47]. On the other hand, there are data leakage problems in many studies, whereby information from the same subject gets into both the training and test data, which can also distort the results [12].

Therefore, in this work, we want to investigate whether we can predict mild depression at an early stage using resting-state EEG data, based on a dataset that better represents the real-world setting and does not focus on subjects with MDD. We take care to ensure subject-independence and use modern DL methods for prediction.

III. METHODOLOGY

The aim of this work is to show a CNN model architecture for the detection of depression that is able to recognize the disease in a early-stage on the basis of resting-state EEG data. In this way, we can show that modern ML methods can support the analysis of the disease, making it possible to take preventive measures. Using a novel 1D CNN architecture, we set the benchmark based on a new dataset. The following section provides an overview of the materials used in the analysis and the methods employed. We then apply a K-Fold Cross Validation during the training process.

Since modern ML methods and CNNs in combination with EEG data have already achieved success in depression detection [17], we first built a CNN architecture adapted to the dataset. To prevent data leakage during the optimization process, we do not apply optimization due to the relatively small dataset, but evaluate the metrics using the trained model.

A. DATA PREPROCESSING

Most EEG devices provide their data in the ‘‘.eeg’’, ‘‘.vmrk’’ and ‘‘.vhdr’’ formats via interfaces. In this work, the EEG data are also available in these formats, and we first load them and have to convert them into the correct format for processing by the CNN. The data is available in raw form, and our aim is to preprocess the data as little as possible to avoid a time-consuming and hardware-intensive process for real-world applications. Minimal preprocessing can have the disadvantage of increasing noise in the data, which can affect the accuracy of a classifier. However, since the removal of noise from the raw EEG data and the associated preprocessing steps significantly increase the computational intensity, it would complicate a real-world application.

Therefore, we first apply only a notch filter to remove the power line noise [48] and resample the data at 250 Hz. In addition, we filter out the data points marked with ‘‘eyes open’’ so that we only work with the ‘‘eyes closed’’ data. Then we remove the base channel from the EEG data. Typically, EEG data represent time series that are of a complex nature and are processed using more complex ML methods such as Long Short-Term Memory networks. However, Buettner et al. [49], for example, showed that spectral analysis can be used to reduce complexity by converting the time series data into frequency domain data. Therefore, we apply a power spectral density (PSD) method according to Welch's approach [50], using a lower starting frequency of 0.5 Hz and not limiting the EEG subbands to the upper end of the spectrum. We use the PSD method of the MNE library [51]. Welch's algorithm works by calculating the distribution of the data using a sliding window, calculating the periodogram in the segments, and then averaging all estimates of the segments [50]. Classically, EEG data is divided into five coarse subbands (delta, theta, alpha, beta, and gamma). However, the method using PSD offers greater information content, as the data is divided into finer sub-bands [52], [53], [54].

The CNN model can now be trained with the preprocessed data. The data is transferred to the model in ".csv" format, whereby the data represents one row per subject and no subject occurs in two rows, thereby avoiding data leakage and guaranteeing subject-independence. One row consists of 62 elements times 1,020 frequency bands.

As a final step, we apply Synthetic Minority Oversampling Technique (SMOTE) to the training data during preprocessing. We made sure that SMOTE was only applied to the training data in each fold individually after splitting the data in order to avoid leakage. With SMOTE, new samples are generated based on the interpolation of neighboring instances of the minority class. On the one hand, this ensures that the data set is balanced and, on the other hand, serves the purpose of avoiding overfitting in the model [55]. Overfitting occurs when there is not enough data available and the model learns the data by heart and does not generalize and thus achieves good results in classifying the training data but does not work on the unseen test data [56].

B. MODEL ARCHITECTURE

The model is adapted to the data of the ".csv" file and thus has an input layer that expects a one-dimensional vector (per subject) containing the 62 channels times 1,020 frequency bands. Figure 2 shows the architecture of our model. By reducing the complexity of the data, we were able to design a simpler architecture with a 1D CNN, which, for example, does not have to consider temporary dependencies.

SIEPTNet was designed by us and represents an architecture that is suitable for the classification of EEG data. In general, more complex CNN architectures outperform simpler ML algorithms for most applications [52]. SIEPTNet has a total of 13 layers. All three convolutional layers are one-dimensional and have a kernel size of three with 64 filters. Each of the convolutional layers is followed by a batch normalization layer. Layers five, nine, and ten are each max pooling layers with a pool size of two and where padding is permitted. A dropout layer with a rate of 0.8 is added after the first and third max pooling layer to avoid overfitting. The last two layers are a standard flat layer and finally a fully connected dense layer with a single output unit and a sigmoid activation function. The model is compiled with an Adam Optimizer [57] and a fixed learning rate of 1e-4 using a binary cross-entropy loss function.

SIEPTNet is orientated towards EEGNet, a compact CNN for the classification of EEG data [58], but on the contrary is a 1D CNN specialized in the classification of EEG data pre-processed by spectral analysis. In the last layer, the model outputs the binary decision whether a person is healthy or in the early stages of depression. Compared to the 2D convolutions used in EEGNet, SIEPTNet can calculate with less complex 1D convolutions and has the advantage of not having to perform a temporal analysis, since spectral data is used.

