SN Computer Science manuscript No.
(will be inserted by the editor)

Accu-Help: A Machine Learning based Smart Healthcare Framework for Accurate Detection of Obsessive Compulsive Disorder

Kabita Patel · Ajaya K. Tripathy * · Laxmi N. Padhy · Sujita K. Kar · Susanta K. Padhy · Saraju P. Mohanty

the date of receipt and acceptance should be inserted later

Received: ** *** 2023 / Accepted: ** *** 2023

Abstract Smart Healthcare becomes one of the popular research areas in recent years. This research proposes to expand the state-of-art of smart healthcare by incorporating solutions for Obsessive Compulsive Disorder (OCD). Classification of OCD by analyzing oxidative stress biomarkers (OSBs) through a machine learning mechanism is a significant development in the study of OCD. However, this procedure requires the collection of OCD class labels from hospitals, collection of corresponding OSBs from biochemical laboratories, integrated and labeled dataset creation, use of suitable machine learning algorithm for designing OCD prediction model, and making these prediction models available for different biochemical laboratories for OCD prediction for unlabeled OSBs. Further, from time to time, with significant growth in the volume of the dataset with labeled samples, redesigning the prediction model is required for further use. The entire process demands distributed data collection, data integration, coordination between the hospital and the biochemical laboratory in real-time, dynamic machine learning model design for OCD prediction, and making the machine learning model available for the biochemical laboratories. Considering these requirements, Accu-Help a fully automated, smart, and accurate OCD detection conceptual model is proposed to help the biochemical laboratories for efficient detection of OCD from OSBs. OSBs are classified into three classes: Healthy Individual (HI), OCD Affected Individual (OAI), and Genetically Affected Individual (GAI). The main component of this proposed framework is the machine learning-based OCD class prediction model design. Accu-Help uses a neural network-based approach with an OCD class prediction accuracy of $86 \pm 2\%$.

Keywords Healthcare Cyber-Physical System (H-CPS) · Smart Healthcare, Internet-of-Medical-Things (IoMT) · Machine Learning · Artificial Neural Network (ANN) · Obsessive Compulsive Disorder (OCD)

Kabita Patel
Dept. of Computer Sci. , Gangadhar Meher University, India
E-mail: kabitapatelgmu@gmail.com

Ajaya K. Tripathy (Corresponding Author)
Dept. of Computer Sci. , Gangadhar Meher University, India
E-mail: ajayatripathy1@gmail.com

Laxmi N. Padhy
Dept. of CSE, Konark Institute of Science and Technology, India.
E-mail: lnpadhy2020@gmail.com

Sujita K. Kar
Dept. of Psychiatry, King George's Medical University, India.
E-mail: drsujita@gmail.com

Susanta K. Padhy
Department of Psychiatry, All India Institute of Medical Sciences, Bhubaneswar, India.
E-mail: psych_susanta@aiimsbhubaneswar.edu.in

Saraju P. Mohanty
Dept. of Computer Sci. and Eng., University of North Texas, USA
E-mail: saraju.mohanty@unt.edu

Conflict of Interest:

Author Ajaya K. Tripathy has received research grants from Odisha Higher Education Programme for Excellence and Equity (OHEPEE) World Bank.

1 Introduction

In the era of the rapid advancement of machine learning, the internet of things, and cyber-physical system

technology, there is a larger possibility of improving the excellence of healthcare technologies. Recently, numerous researchers have demonstrated their interest in designing and developing smart healthcare technologies to resolve different issues in the healthcare system. For example, to automate seizure detection from EEG the authors in [1] have proposed an IoT-based System using Machine Learning. A H-CPS is proposed in [2] to detect Blood Alcohol Concentration using machine learning. A smart healthcare system is proposed in [3] to detect and monitor diseases to provide real-time support to patients.

Barely attention is given in the literature to designing H-CPS for mental illnesses like anxiety, obsession, compulsion, or obsessive-compulsive disorder.

This research focuses on obsessive-compulsive disorder (OCD). OCD is one of the classes of anxiety illness [4]. Persons with OCD have a mental condition of obsession and compulsion. Obsessions are unpleasant and unwanted feelings that automatically come into mind. Obsession makes an individual very uncomfortable, anxious, and fearful. For the sake of correction, individuals with obsession carry out repeated activities called compulsion. For example, washing hands frequently to overcome the thought of contamination. Here, contamination feeling is the obsession, and washing hands repeatedly is the compulsion.

One of the effective treatments for OCD is Cognitive-behavioral therapy (CBT). CBT is a general kind of conversion therapy. This therapy helps individuals to learn the way to recognize the patterns and alter negative feelings or emotions. However, OCD is detected from behavioral symptoms analysis, but the OCD behavior is observable at a letter stage. Earlier detection can help the patient for quick recovery using CBT treatment. However, in general, OCD-affected persons have less trust in the detection process through symptom analysis in the preliminary stage. As long as this OCD problem is not harming their day-to-day activity, they are not accepting the disease and abstaining from treatment. Therefore, the treatment starts at a later stage and the recovery period becomes longer. In such a situation, OCD detection using an intelligent machine may create trust in the diagnosis process of OCD at an early stage. As a result, CBT can be adapted in the early stage for better results.

For automated and accurate detection of OCD, a Healthcare Cyber-Physical System (H-CPS) is proposed called Accu-Help. H-CPS is an information and communication technology-based infrastructure in which the healthcare process is supervised and controlled using smart systems. Accu-Help is designed as an H-CPS which uses a machine learning-based smart healthcare framework for the accurate detection of OCD. The accountability of Accu-Help includes the collection of OCD class labels from hospitals, collection of corresponding OSBs from biochemical laboratories, integrated database creation, use of suitable machine learning algorithm for designing OCD prediction model, and making these prediction models available for different biochemical laboratories for OCD prediction for unlabeled OSBs. With the growth of the size of the OCD database containing labeled samples, redesigning the OCD prediction model is required to enhance the robustness of Accu-Help. The entire process requires distributed data collection, data integration, coordination between the hospital and biochemical laboratory, dynamic machine learning-based OCD prediction mode design, and making the prediction model available over the cloud for easy access and integration.

The core part of Accu-Help is a machine learning-based model which can provide the intelligence required for the prediction of OCD. Accu-Help can collect OSBs and OCD class labels from different h ospitals a nd different biochemical laboratories located in different geographical locations. OSBs along with the class labels are used for ML model design for OCD classification and made available online. OSBs estimated from new individuals' blood samples in a biochemical laboratory can be given to Accu-Help for OCD class detection.

A novel hyperparameter-optimized neural network for OCD identification o r c lassification us ing oxidative stress biomarkers is used in the Accu-Help. The usefulness of the proposed approach is compared with the usefulness of some of the popular classification approaches. Experimental results reveal that the suggested mechanism is better contrary to others.

The rest of the article is organized as follows. The state of the art of the problem at hand is summarized in Section 2. The novelty of the article is presented in Section 3. In Section 4 a cyber-physical system is proposed to handle the whole system of OCD detection. Various popular machine learning methods are used for OCD detection in Section 5. Section 6 presents the proposed hyperparameter-optimized neural network for OCD classification through oxidative stress biomarkers. The dataset description, experimental result analysis, and comparative studies are performed in Section 7. Section 8 gives the future work direction and concludes the article.