C. PROCESS OF TRAINING

For the training of the SIEPTNet model, the subjects in the database are divided into training and testing splits with a ratio of 90/10. Since we use cross validation, a validation set is not required. In addition, we do not optimize the model, which guarantees that there is no data leakage when training with the test data.

We use a K-fold cross validation with 10 folds. The multiple folds ensure that the performance of the model is generalizable and does not depend on randomly lucky splits of the data set. In K-fold cross validation, the data is divided into K segments of equal size, one of which is left out for testing in each iteration. The overall accuracy of the model is then averaged across the folds to obtain a valid result. To ensure that each fold contains the same percentage distribution of classes, we used a stratified K-fold. This also increases the generalizability of the results [59].

With each iteration, the model is trained with 90% of the subjects and tested with 10% of unseen subjects. Since we want to avoid data leakage, we tested the model without optimization. Often the data is not cleanly separated and information from the test data is transferred to the model during hyperparameter tuning, for example [17]. However, since we want to achieve strict subject independence, we have not optimized or tuned the model and can therefore reliably determine via the 10 folds whether depression can be predicted in early stages with unknown subjects. The model is reloaded for each fold and therefore contains no information.

D. EVALUATION METRICS AND PROCEDURE

During training, several evaluation metrics are measured and saved in each fold, from which the final performance metrics are derived by averaging. We used the following metrics for this: Accuracy, Balanced accuracy, True positive rate (sensitivity or recall), True negative rate (Specificity), and the Area Under the Curve (AUC) score.

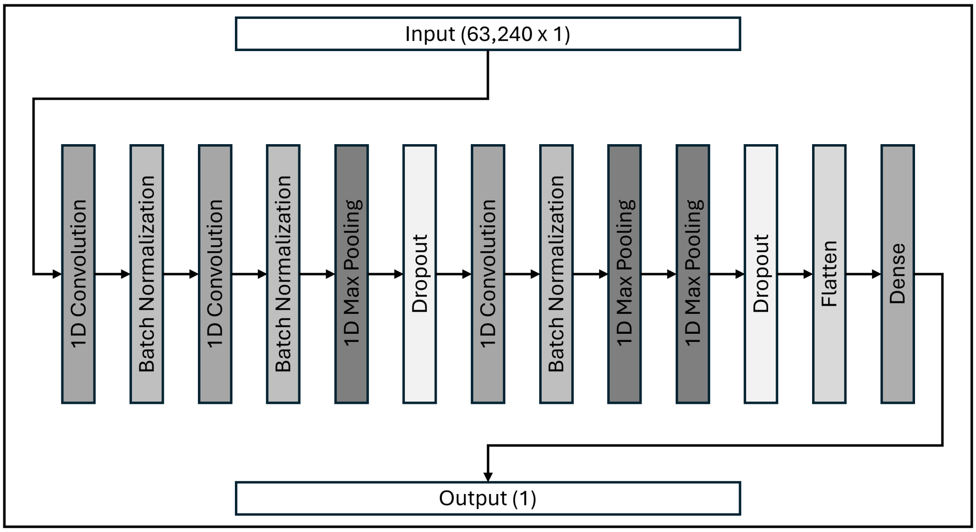

FIGURE 2. SIEPTNet architecture.

First, the accuracy of the model indicates how effective it is overall without taking into account the distribution of classes in the dataset [60]. Balanced accuracy is used to calculate these imbalances. It represents the arithmetic mean of sensitivity and specificity [61], [62]. While the true positive rate indicates how well positive classes are correctly classified [63], the true negative rate indicates how accurately negative classes are assigned [60].

In addition to the evaluation metrics presented, we use the averaged confusion matrix across the 10 folds to gain deeper insights into the classification errors.

E. DATASET

As Cukic et al. [12] noted in their review, there is no standardized dataset for the detection of depression from EEG data in general. As described in the research background, most work deals with the detection of MDD. In this work, however, our goal is to achieve early detection of depression, which is why many datasets are not suitable for us.

We therefore chose a dataset that is part of the Leipzig Study for Mind-Body-Emotion Interactions (LEMON) dataset [64] to demonstrate the quality and reliability of our approach. Although the dataset does not focus on depression detection, it contains the resting-state EEG data of 203 subjects, which were recorded using 62 electrodes. There are 16 minutes of resting-state EEG data per subject, eight of which were recorded and labeled with eyes closed and eight with eyes open.

Although the LEMON dataset is not designed for depression detection, many past and present psychiatric syndromes were assessed using the Standardized Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders. The severity of the depression was also measured using the Hamilton Depression Scale.

Table 2 shows the distribution of the summed Hamilton scores achieved. As explained in the Research Background, the Hamilton score consists of several items that are assessed individually and then added together to form a score. This summed score determines the severity of the depression. A total score of less than seven indicates no depression, while scores between seven and 17 indicate mild depression. Scores above 17 correspond to moderate and severe depression [65]. In order to create a binary classification problem, we adhered to the cut-off values and grouped scores 0-5 into the healthy group and subjects with scores greater than six into the mild depression group. We leave out the six subjects with a score of six to have a clear separation.

TABLE 2. Hamilton score distribution.

Score	0	1	2	3	4	5	6	7	8	9	10	11	14
Count	58	45	34	25	20	13	6	6	2	3	3	2	1

In the LEMON data set, EEG data was measured from a total of 211 subjects. After we merged the EEG data with the Hamilton scores according to subject id, 203 subjects remained as a database from which both EEG data and Hamilton scores were available [64]. Of these 203 subjects, 187 are in the healthy group while 16 suffer from moderate depression. Even though the data set thus contains a small number of affected subjects, it is very well suited for our research goal of early detection of depression. In the world, around 4.4% of humanity suffers from depression [1], while around 7% of the subjects in the data set have mild depression. The data set therefore represents the real world scenario quite well and thus serves as a good basis for investigating whether depression can be recognized in the early stages in realistic scenarios.

IV. RESULTS

For the development of our CNN we used Google Colaboratory (Colab), a cloud service from Google based on Jupyter Notebook. In Colab we used an NVIDIA Tesla T4 GPU with up to 15 gigabytes of GPU-RAM and up to 12 gigabytes of system RAM.

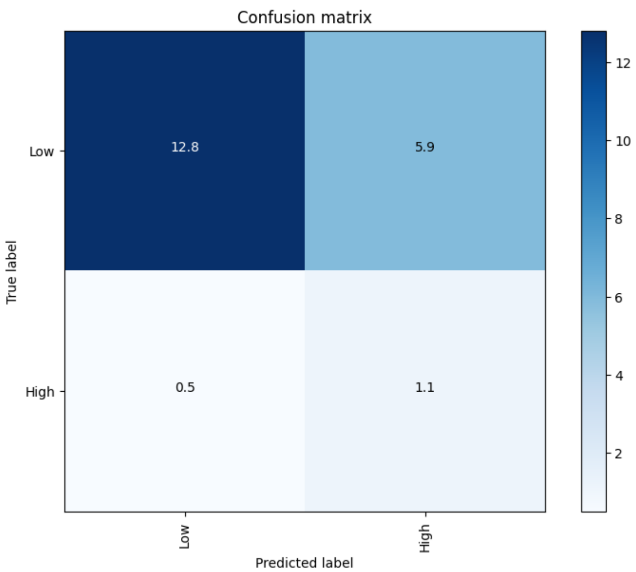

FIGURE 3. Confusion Matrix - 10-fold cross-validated.

TABLE 3. Evaluation indicators of our approach.

Performance indicator	Value
Accuracy	68.57%
Balanced accuracy	69.21%
True positive rate (Sensitivity)	70.00%
True negative rate (Specificity)	68.42%

Figure 3 shows the confusion matrix with the average results of the ten folds. To calculate the values, we simply added up the individual confusion matrices of the iterations and divided them by ten, which gave us a reliable average result. The matrix shows how many subjects from the test splits of the cross validation were assigned to which class and is therefore representative for the performance of the prediction.

The individual values of the confusion matrix can also be used to calculate the more detailed evaluation metrics, which are shown in table 3. Again, these are the average values across the 10 folds, as this guarantees that we do not present any lucky splits. Due to the unbalanced dataset, the balanced accuracy value of 69.21% is decisive for the actual performance of the classifier.

V. DISCUSSION

To the best of our knowledge, we are the first to use a CNN for resting-state EEG-based classification of healthy controls and subjects with mild depression. We are therefore setting a new benchmark for this case with a balanced accuracy of 69.21%. Since we have a strong imbalance between the distribution of subjects across the two classes in the underlying data, balanced accuracy is a very meaningful metric in this case, as it takes this imbalance into account.

The confusion matrix in figure 3 also shows once again that the test splits also contain more healthy control subjects than subjects with mild depression. However, because we stratify the data in the ten folds, the distribution in the test split also represents the distribution of the entire dataset and thus approximately the real population. It can also be seen that of the average 1.6 subjects with mild depression per fold, around 68% were correctly classified, which is also evident from the true negative rate. At the same time, of an average of 18.7 healthy control subjects, around 70% were correctly classified, which is also reflected in the true positive rate. For mental health data, the true negative rate is particularly significant. The impact of a patient who would be wrongly classified as a person with a psychiatric disability is lower in cases such as depression than a patient who is wrongly classified as healthy, which is why these metrics need to be monitored closely. The slightly different results of the confusion matrix and the actual performance indicators are due to the automatic truncation of the decimal places in the confusion matrix. The values show that both within the healthy controls and among the subjects with mild depression, the classifier makes equally good decisions, which shows that no class performs excessively well, which would influence the overall result. In our scenario, the more serious error would be to classify a subject with mild depression as healthy. However, this error does not occur conspicuously more frequently, which once again confirms the quality of the model.