2 Related Prior Works

The related research approaches are categorized into two parts. In Section 2.1, few recent smart healthcare systems are presented. Further, in Section 2.2, OCD

detection-related approaches are presented and analyzed.

2.1 Smart Healthcare Systems

In recent years, there has been an increasing interest among several researchers are seen to attempt of developing a smart healthcare system to automatically diagnose, and manage different expects of diseases [18, 19]. However, the healthcare system proposed in this article mainly focuses on OCD and the diagnosis of OCD from oxidative stress biomarkers. There are several smart healthcare systems proposed in the literature. The research work in [20] proposes a model to automate the process of automatic monitoring of patients and biomedical devices. To automate seizure detection from EEG the authors in [1] have proposed an IoT-based System using Machine Learning. A smart healthcare system is proposed in [2] to detect Blood Alcohol Concentration using machine learning. In [21] the authors have proposed a healthcare system to handle the mobility of individuals during a pandemic. Even though several smart healthcare systems are proposed in the literature, as per the knowledge of the authors' none of the approaches has proposed a smart healthcare system to automate the process of OCD detection.

2.2 Related Prior Research

In recent years several researchers have analyzed several aspects of OCD such as prediction, monitoring treatment effectiveness, and severity prediction. Artificial intelligence approaches were also well utilized to resolve different problems related to OCD. Several approaches focus on the prediction of OCD patients by studying neuropsychological biomarkers, genetic biomarkers, MRI biomarkers [5], fMRI biomarkers, DTI and EEG biomarkers [6, 7], and EEG biomarkers [8]. Artificial intelligence approaches have been used in the detection of OCD treatment effectiveness in [10], OCD severity forecast [11] and OCD severity reduction forecast model has been designed in [22, 23]. In general, the EEG analysis is performed by power and source analysis. EEG channel one and four analyses have been performed for OCD detection in [6]. An intracortical EEG signal is analyzed in [9] and observed hyper-activation for OCD individuals. MRI and fMRI have been analyzed in many studies to diagnose OCD [12, 24–27]. However, the collection of such biomarkers involves high-end machines which may not be available in most places. The OCD research literature on neuroimaging biomarkers and their limitations are summarized in Table 1. In absence of biological markers such as EEG, MRI, fMRI, and DTI, an alternative mechanism should be adopted.

The study in [17] suggests the act of oxidative stress biomarkers in OCD. Oxidative stress biomarkers (OSBs) can be measured from an individual's blood samples. This study observed different patterns of OSBs in all three OCD classes (HI, GAI, and OAI. The OSBs considered for the study are superoxide dismutase (SD), Glutathione Peroxidase (GP), Catalase (CAT), Malondialdehyde (MAL), and serum cortisol (SC). Several studies are found in the literature on the analysis of OSBs to understand OCD. In some studies [14, 17] COR is found to be normal whereas in some other studies [13, 28] it is found to be high in the case of OCD individuals. As per [14, 15, 17, 29], CAT, GPX, and SOD are lower and MDA is higher in OCD whereas levels of CAT, GPX, and SOD were higher in OCD individuals in other studies [14, 16]. In [25] the authors have made a careful study of OSBs and it is observed that these markers act as a major role in OCD individuals [25]. In [25] it is also observed that the population mean and the standard deviation is significantly different in the case of OCD individuals. However, it is not possible to predict OCD through the statistical analysis of each marker. The OCD research literature on OSBs and their limitations are summarized in Table 2. This research aims to design an intelligent predictive method to predict the existence of OCD in an individual through oxidative stress biomarkers. Thus, it is essential not only to develop a mechanism to detect OCD without EEG, MRI, fMRI, and DTI but also to classify all three classes.

3 Novel Contributions of the Article

Segregating the HIs from the OAI is considered the OCD detection or classification process. In regular practice, the mechanism of OCD detection is carried out by symptom observation. Generally, the symptoms are visible at the latter stage of the disease. As the symptoms are not visible in the initial stage of the disease, early recognition is not possible. Further, in the case of an early or moderate stage of OCD, the symptoms are mild and in general, the patient doesn't trust the detection through symptoms observation. As a result, they obstruct themselves from taking treatment. However, laboratory detection of OCD may build extra trust in the mind of OCD patients even at the early stage and as a result, they may accept the treatments in the early stages. Accepting treatment at the early stage may result in quick recovery.

In the literature, several works have been carried out to detect OCD by applying machine learning to

Table 1 OCD research literature on neuroimaging biomarkers and their limitations

Research	**Analysis and results**	**Limitations**
OCD severity detection [5]	Data Analyzed: MRI, DTI, and neuropsychological data. The technique used: Applied machine learning for OCD severity detection Result: Detection Ability of 90% in the training set and 70% in the testing set.	Collection of such biomarkers requires high-end machines near the patient. Without biological markers such as EEG, MRI, fMRI, and DTI, an alternative mechanism should be adopted.
Classification of OCD [6]	Data Analyzed: EEG data and hemispheric dependency data. The technique used: Support vector machine (SVM) classifiers. Result: Achieved an OCD classification accuracy of $85 \pm 5.2\%$.	This approach does not able to classify GAI. Further, in the absence of biological markers such as EEG, MRI, fMRI, and DTI, an alternative mechanism should be adapted.
Classify trichotillomania and OCD [7]	Data Analyzed: EEG biomarkers. The technique used: SVM with ant colony optimization. Result: Achieved a classification accuracy of 81.04%.	This approach is not able to address the OCD classification problem and is not able to identify the GAI group.
EEG source analysis in OCD [8, 9]	Data Analyzed: EEG biomarkers. The technique used: Compared resting state using standardized low-resolution electromagnetic tomography. Result: Observed that there is a medial frontal hyperactivation in OCD.	This approach is not able to classify GAI. Further, an alternative mechanism should be adapted in the absence of biological markers such as EEG, MRI, fMRI, and DTI.
Prediction of OCD treatment response [10]	Data Analyzed: Symptoms dimension, neuropsychologic act, and epidemiologic parameters. The technique used: Multilayer perceptrons. Result: 93.3% of correct classification of cases achieved.	This approach is not able to address the OCD classification problem. Does not able to identify GAI group.
Predicting OCD severity [11]	Data Analyzed: MRI data. The technique used: support vector regression. Result: Concluded that Support Vector Regression can predict OCD symptom severity.	This approach is not able to address the OCD classification problem. Does not able to identify GAI group.
fMRI pattern recognition in OCD [12]	Data Analyzed: fMRI data. The technique used: Multivariate pattern classification techniques. Result: Neurobiological markers provide reliable diagnostic information about OCD.	This approach is not able to classify GAI. Further, an alternative mechanism should be adapted in the absence of biological markers such as EEG, MRI, fMRI, and DTI.

Table 2 OCD research literature on oxidative stress biomarkers and their limitations

Research	**Findings**	**Limitations**
Urinary free cortisol (UFC) Cortisol level analysis in OCD [13]	UFC of both the groups was compared and the OAI group had significantly higher UFC levels than the HI group.	This study analyzed the cortisol level but had not come up with a threshold as OCD detector.
MAL, SD, GP, and CAT levels in patients with OCD [14]	Higher MAL, SD, GP, and CAT activity but the differences were not big. However, it is observed that OCD is linked with free radicals.	This study analyzed oxidative stress biomarkers but had not come up with a threshold as an OCD detector.
Analysis of free radical metabolism and antioxidants in OCD [15]	A higher level of MAL was observed in the OCD group.	This study analyzed MAL but had not come up with a threshold as an OCD detector.
Analysis of oxidative stress of OCD patients [16]	Oxidative stress biomarkers imbalance was observed in the OCD group.	This study analyzed oxidative/ antioxidative status but had not come up with a threshold as an OCD detector.
Oxidative stress biomarkers analysis in three groups (HI, GAI, and OAI) [17]	Levels of CAT, SD and GP in all three groups are significantly different.	This study analyzed oxidative stress in all three groups but had not come up with a mechanism to segregate these three groups.
Accu-Help: A Machine Learning-based smart healthcare framework for accurate detection of OCD class (HI, GAI, or OAI).	Hyperparameter optimized Neural networks have achieved a prediction accuracy of $86 \pm 1\%$.	This approach is successful in applying Neural Networks on OSBs to predict OCD class. However, the training process is computationally costly.

biomarkers such as MRI, fMRI, EEG, and DTI. However, the collection of such biomarkers involves high-end machines which may not be available everywhere. The study in [30] suggests a link between oxidative stress biomarkers in OCD and these biomarkers can be measured from blood samples. Further, the recent study performed in [17] suggests the existence of a significantly distinguished pattern in the biomarkers of the first-degree relatives of OCD-affected individuals. By identifying the individuals with genetic linkage with OCD, preventive suggestions may be recommended to such individuals to stay away from OCD. Therefore identifying the genetically affected individuals (GAI) with OCD has greater importance in medical science [17]. Therefore, this study aims at the identification or segregation of the three groups (HI, GAI, and OAI). In this article, we define the OCD classification problem as, classifying the given OSBs of an individual into one of the three classes HI, GAI, and OAI.

3.1 Research Questions

This research aims to design a Healthcare Cyber-Physical System (Accu-Help) to automate the OCD classification process to strengthen the concept of smart healthcare. The issues that Accu-Help primarily focuses on are:

- In the literature, several approaches are found for OCD detection using artificial intelligence by analyzing neuroimaging biomarkers. A collection of such biomarkers involve high-end equipment and the individual needs to be physically present near the equipment for data collection. However, such equipment may not be available everywhere and the collection of such markers becomes a challenge. As a result, these OCD detections process may not be widely available.
- It is observed by many researchers that oxidative stress has a significant role in OCD. The majority of these studies focused on comparing the population mean, mode, and standard deviation of OSBs among the HI group and the OAI group. However, very less attention is given to detecting OCD through oxidative stress biomarkers.
- Several studies observed the genetic linkage to OCD. Identifying individuals with a genetic link to OCD has a greater importance in medical science. However, in the literature, less importance is given to detect GAI individuals.
- From the above discussion, it is observed that classifying an individual into one of the three classes (HI, GAI, or OAI) has greater importance. It is also observed that artificial intelligence can play a major role in addressing this issue and OSBs can be used as a useful biomarker in solving this problem. In the literature, less importance is given to designing a machine learning prediction model to detect the OCD class (HI, GAI, or OAI) from given OSBs.
- Designing an H-CPS to automate the OCD detection process has greater importance. This H-CPS should be essentially designed to support distributed system to collect labeled OSBs from hospitals and biochemical laboratories, design a machine learning prediction model, and make it available for future use by biochemical laboratories for OCD detection by just giving the OSBs of an individual as input.

3.2 Proposed Solution

As a solution to the problems highlighted in Section 3.1, Accu-Help an H-CPS is proposed. OSBs such as SD, GP, CAT, MAL, and SC are estimated from the blood samples of an individual. These biomarkers are passed to a machine learning prediction model designed by the Accu-Help environment to identify the class (one among HI, GAI, and OAI) by analyzing OSBs. Accu-Help collects labeled OSBs from hospitals and biochemical laboratories. In this study, OCD detection signifies the identification of the class (one of the three classes: HI, GAI, and OAI) from someone's oxidative stress biomarkers. The core part of the Accu-Help is the artificial intelligence part designed based on an artificial neural network for OCD prediction.

3.3 Research Objective

The idea behind the proposed H-CPS is conceptualized by taking into account the process of OCD detection, the ease of involvement of human/machinery components of this process, easy coordination and cooperation among them, and its ancillary impact on the community. The main objectives that are targeted through Accu-Help are:

OCD Individuals Health In the era of artificial intelligence, the detection of OCD through machine learning can put extra faith in the mind of OCD patients. As a result, the patients may start their treatment at an early stage and recover with reduced treatment duration.

Genetically Affected OCD individuals Health This H-CPS aims to segregate HI, GAI, and OAI. GAI individuals can be benefited by taking preventive advice from experts to stay away from OCD.

Early Detection of OCD Accu-Help also aims to predict OCD even in the early stage when symptoms are not so significant. As a result, early treatment can be recommended.

OCD Detection at Biochemical Laboratories As Accu-Help aims to detect from OSBs, the biochemical laboratories can take help of Accu-Help to detect OCD from OSBs without involvement of a doctor.

Technological Advancement In general, behavioral analysis is the popular method of OCD detection. However, this detection happens in the latter stage of the disease, and OCD patient has less faith in this detection process. As a result, they only accept treatment only at the latter stage when the disease affects their day-to-day living. The idea of using machine learning methods through an H-CPS system to analyze OSBs for the detection of HI, GAI, or OAI can be a significant development in OCD studies.

3.4 Practical Use of Proposed Approach

In normal practice, OCD detection is carried out through symptom analysis. However, mild symptoms can be observed in suspected individuals or genetically linked individuals. At the initial stage, individuals have less faith in OCD detection. Further, they don't have any mechanism to make confirmatory tests. As a result, they are restricting them to take preventive treatment. At the latter stage, when the symptoms become significant and OCD negatively affects their day-to-day life they approach the doctor for treatment. As a result, in some cases the OCD becomes chronic and the treatment becomes lengthy. To overcome such a situation, laboratory detection of OCD can be adopted to create extra faith in the early detection process. The proposed approach aims to mitigate this problem. In the proposed approach, blood samples are collected remotely and Oxidative Stress Biomarkers are estimated in a laboratory. Further, the Oxidative Stress Biomarkers are sent to the machine learning model to identify the class label of the sample. The outcome of the machine learning model can be one of the three class labels: 1) Healthy Individual (HI), 2) Individual at the initial stage of OCD or Genetically Affected (GAI), and 3) Individual affected with OCD and is at advance stage (OAI). In case it is detected with GAI, a preventive treatment can be suggested.