Compared to many previous works, we have taken great care to avoid data leakage through cross-validation. Many of the previous works have not been able to guarantee 100% subject-independence through optimization attempts or training with multiple samples of a single subject, which has led to the model being partially trained with data that is very similar to the test data, which can affect performance [17]. For this reason, we deliberately decided not to optimize the model and to use the entire data per subject as one sample, which guarantees that our results are subject-independent. In this way, we have ensured that independence is validated prospectively in order to analyze performance accurately.

Besides our work, there is only the approach of Thulasi et al. [13] who also use EEG data generated without stimuli to predict mild depression. The authors simply report the standard accuracy, which is not entirely meaningful due to the imbalance of the dataset. Nevertheless, we still exceed the result of 68.87% with our method. The other papers that also use resting-state data do not make a binary classification between healthy controls and subjects with mild depression, which makes a comparison difficult. However, like us, another paper used minimal preprocessing and did not remove artifacts, for example, and they distinguished mild depression from MDD with an accuracy of 66.93%. Again, we outperform the reported value with our classifier.

Even if we do not achieve optimal results in the classification of mild depression and healthy controls, our work and the comparison with the state-of-the-art should show that our approach achieves competitive results. In this work, we deliberately decided not to carry out any elaborate preprocessing in order to save computing power and complexity, which in turn increases practicability and usability in the real world. Due to the minimal preprocessing, however, the EEG data with which the model is trained contains more artifacts, which can negatively influence the performance of the classifier [66], [67]. In the future, the architecture shown could be further optimized with computationally more efficient preprocessing methods to remove noise and thus achieve higher accuracy. Similarly, with a larger dataset containing more subjects, the model could achieve enhanced robustness and generalizability, as subject independence would be strengthened.

A. PRACTICAL IMPLICATIONS

As already explained, the nature of the disease makes it difficult for patients to seek help themselves and social stigmas make it difficult to communicate depression publicly [22], [23]. However, as depression can develop from a mild stage into a severe stage [24], early detection is vital [25].

At the same time, over 450 million people worldwide are affected by some form of mental disorder [1], creating a shortage of mental healthcare providers and psychiatrists, which contributes to their overload [68]. In the healthcare sector, the idea of remote rehabilitation has been around for some time, where mobile sensors can be used, for example, to monitor the physiological parameters of patients [69]. The problem with EEG sensing devices, however, was that they were difficult to use and too expensive for private use [70], [71]. However, over the last few years, more and more companies have been involved in the development of more practical and cheaper wireless EEG devices, which has led to several consumer devices now being on the market [72].

These devices could enable mobile (self-)assessment of the disease. The approach shown in our work could contribute to making such mobile sensors possible. The results shown demonstrate the potential for early detection of depression, while at the same time taking care to work in a resource-saving manner, which enables the use on mobile sensors. These sensors could then, for example, enable a preliminary check for the early detection of mild depression, on the basis of which healthcare facilities could take initial preventive measures, while the patients could remain mainly anonymous.

A possible next step would be to integrate the approach shown into a web environment so that patients can access it remotely and obtain information about their depressive status based on their EEG [12]. This provision could lead to more frequent and earlier recognition of depression, which would enable timely treatment and at the same time relieve the burden on the healthcare system.

VI. CONCLUSION

In this work, we have demonstrated a novel method for DL-based early detection of mild depression based on resting-state EEG data. With SIEPTNet, we have demonstrated an innovative 1D CNN architecture that achieves promising results with a balanced accuracy of almost 70% and, at the same time, works in a resource-saving manner, which enables it to be used in the real world.

Thereby, we were able to close the previously identified research gap, which requires a solution for the reliable detection of mild depression using DL methods. We were also able to positively answer the research question of whether DL methods are suitable for predicting mild depression with resting-state EEG data. The state-of-the-art has also shown that many preliminary studies are not subject-independent, which is necessary for use in the real world [17]. Due to the data set used, which reflects the conditions of a real scenario well, we were able to guarantee subject-independence through strict subject-based cross validation.

A. LIMITATIONS

Even if our results are promising and we can thus answer the research question formulated at the beginning and close the research gap identified, there are limitations. As part of the study, we deliberately decided to use a dataset that does not focus on depression so that we can depict a scenario that is as real as possible. However, this only allowed us to build a binary classifier for the prediction of a mild depression in the study. In addition to the two stages of mild and no depression, more severe states also exist, whose influence on our model might be investigated. Nevertheless, it is difficult to identify a suitable dataset that contains enough patients with severe depression who are not yet undergoing treatment, which would falsify the results due to medication [47]. Another limitation is the validation of our architecture. For the use case of mild depression detection, we have already successfully demonstrated its suitability. Nevertheless, it would be insightful if the same architecture could be tested in other application areas with resting-state EEG data.

B. FUTURE WORK

In future research, we want to tackle these limitations ourselves and implement further research ideas on the subject.