4 Accu-Help: A Cyber-Physical System for Accurate Detection of OCD

In recent years many cyber-physical systems are proposed to achieve smart health care [2], and many more. This section proposes a conceptual cyber-physical system for organizing and managing the OCD detection process. The main component of this system is a machine learning model which can solve the OCD classification/detection problem. The efficiency of the machine learning model is mainly dependent on the labeled samples collected as training data and the learning approach. Upon completion of training and validation of the machine learning model, the model should be accessible remotely to make the model widely available and reachable to different biochemical laboratories situated in different geographical locations. Considering all these aspects in mind, the proposed health care cyber-physical system is conceptually divided into three components. The three components are 1) Labeled data collection (LDC), 2) Machine learning model design (MLMD), and 3) Machine learning model remote access (MLMRA).

4.1 Labeled data collection (LDC)

At the time of manual detection of OCD by a psychiatric doctor blood samples are collected from OAIs and their GAIs. The blood samples have to be sent to a biochemical laboratory for oxidative stress biomarkers (such as SD, GP, CAT, MAL, and SC) estimation. The estimated biomarkers along with their class labels have to upload to the central data cloud data server.

4.2 Machine learning model design (MLMD)

Periodically, using the updated dataset machine learning model training and validation are performed in a cloud server and produce batter prediction models from time to time. The improved version of the machine learning prediction model is to be made online available for OCD detection and is called OCD prediction model.

4.3 Machine learning model remote access (MLMRA)

To diagnose OCD in an individual, a blood sample of the individual is needed to be sent to a biochemical laboratory for oxidative stress biomarkers estimation. The estimated biomarkers are to be given to the OCD prediction model which is available online. The OCD

prediction model in the cloud classifies the given samples into one of the three classes HI, GAI, or OAI.

The conceptual design of the cyber-physical system is presented in Fig. 1. The primary objective is the construction of the OCD prediction model which is described in detail in the following sections.

5 Classical Machine Learning Models for OCD detection

For the problem at hand, classification algorithms are suitable to solve the problem. Classification algorithms can be used to predict the class labels of unknown samples. The conceptual view of the OCD class prediction process is presented in Fig. 2.

Among several important classifications approaches k-nearest neighbor, logistic regression, linear discriminant analysis, and neural networks are some of the popular ones. A brief description of these approaches and the OCD detection effectiveness is described in this section as follows.

5.1 Logistic Regression and OCD Detection

Given the OCD dataset containing N number of oxidative stress biomarkers set along with the OCD class labels $\{OSBS_i, CL_i\}_{i=1}^{N}$. Where $CL_i \in \{HI, GAI, OAI\}$ and $OSBS_i =< SD_i, GP_i, CAT_i, MAL_i, andSC_i >$ represents the oxidative stress biomarker set of the i^{th} sample containing five biomarkers called SD, GP, CAT, MAL, and SC. The OSBs are real valued parameters and $OSBS_i \in \Re^5$. The objective is to learn from the dataset and design a prediction model. For a given new $OSBS_i$ determine the class label.

Procedure: Given the training dataset $\{S_i, y_i\}_{i=1}^{N}$, Data point $S_i \in \Re^p$ where, p is the number of predictors, class label $y_i \in \{1, 2, ..., M)$. The training dataset contain N number of samples and the number of class levels are M.

`Objective:` For a given new $s \in \Re^p$ determine the probability of $y \in \{1, 2, ..., M\}$ such that $s \in class\ y$.

`Assumption:` The predictors are drawn from a probability distribution having

$$Pr(y = 1|s) = \frac{e^{\beta^\tau s}}{1 + e^{\beta^\tau s}} = p(s; \beta)....(say), \tag{1}$$

β^τ is expressed as the following:
$\beta^\tau = (\beta_0, \beta_1, \beta_2, ..., \beta_p)$ and $s^T = (1, s_1, s_2, ..., s_p)$.

$$\beta^\tau s = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + ... + \beta_p s_p \tag{2}$$

For a two class classification problem,

$$Pr(y = 0|s) = 1 - p(s; \beta) \tag{3}$$

$\therefore$ To know the probability that a given s is from a class 0 or 1 can be calculated if β is known. β can be calculated using the maximum likelihood method:

$$Pr(y|s) = \prod_{k=1}^{N} Pr(y_k|s_k) \tag{4}$$

For a two class dataset:

$$Pr(y|s) = \prod_{k=1}^{N} p(s_k; \beta)^{y_k} (1 - p(s_k; \beta))^{1-y_k} \tag{5}$$

where, $p(s_k; \beta) = \frac{e^{\beta^\tau s_k}}{1+e^{\beta^\tau s_k}} = \frac{1}{1+e^{-\beta^\tau s_k}} = g(\beta^\tau x)...(say)$

$$Let,\ \ L(\beta) = Pr(y|s) = \prod_{k=1}^{N} p(s_k; \beta)^{y_k} (1 - p(s_k, \beta))^{1-y_k} \tag{6}$$

β can be determined by maximizing the likelihood of the occurrences of all the events in the dataset. $\therefore$ $\beta^\tau = (\beta_0, \beta_1, \beta_2, ..., \beta_p)$ are the values for which $L(\beta)$ is maximum. But, maximizing $L(\beta)$ is same as maximizing $l(\beta) = log(L(\beta))$. To find the β value for which $l(\beta)$ is maximum make $\nabla l(\beta) = 0$.

$\Rightarrow \frac{\partial l}{\partial \beta_j} = 0$, for $j = 0, 1, ..., p$

$\Rightarrow \sum_{i=1}^{N} (y_i - p(x_i, \beta)) x_{ij} = 0$

To get β, one has to solve $\sum s_k(y_k - p(s_k, \beta)) = 0$ and the solution is described in Algorithm 1. For a detailed description of Logistic Regression analysis one can go through [31].

Algorithm 1: β Estimation from the given OCD dataset.

1 **Input:** OCD dataset.
1. Make initial prediction of β, lets call it β^0 and $k = 0$
2. $\beta^{k+1} = \beta^k + \alpha_k \nabla l(\beta^k)$, where α_k is a small value called learning rate.
3. while ($\left\| \beta^{k+1} - \beta^k \right\| < \epsilon$)
4. $\quad \beta^k = \beta^{k+1}$
5. $\quad \beta^{k+1} = \beta^k + \alpha_k \nabla l(\beta^k)$ // Estimate $\nabla l(\beta^k)$ with the help of OCD dataset.
6. Return β^{k+1}

This algorithm is simulated using $R-Programming$ and experimented with OCD dataset containing OSBs and class labels. The experimental outcomes revels that this algorithm achieves an Overall OCD classification Accuracy of 0.777789, Precision of 0.768943, Recall of 0.771247, and F1-Score of 0.770093.

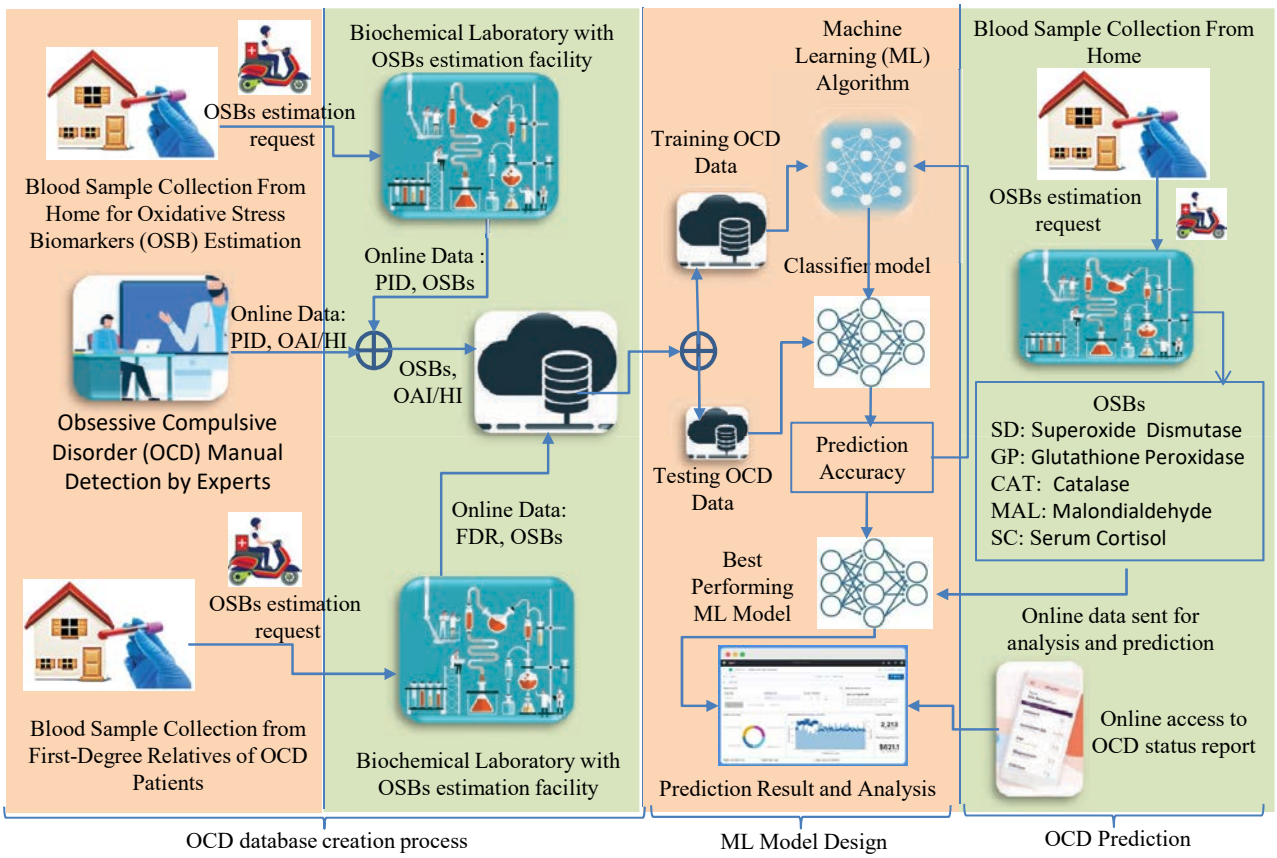

Fig. 1 Conceptual Healthcare Cyber-Physical System (H-CPS) for Obsessive Compulsive Disorder (OCD) Detection

5.2 Linear discriminant analysis and OCD Detection

Given the OCD dataset containing N number of oxidative stress biomarkers set along with the OCD class labels $\{OSBS_i, CL_i\}_{i=1}^{N}$. Where $CL_i \in \{HI, GAI, OAI\}$ and $OSBS_i =< SD_i, GP_i, CAT_i, MAL_i, and SC_i >$ represents the oxidative stress biomarker set of the i^{th} sample containing five biomarkers called SD, GP, CAT, MAL, and SC. The OSBs are real valued parameters and $OSBS_i \in \Re^5$. The objective is to learn from the dataset and design a prediction model. For a given new $OSBS_i$ determine the class label.

Procedure: Let the training dataset $D = \{S_i,\ y_i\}_{i=1}^{N}$, a data point $S_i \in \Re^p$ where p is the number of predictors, and class label $y_i \in \{1,\ 2, ...,\ M\}$.

The training dataset contains N number of samples and the number of class levels is M.

`Objective:` For a given new sample $s \in \Re^p$ determine the the class of s.

`Classification method:` Assign the class label to x for which class probability of occurrence is maximum. i.e., Class label of s is $l = ArgMax_{1 \le k \le M}\ p(y = k|S = s)$.

$$p(y = k|S = s) = \frac{p(S = s|y = k)p(y = k)}{p(S = s)} \tag{7}$$

This can be denoted as $p_k(s) = \frac{f_k(s)\prod_k}{p(s)}$, where $f_k(s)$ is the probability of $S = s$ given y, $\prod_k$ is the probability of $y = k$, and $p(s)$ is the probability of occurrences of s. It is assumed that the probability of $S = s$ for a given $y = k$ is normally distributed. i.e., $f_k(s) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(s-\mu_k)^2}{2\sigma_k^2}}$, here it is assumed that $s \in \Re$ where σ_k is the standard deviation of class k. Again it is assumed that the standard deviation in each class are equal i.e., $\sigma_1 = \sigma_2 = ... = \sigma_M = \sigma$.

$$\therefore f_k(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(s-\mu_k)^2}{2\sigma^2}} \tag{8}$$

But it is known that
$ArgMax_{1 \le k \le M}\ p_k(s) = ArgMax_{1 \le k \le M}\ log\,(p_k(s))$.
$\therefore$ Let us use $log\,(p_k(s))$ to find out for which value of k, $p_k(s)$ is maximum.

$$log(p_k(s)) = log\left[\frac{\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(s-\mu_k)^2}{2\sigma^2}} \prod_k}{p(s)}\right]$$

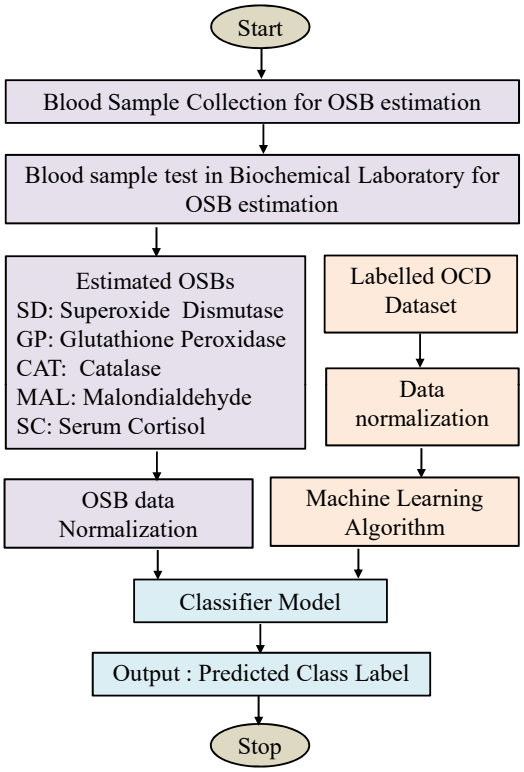

Fig. 2 Proposed OCD prediction model through oxidative stress biomarkers (OSBs) analysis.