A major point is the dataset, where we want to use mobile (consumer) EEG sensors to collect a realistic dataset that includes both mild and no depression subjects, but also more severe patients, and thus test our approach. We want to investigate the influence of a more severe course of the disease on our model and perhaps develop a more holistic model for predicting all stages that also works with mobile EEG sensing devices.

In a further study, we also want to increase the interpretability of the model by using explainable artificial intelligence methods such as DeepLIFT [73] to understand the decisions of our model and thus possibly generate insights for understanding the disease and other mental health diagnostics in general. Even if the explanation of EEG-based DL models has not yet been studied much, there are promising approaches [74].

And last but not least, in a series of further studies, we want to test our newly developed CNN architecture for other resting-state EEG-based application areas such as working memory or personality prediction in order to increase its validity.

REFERENCES

[1] *Depression and Other Common Mental Disorders: Global Health Estimates*, World Health Org., Geneva, Switzerland, 2017.

[2] J. J. Newson and T. C. Thiagarajan, "EEG frequency bands in psychiatric disorders: A review of resting state studies," *Frontiers Hum. Neurosci.*, vol. 12, p. 521, Jan. 2019.

[3] M. Wolfersdorf, "Depression and suicide," *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, vol. 51, pp. 443–450, Jan. 2008.

[4] G. Li, D. Fife, G. Wang, J. J. Sheehan, R. Bodén, L. Brandt, P. Brenner, J. Reutfors, and A. DiBernardo, "All-cause mortality in patients with treatment-resistant depression: A cohort study in the US population," *Ann. Gen. Psychiatry*, vol. 18, no. 1, pp. 1–8, Dec. 2019.

[5] K. M. Schafer, A. Lieberman, A. C. Sever, and T. Joiner, "Prevalence rates of anxiety, depressive, and eating pathology symptoms between the pre- and peri-COVID-19 eras: A meta-analysis," *J. Affect. Disorders*, vol. 298, pp. 364–372, Feb. 2022.

[6] G. M. Goodwin, "Depression and associated physical diseases and symptoms," *Dialogues Clin. Neurosci.*, vol. 8, no. 2, pp. 259–265, Jun. 2006.

[7] C. M. Gillan and R. Whelan, "What big data can do for treatment in psychiatry," *Current Opinion Behav. Sci.*, vol. 18, pp. 34–42, Dec. 2017.

[8] S. Yasin, S. A. Hussain, S. Aslan, I. Raza, M. Muzammel, and A. Othmani, "EEG based major depressive disorder and bipolar disorder detection using neural networks: A review," *Comput. Methods Programs Biomed.*, vol. 202, Apr. 2021, Art. no. 106007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260721000821

[9] P. Bech, P. Allerup, L. F. Gram, N. Reisby, R. Rosenberg, O. Jacobsen, and A. Nagy, "The Hamilton depression scale: Evaluation of objectivity using logistic models," *Acta Psychiatrica Scandinavica*, vol. 63, no. 3, pp. 290–299, Mar. 1981.

[10] R. M. Bagby, A. G. Ryder, D. R. Schuller, and M. B. Marshall, "The Hamilton depression rating scale: Has the gold standard become a lead weight?" *Amer. J. Psychiatry*, vol. 161, no. 12, pp. 2163–2177, Dec. 2004.

[11] C. Bernard and W. Anita, "Catalog of biases in questionnaires," *Prev. Chronic. Dis.*, vol. 2, no. 1, pp. 1–13, Jan. 2005.

[12] M. Čukić, V. López, and J. Pavón, "Classification of depression through resting-state electroencephalogram as a novel practice in psychiatry: Review," *J. Med. Internet Res.*, vol. 22, no. 11, Nov. 2020, Art. no. e19548.

[13] K. Thulasi, S. Balakrishnan, Y. J. Suan, Y. X. Qing, M. B. Malarvili, R. K. Murugesan, and P. Devandran, "Detecting mild depression from EEG signal in a non-clinical environment using machine learning technique," in *Proc. IEEE 2nd Nat. Biomed. Eng. Conf. (NBEC)*, Sep. 2023, pp. 100–105.

[14] M. Bachmann, L. Päeske, K. Kalev, K. Aarma, A. Lehtmets, P. Ööpik, J. Lass, and H. Hinrikus, "Methods for classifying depression in single channel EEG using linear and nonlinear signal analysis," *Comput. Methods Programs Biomed.*, vol. 155, pp. 11–17, Mar. 2018.

[15] W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik, "A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD)," *Med. Biol. Eng. Comput.*, vol. 56, no. 2, pp. 233–246, Feb. 2018.

[16] S. Mahato and S. Paul, "Classification of depression patients and normal subjects based on electroencephalogram (EEG) signal using alpha power and theta asymmetry," *J. Med. Syst.*, vol. 44, no. 1, pp. 1–8, Jan. 2020.