$= log\frac{1}{\sqrt{2\pi}\sigma_k} - \left[\frac{s^2}{2\sigma^2} + \frac{\mu_k^2}{2\sigma^2} - \frac{2s\mu_k}{2\sigma^2}\right] + log\prod_k - log\, p(s)$.
$\Rightarrow Arg\ Max_{1\le k\le M}\ (p_k(s)) = Arg\ Max_{1\le k\le M}\ (log\prod_k - \frac{\mu_k^2}{2\sigma^2} + \frac{s\mu_k}{\sigma^2})$ as other terms are independent of k.
Let us denote it as

$Arg\ Max_{1\le k\le M}\ (p_k(s)) = Arg\ Max_{1\le k\le M}\ (\delta_k)$

$\because \prod_k$, μ_k, and σ are unknown, we can use some indirect way to estimate it. Let us define these estimated values as $\hat{\prod_k} = \frac{Numberofelementsink^{th}claaa}{Totalnumberofelements} = \frac{n_k}{N}$, $\hat{\mu_k} = \frac{1}{n_k}\sum_{i:y_i=k} x_i$ and $\hat{\sigma} = \sum_{k=1}^{M}\left(\frac{1}{N-M}\sum_{i:y_i=k}(s_i - \hat{\mu_k})\right)$.

Compute δ_k for all k and find the k for which δ is maximum. That k value is the class label generated by linear discriminant analysis for s.

For the case where $p > 1$ i.e., multiple predictors are there then the shape of linear discriminant analysis is as follows.

Suppose s_1 and s_2 are two random variables independent and normally distributed then the joint probability density function can be written as:

$f(s) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-(s_1-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-(s_2-\mu_2)^2}{2\sigma_2^2}}$
$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(s_1-\mu_1)^2}{\sigma_1^2} + \frac{(s_2-\mu_2)^2}{\sigma_2^2}\right]}$

In general for a p random, independent and normally distributed variables/ predictors:

$$f(s) = \frac{1}{2\pi^{\frac{p}{2}}\left|\sum\right|^{\frac{p}{2}}} e^{-\frac{1}{2}(s-\mu)^\tau \sum^{-1}(x-\mu)} = p_k(s) \quad (9)$$

For $p = 2$: $s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, and
$\sum^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix}$
But we know that,

$$\underset{1\le k\le M}{ArgMax}\ p_k(s) = \underset{1\le k\le M}{ArgMax}\ log\ (p_k(s)) \quad (10)$$

and

$$\underset{1\le k\le M}{ArgMax}\ log\ (p_k(s)) = \underset{1\le k\le M}{ArgMax}\ \delta_k \quad (11)$$

$$where\ \delta_k = s^\tau \sum^{-1} \mu_k - \frac{1}{2}\mu_k^\tau \sum^{-1} \mu_k + log(\prod_k) \quad (12)$$

In equation (12), $s^\tau = (s_1, s_2, ..., s_p)$,
$\mu_k^\tau = (\mu_1, \mu_2, ..., \mu_p)$,
$\prod_k = \frac{Number\ of\ elements\ in\ k^{th}\ class}{Total\ number\ of\ elements}$,
and $\sum = \begin{pmatrix} \sigma_1^2 & 0 & ... & 0 \\ 0 & \sigma_2^2 & ... & 0 \\ . & . & & . \\ . & & ... & . \\ . & . & & . \\ 0 & . & ... & \sigma_p^2 \end{pmatrix}$.

Classification label for s is $l = \underset{1\le k\le M}{Arg\ Max}\ \delta_k$, i.e., compute δ_k for all k for which δ is maximum. That k value is the class label generated by linear discriminant analysis for s from multi dimensional feature space. For a detailed description of Linear discriminant analysis one can go through [32].

This algorithm is simulated using $R-Programming$ and experimented with OCD dataset containing OSBs and class labels. The experimental outcomes revels that this algorithm achieves an Overall OCD classification Accuracy of 0.821431, Precision of 0.833333, Recall of 0.828347, and F1-Score of 0.830827.

5.3 K Nearest neighbor and OCD Detection

Given the OCD dataset containing N number of oxidative stress biomarkers set along with the OCD class labels $\{OSBS_i, CL_i\}_{i=1}^N$. Where $CL_i \in \{HI, GAI, OAI\}$ and $OSBS_i =< SD_i, GP_i, CAT_i, MAL_i, andSC_i >$ represents the oxidative stress biomarker set of the i^{th} sample containing five biomarkers called SD, GP, CAT, MAL, and SC. The OSBs are real valued parameters and $OSBS_i \in \Re^5$. The objective is to learn from the dataset and design a prediction model. For a given new $OSBS_i$ determine the class label.

Procedure: Let the training dataset $\{S_i, y_i\}_{i=1}^{N}$, Data point $S_i \in \Re^p$ where, p is the number of predictors, class label $y_i \in \{1, 2, ..., M)$. The training dataset contain N number of samples and the number of class levels are M.

`Objective:` For a given new $s \in \Re^p$ determine the the class of s.

The procedure of k-nearest neighbor is described in Algorithm 2. For a detail description of K Nearest neighbor one can go through [33].

Algorithm 2: k Nearest Neighbor for OCD Detection

1 **Input:** Given an OCD dataset and a new OSB sample that is to be classified.
1. Let k be a positive integer and s be a new sample to be classified.
2. Evaluate the similarity of s compare to all samples of the OCD dataset using the function $distance(s, s_j) \forall j = 1, 2, ..., N$. The $distance()$ function uses euclidian distance.
3. Sort the distances in ascending order.
4. Consider the first k distances.
5. Identify the k samples corresponding to these k lowest distances.
6. Let c_i is the number of samples from i^{th} class among these k points.
7. The new sample s is classified as c_i if $c_i > c_j \forall j \neq i$

This algorithm is simulated using $R-Programming$ and experimented with OCD dataset containing OSBs and class labels. The experimental outcomes revels that this algorithm achieves an Overall OCD classification Accuracy of 0.785741, Precision of 0.814132, Recall of 0.786613, and F1-Score of 0.800102.

6 The Proposed Novel Hyperparameters Optimized Neural Network(HONN) for OCD Detection

The neural network approach is widely adapted by researchers to solve various classification problems. However, the performance of an artificial neural network model is highly relies on the selection hyperparameters such as number of computational unit layers in the network, activation function, learning rate, and number of computational units in each layer. In this research work, we propose an approach to optimize such hyperparameters of artificial neural network for OCD classification.