[17] C.-T. Wu, H.-C. Huang, S. Huang, I.-M. Chen, S.-C. Liao, C.-K. Chen, C. Lin, S.-H. Lee, M.-H. Chen, C.-F. Tsai, C.-H. Weng, L.-W. Ko, T.-P. Jung, and Y.-H. Liu, "Resting-state EEG signal for major depressive disorder detection: A systematic validation on a large and diverse dataset," *Biosensors*, vol. 11, no. 12, pp. 499–524, Dec. 2021.

[18] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, Mar. 2013.

[19] M. Sharma, P. V. Achuth, D. Deb, S. D. Puthankattil, and U. R. Acharya, "An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals," *Cognit. Syst. Res.*, vol. 52, pp. 508–520, Dec. 2018.

[20] D. M. Khan, K. Masroor, M. F. M. Jailani, N. Yahya, M. Z. Yusoff, and S. M. Khan, "Development of wavelet coherence EEG as a biomarker for diagnosis of major depressive disorder," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4315–4325, Mar. 2022.

[21] M. BalaKrishnan, P. Colditz, and B. Boashash, "A multi-channel fusion based newborn seizure detection," *J. Biomed. Sci. Eng.*, vol. 7, no. 8, pp. 533–545, 2014.

[22] S. B. Harvey, S. Øverland, S. L. Hatch, S. Wessely, A. Mykletun, and M. Hotopf, "Exercise and the prevention of depression: Results of the HUNT cohort study," *Amer. J. Psychiatry*, vol. 175, no. 1, pp. 28–36, Jan. 2018.

[23] A. Ksibi, M. Zakariah, L. J. Menzli, O. Saidani, L. Almuqren, and R. A. M. Hanafieh, "Electroencephalography-based depression detection using multiple machine learning techniques," *Diagnostics*, vol. 13, no. 10, p. 1779, May 2023.

[24] H.-P. Volz and P. Laux, "Potential treatment for subthreshold and mild depression: A comparison of St. John's wort extracts and fluoxetine," *Comprehensive Psychiatry*, vol. 41, no. 2, pp. 133–137, Mar. 2000.

[25] L. Duan, H. Duan, Y. Qiao, S. Sha, S. Qi, X. Zhang, J. Huang, X. Huang, and C. Wang, "Machine learning approaches for MDD detection and emotion decoding using EEG signals," *Frontiers Hum. Neurosci.*, vol. 14, p. 284, Sep. 2020.

[26] X. Li, R. La, Y. Wang, J. Niu, S. Zeng, S. Sun, and J. Zhu, "EEG-based mild depression recognition using convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 57, no. 6, pp. 1341–1352, Jun. 2019.

[27] A. R. Keeler, J. T. Siegel, and E. M. Alvaro, "Depression and help seeking among Mexican–americans: The mediating role of familism," *J. Immigrant Minority Health*, vol. 16, no. 6, pp. 1225–1231, Dec. 2014.

[28] L. M. McDermott and K. P. Ebmeier, "A meta-analysis of depression severity and cognitive function," *J. Affect. Disorders*, vol. 119, nos. 1–3, pp. 1–8, Dec. 2009.

[29] X. Li, R. La, Y. Wang, B. Hu, and X. Zhang, "A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography," *Frontiers Neurosci.*, vol. 14, Apr. 2020, Art. no. 477153.

[30] X. Li, B. Hu, S. Sun, and H. Cai, "EEG-based mild depressive detection using feature selection methods and classifiers," *Comput. Methods Programs Biomed.*, vol. 136, pp. 151–161, Nov. 2016.

[31] J. W. Kanter, A. M. Busch, C. E. Weeks, and S. J. Landes, "The nature of clinical depression: Symptoms, syndromes, and behavior analysis," *Behav. Analyst*, vol. 31, no. 1, pp. 1–21, Apr. 2008.

[32] D. R. Strunk, B. J. Pfeifer, and I. D. Ezawa, "Depression," in *Handbook of Cognitive Behavioral Therapy: Applications*, A. Wenzel, Ed. American Psychological Association, 2021, pp. 3–31.

[33] M. Hamilton, "A rating scale for depression," *J. Neurol. Neurosurg.*, vol. 23, no. 1, pp. 56–62, 1960.

[34] I. W. Miller, S. Bishop, W. H. Norman, and H. Maddever, "The modified Hamilton rating scale for depression: Reliability and validity," *Psychiatry Res.*, vol. 14, no. 2, pp. 131–142, Feb. 1985.

[35] S. Debener, F. Minow, R. Emkes, K. Gandras, and M. de Vos, "How about taking a low-cost, small, and wireless EEG for a walk?" *Psychophysiology*, vol. 49, no. 11, pp. 1617–1621, Nov. 2012.

[36] A. Dev, N. Roy, M. K. Islam, C. Biswas, H. U. Ahmed, M. A. Amin, F. Sarker, R. Vaidyanathan, and K. A. Mamun, "Exploration of EEG-based depression biomarkers identification techniques and their applications: A systematic review," *IEEE Access*, vol. 10, pp. 16756–16781, 2022.

[37] S. Sun, J. Li, H. Chen, T. Gong, X. Li, and B. Hu, "A study of resting-state EEG biomarkers for depression recognition," 2020, *arXiv:2002.11039*.

[38] J. J. Allen and S. J. Reznik, "Frontal EEG asymmetry as a promising marker of depression vulnerability: Summary and methodological considerations," *Current Opinion Psychol.*, vol. 4, pp. 93–97, Aug. 2015.

[39] F. S. D. A. Neto and J. L. G. Rosa, "Depression biomarkers using non-invasive EEG: A review," *Neurosci. Biobehavioral Rev.*, vol. 105, pp. 83–93, Oct. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763419303823

[40] S. Aleem, N. U. Huda, R. Amin, S. Khalid, S. S. Alshamrani, and A. Alshehri, "Machine learning algorithms for depression: Diagnosis, insights, and research directions," *Electronics*, vol. 11, no. 7, p. 1111, Mar. 2022.

[41] H. Cai, J. Han, Y. Chen, X. Sha, Z. Wang, B. Hu, J. Yang, L. Feng, Z. Ding, Y. Chen, and J. Gutknecht, ''A pervasive approach to EEG-based depression detection,'' *Complexity*, vol. 2018, no. 1, 2018, Art. no. 5238028.

[42] P. Kaushik, H. Yang, P. P. Roy, and M. van Vugt, ''Comparing resting state and task-based EEG using machine learning to predict vulnerability to depression in a non-clinical population,'' *Sci. Rep.*, vol. 13, no. 1, p. 7467, May 2023.

[43] M. A. Villarroel and E. P. Terlizzi, ''Symptoms of depression among adults: United States, 2019,'' *NCHS Data Brief*, vol. 379, pp. 1–8, Sep. 2020.

[44] S. Mahato, N. Goyal, D. Ram, and S. Paul, ''Detection of depression and scaling of severity using six channel EEG data,'' *J. Med. Syst.*, vol. 44, no. 7, pp. 1–12, Jul. 2020.

[45] M. Li, Y. Liu, Y. Liu, C. Pu, R. Yin, Z. Zeng, L. Deng, and X. Wang, ''Resting-state EEG-based convolutional neural network for the diagnosis of depression and its severity,'' *Frontiers Physiol.*, vol. 13, Oct. 2022, Art. no. 956254.

[46] M. Li, J. Zhang, J. Song, Z. Li, and S. Lu, ''A clinical-oriented non-severe depression diagnosis method based on cognitive behavior of emotional conflict,'' *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 1, pp. 131–141, Feb. 2023.

[47] A. Baskaran, R. Milev, and R. S. McIntyre, ''The neurobiology of the EEG biomarker as a predictor of treatment response in depression,'' *Neuropharmacology*, vol. 63, no. 4, pp. 507–513, Sep. 2012.

[48] M. Ferdjallah and R. E. Barr, ''Adaptive digital notch filter design on the unit circle for the removal of powerline noise from biomedical signals,'' *IEEE Trans. Biomed. Eng.*, vol. 41, no. 6, pp. 529–536, Jun. 1994.

[49] R. Buettner, J. Frick, and T. Rieg, ''High-performance detection of epilepsy in seizure-free EEG recordings: A novel machine learning approach using very specific epileptic EEG sub-bands,'' in *Proc. 40th Int. Conf. Inf. Syst.*, 2019, p. 2536.

[50] P. Welch, ''The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,'' *IEEE Trans. Audio Electroacoustics*, vol. AE-15, no. 2, pp. 70–73, Jun. 1967.

[51] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, ''MEG and EEG data analysis with MNE-Python,'' *Frontiers Neurosci.*, vol. 7, p. 267, Dec. 2013.

[52] H. Baumgartl, S. Bayerlein, and R. Buettner, ''Measuring extraversion using EEG data,'' in *Information Systems and Neuroscience*. Berlin, Germany: Springer, 2020, pp. 259–265.

[53] J. Breitenbach, H. Baumgartl, and R. Buettner, ''Detection of excessive daytime sleepiness in resting-state EEG recordings: A novel machine learning approach using specific EEG sub-bands and channels,'' in *Proc. AMCIS*, 2020, pp. 1–9.

[54] J. Breitenbach, A. Uzun, T. Rieg, H. Baumgartl, and R. Buettner, ''A novel machine learning approach to working memory evaluation using resting-state EEG data,'' *PACIS*, vol. 21, pp. 1–12, Jan. 2021.

[55] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ''SMOTE: Synthetic minority over-sampling technique,'' *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[56] X. Ying, ''An overview of overfitting and its solutions,'' *J. Phys., Conf.*, vol. 1168, Feb. 2019, Art. no. 022022.

[57] D. P. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' 2014, *arXiv:1412.6980*.

[58] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, ''EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces,'' *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[59] L. Breiman, ''Heuristics of instability and stabilization in model selection,'' *Ann. Statist.*, vol. 24, no. 6, pp. 2350–2383, Dec. 1996.