6.1 Neural Network and OCD Detection

Given the OCD dataset containing N number of oxidative stress biomarkers set along with the OCD class labels $\{OSBS_i, CL_i\}_{i=1}^{N}$. Where $CL_i \in \{HI, GAI, OAI\}$ and $OSBS_i =< SD_i, GP_i, CAT_i, MAL_i, and SC_i >$ represents the oxidative stress biomarker set of the i^{th} sample containing five biomarkers called SD, GP, CAT, MAL, and SC. The OSBs are real valued parameters and $OSBS_i \in \Re^5$. The objective is to learn from the dataset and design a prediction model. For a given new $OSBS_i$ determine the class label.

Procedure: A generalized neural network comprises several layers of computational computational units called Input Layer (IL), zero or more Hidden Layer (HL) and Output Layer (OL).The front-end layer of computational units is known as the IL, the backend layer of computational units is called the OL, and the computational units layers between the IL and OL are called HL. The IL feeds the input to the layer next to it, the computational results of the first HL computational units become the input to the next HL, and so on. Each computational unit performs a weighted sum of the inputs and passes to its activation function (AF). The AF is a continuous nonlinear function. Each OL computational units has a specific target value to produce for an associated input, the comparison of actual output and target value estimates an error signal. The error signals computed at the OL computational units and the associated weights are used to estimate the error signal at the computational units of the previous layer. This way the error signal propagates backward layer-by-layer. The error signal is nothing but the gradient of the error function concerning the associated input weights of the computational unit. The general structure of a neural network is presented in Fig. 3.

A set of patterns with the class labels is given as the training dataset. The training dataset is defined by

$$TP = \{X^m, t^m\}_{m=1}^{N}. \tag{13}$$

where $X^m \in \mathbb{R}^p$ is a p dimensional vector that represents the m^{th} pattern and t^m denotes the class label of X^m. Let y_i^m denote the output of computational unit i at the OL as a result of the input X^m at the IL. The error signal generated at computational unit i of the OL is defined by

$$\xi_i^m = t_i^m - y_i^m \tag{14}$$

where t_i^m is the i^{th} component of the desired response vector t^m. $\therefore$ the error energy of computational unit i is defined by

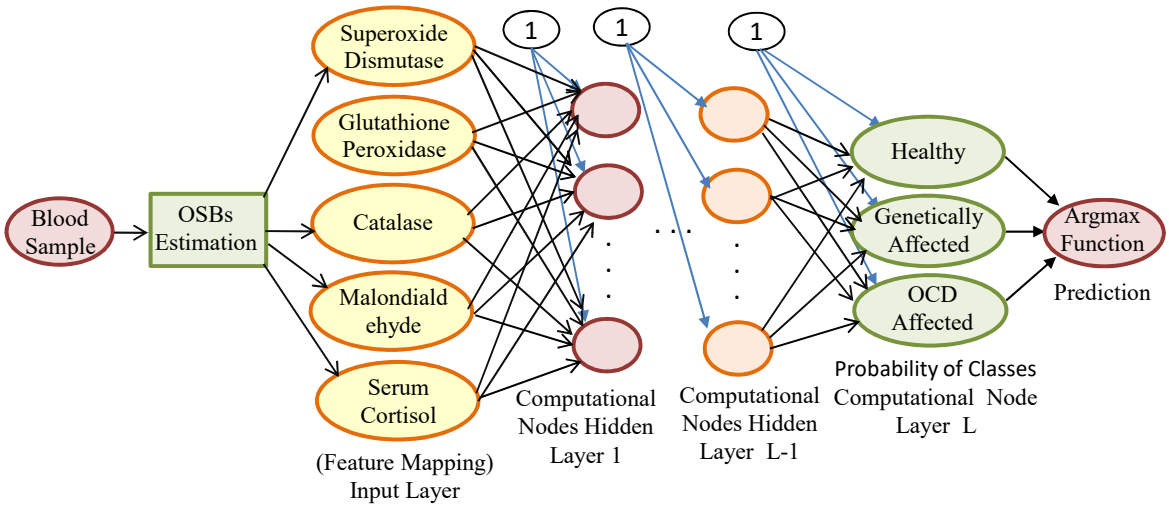

Fig. 3 Proposed ANN for automatic OCD Detection

$$E_i^m = \frac{1}{2}(\xi_i^m)^2 \tag{15}$$

Total error energy at the output layer is defined by

$$E^m = \sum_i E_i^m = \frac{1}{2}\sum_i (\xi_i^m) \tag{16}$$

The average error energy for all training samples is defined by

$$E_{av} = \frac{1}{N}\sum_{m=1}^{N}\sum_i (\xi_i^m)^2 \tag{17}$$

Let the computational unit i being fed by the computational units of the previous layer, The total input to computational unit i is defined by

$$v_i^m = \sum_{j=0}^{n} \omega_{ji} y_j^m \tag{18}$$

where n is the number of inputs (excluding the bias) to computational unit i. The weight ω_{ji}^m is associated with the link between j^{th} computational unit of the previous layer and i^{th} computational unit of the current layer. The weight ω_{0i}^m associated with the constant input $y_0 = 1$ represents the bias b_i applied to computational unit i.

The activation function of the computational unit i produces the output

$$y_i^m = \psi(v_i^m) \tag{19}$$

The weight correction $\Delta\omega_{ji}^m$ is applied to the weight ω_{ji}^m, proportional to $\frac{\partial E^m}{\partial \omega_{ji}^m}$.

$$\frac{\partial E^m}{\partial \omega_{ji}^m} = -\xi_i^m \psi_i^{'}(v_i^m) y_i^m \tag{20}$$

where $\psi_i^{'}(v_i^m) = \frac{\partial y_i^m}{\partial v_i^n}$, $\Delta\omega_{ji}^m = -\rho\frac{\partial E^m}{\partial \omega_{ji}^m} = \rho\gamma_i^m y_j^m$, ρ is the step size and $\gamma_i^m = \xi_i^m \psi_j^{'}(v_i^m)$.

For the hidden layer computational units, there is no target response. Hence, direct estimation of error signal at hidden layer computational unit is not possible. However, the error signal of the hidden layer computational unit can be estimated using the error signal of the succeeding layer computational units and the weight associated with them. Therefore, the error signal propagates backward.

The local gradient of the error signal at a hidden layer computational unit i is defined by

$$\gamma_i^m = \psi_i^{'}(v_i^m)\sum_k \gamma_k^m \omega_{ik}^m \tag{21}$$

where k denotes a computational unit in the succeeding layer computational unit, $omega_{ik}^m$ is the weight associated between computational unit i and k and γ_k^m is the local gradient of error signal at computational unit k. The local gradient of the error signal is computed from output layer computational units to the first hidden layer computational units in a backward direction.

$\therefore$ In general, the weight correction $\Delta\omega_{ji}^m$ formula can be defined as

$$\Delta\omega_{ji}^m = \rho\gamma_i^m y_j^m \tag{22}$$

where ρ is the learning rate or step size, γ_i^m is the local gradient and y_j^m is the input signal of computational unit j. The computation of γ_i^m depends on whether

neuron i is a hidden layer neuron or an output layer neuron.

This process of weight correction is performed several times by passing different samples each time from the training dataset until the average error comes down to an acceptable range.