[60] M. Sokolova and G. Lapalme, ''A systematic analysis of performance measures for classification tasks,'' *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

[61] Q. Wei and R. L. Dunbrack, ''The role of balanced training and testing data sets for binary classifiers in bioinformatics,'' *PLoS ONE*, vol. 8, no. 7, Jul. 2013, Art. no. e67863.

[62] D. Chicco, N. Tötsch, and G. Jurman, ''The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation,'' *BioData Mining*, vol. 14, no. 1, pp. 1–22, Feb. 2021.

[63] C. Sharpe, T. Wiest, P. Wang, and C. C. Seepersad, ''A comparative evaluation of supervised machine learning classification techniques for engineering design applications,'' *J. Mech. Design*, vol. 141, no. 12, Dec. 2019, Art. no. 121404.

[64] A. Babayan, M. Erbey, D. Kumral, J. D. Reinelt, A. M. Reiter, J. Röbbig, H. L. Schaare, M. Uhlig, A. Anwander, P.-L. Bazin, and A. Horstmann, ''A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults,'' *Sci. data*, vol. 6, no. 1, pp. 1–21, 2019.

[65] L. L. Onega and I. L. Abraham, ''Factor structure of the Hamilton rating scale for depression in a cohort of community-dwelling elderly,'' *Int. J. Geriatric Psychiatry*, vol. 12, no. 7, pp. 760–764, Jul. 1997.

[66] D.-C. Li, C.-W. Liu, and S. C. Hu, ''A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets,'' *Artif. Intell. Med.*, vol. 52, no. 1, pp. 45–52, May 2011.

[67] X. Jiang, G.-B. Bian, and Z. Tian, ''Removal of artifacts from EEG signals: A review,'' *Sensors*, vol. 19, no. 5, p. 987, Feb. 2019.

[68] F. Sholevar, T. Butryn, L. Bryant, and C. Marchionni, ''The shortage of psychiatrists and other mental health providers: Causes, current state, and potential solutions,'' *Int. J. Academic Med.*, vol. 3, no. 1, pp. 5–9, 2017.

[69] K. M. Esquivel, ''Remote rehabilitation: A solution to overloaded & scarce health care systems,'' *Trends Telemedicine E-Health*, vol. 1, no. 1, pp. 1–19, Aug. 2018.

[70] S. Noachtar and J. Rémi, ''The role of EEG in epilepsy: A critical review,'' *Epilepsy Behav.*, vol. 15, no. 1, pp. 22–33, May 2009.

[71] V. Mihajlovic, B. Grundlehner, R. Vullers, and J. Penders, ''Wearable, wireless EEG solutions in daily life applications: What are we missing?'' *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 6–21, Jan. 2015.

[72] A. Biondi, V. Santoro, P. F. Viana, P. Laiou, D. K. Pal, E. Bruno, and M. P. Richardson, ''Noninvasive mobile EEG as a tool for seizure monitoring and management: A systematic review,'' *Epilepsia*, vol. 63, no. 5, pp. 1041–1063, May 2022.

[73] A. Shrikumar, P. Greenside, and A. Kundaje, ''Learning important features through propagating activation differences,'' in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 3145–3153.

[74] A. Sujatha Ravindran and J. Contreras-Vidal, ''An empirical comparison of deep learning explainability approaches for EEG using simulated ground truth,'' *Sci. Rep.*, vol. 13, no. 1, p. 17709, Oct. 2023.

PASCAL PENAVA was born in May1999. He received the B.S. degree in information systems from Friedrich Alexander University, Erlangen–Nuremberg, Germany, in 2021, and the M.S. degree in digitalization and entrepreneurship from Bayreuth University, Bayreuth, Germany, in 2023. Currently, he is pursuing the Ph.D. degree with the Chair of Hybrid Intelligence, Helmut-Schmidt-University/University of the Federal Armes Forces Hamburg, Hamburg. His research interests include the development of EEG based BCIs and machine learning based analyses of time-series data.

RICARDO BUETTNER (Senior Member, IEEE) received the Dipl.-Inf. degree in computer science and the Dipl.-Wirtsch.-Ing. degree in industrial engineering and management from the Technical University of Ilmenau, Germany, the Dipl.-Kfm. degree in business administration from the University of Hagen, Germany, the Ph.D. degree in information systems from the University of Hohenheim, Germany, and the Habilitation (venia legendi) degree in information systems from the University of Trier, Germany. He is currently a Chaired Professor of Hybrid Intelligence, Helmut-Schmidt-University/University of the Federal Armes Forces Hamburg, Germany. He has published more than 150 peer-reviewed articles, including articles in *Electronic Markets*, *AIS Transactions on Human–Computer Interaction*, *Personality and Individual Differences*, *European Journal of Psychological Assessment*, *PLOS One*, and IEEE Access. He has received 18 international best paper, the best reviewer, and service awards and award nominations, including the Best Paper Awards by *AIS Transactions on Human–Computer Interaction*, *Electronic Markets* journal, and HICSS, for his work.