The procedure for the neural network training ia described in Algorithm 3. For a detail description of the neural network training one can go through [34].

This algorithm is simulated using $R-Programming$ and experimented with OCD dataset containing OSBs and class labels. The experimental outcomes revels that this algorithm achieves an Overall OCD classification Accuracy of 0.833333, Precision of 0.839907, Recall of 0.83482, and F1-Score of 0.836035.

Algorithm 3: OCD Detection Network weight set learning (OCD Training Dataset TP, Network Model M, Activation Function ψ, Step size ρ, Number of Epochs EP)

Note: The OCD Training Dataset TP represents the oxidative stress biomarkers (SD, GP, CAT, MAL, and SC) of N number of individuals. The network model $M(n_0, n_1, ..., n_l, ..., n_L)$ represents the $L+1$ number of layers and n_l numbers of computational units in layer l. The input layer is represented by $l = 0$ and the output layer is represented by $l = L$. The training dataset $TP = \{X^m, t^m\}_{m=1}^N$ consists of N number of training patterns. The training patterns belong to n_0-dimensional real space, that is $X^m \in \mathbb{R}^{n_0}$. The weight ω_{ji}^l is the weight associated with the link between the j^{th} computational unit of layer $l-1$ and the i^{th} computational unit of layer l. The weight ω_{0i}^l signifies the bias of the i^{th} computational unit of the layer l.

Initialization: Initialize the weight set from a normal distribution whose mean is zero and standard deviation is a small number.

for *Epoch* $= 1$ *to* EP **do**
- **for** $n = 1$ *to* N **do**
  - Choose a random sample X from TP (i.e., OCD dataset) and set it as input to the network.
  - **for** $l = 1$ *to* L **do**
    - **for** *each computational unit* i *in layer* l **do**
      - $v_i^l = \sum_{j=1}^{m_{l-1}} \omega_{ji}^l y_i^{l-1}$ where v_i^l is the input to computational unit i in layer l, m_{l-1} is the number of computational units in layer $l-1$, y_i^{l-1} is the output of computational unit i in layer $l-1$, and $y_i^0 = X_i$.
    - $y_i^l = \psi_i^l(v_i^l)$
  - **for** $i = 1$ *to* m_L **do**
    - $\xi_i^n = t_i - y_i^L$
  - **for** $i = 1$ *to* m_L **do**
    - $\gamma_i^L = \xi_i \psi_i^{l'}(v_i^L)$
  - **for** $l = L-1$ *to* 1 **do**
    - **for** $i = 1$ *to* m_l **do**
      - $\psi_i^{l'}(v_i^l) \sum_{k=1}^{m_{l+1}} \gamma_k^{l+1} \omega_{ik}^{l+1}$
  - **for** $l = 1$ *to* L **do**
    - **for** $j = 0$ *to* m_{l-1} **do**
      - **for** $i = 1$ *to* m_l **do**
        - $\Delta\omega_{ji} = \rho \gamma_i^l y_j^{l-1}$
        - $\omega_{ji} = \omega_{ji} + \Delta\omega_{ji}$
  - **for** $n = 1$ *to* N **do**
    - Using X^n as input calculate y_n^L and estimate ξ_n.
  - $\xi_{av} = \frac{1}{N} \sum_{n=1}^N \xi_n$

Return(Weight set W)

6.2 Hyperparameter Optimization Procedure

The OCD class prediction accuracy by a neural network mainly depends on various hyperparameters. It is crucial to find optimal hyperparameters to increase the performance of a network. The approach proposed in this article adopts a finite hyperparameter set guestimating approach called HONN.

HONN ModelArchitecture Select a set of hyperparameters that needs to be optimized. For each hyperparameter choose a finite list of possible values. Initialize the hyperparameters of the neural network model by taking one value from the lists of each parameters list. Perform the neural network training by providing the training dataset. Once the neural network is trained, test it with the help of a test dataset and record the test accuracy. If the accuracy achieved in this step is better than the past models then preserve the parameters set. Repeat this process of new hyperparameter set selection from the list, training, and testing process for all combinations of hyperparameter sets. Finally, return the parameters that are preserved as best-performing ones. The conceptual architecture of the HONN is presented in Fig. 4.

Out of many hyperparameters, this approach tries to optimize the activation function, the number of layers in the network, the number of computational units in each layer, and number of epoch. The activation functions considered for different computational units of the neural network are Logistic $\left(f(v) = \frac{1}{1+e^{-v}}\right)$, Tanh $\left(f(v) = \frac{2}{1+e^{-2v}} - 1\right)$, ArcTan $\left(tan^{-1}(v)\right)$, and Softplus $(log_e(1+e^v))$. Apart from the input and output layer number of layers considered are 0, 1, and 2. The number of computational units in each layer ranges from 3

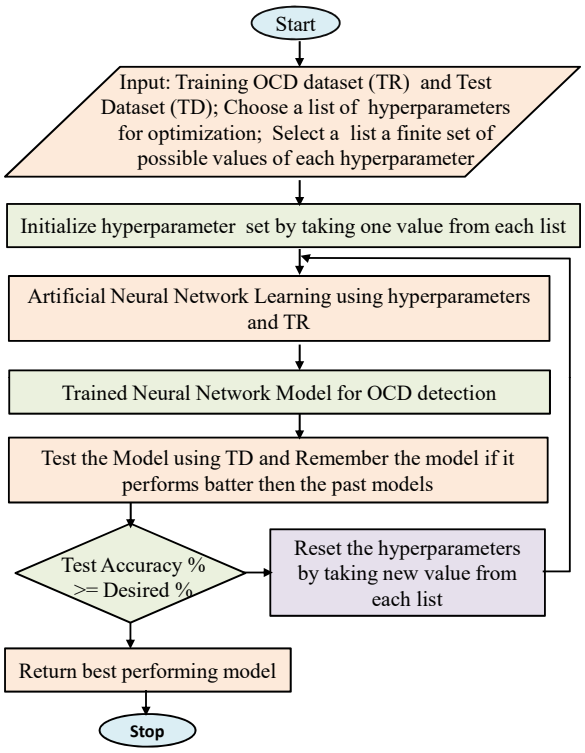

Fig. 4 Proposed flow f or n ovel H yperparameters Optimized Neural Network (HONN) modeling.

to 15. The hyperparameter optimization or model selection approach is presented in Algorithm 4.

This article uses k-Nearest Neighbor (KNN), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machine with Radial Basis Func-tion kernel function (SVMR), Support Vector Machine with Linear kernel function (SVML), Random Forest (RF), and HONN) for OSBs classification for OCD class detection. The working principle comparative study of these approach are presented in Table 3.

6.3 Experimental Setup

For a better conformation of the model accuracy a $k-fold$ cross-validation approach is adopted. k round, of experiments are performed and in each round the model accuracy is estimated. The average over all the k experiments accuracy results is considered as the model accuracy. In the $k-fold$ cross-validation approach, the given OCD dataset is randomly shuffled and divided into k nearly equal size partitions. That is the given dataset $D = \{D_1, D_2, ..., D_k\}$. In the i^{th} round of the experiment, the training dataset $TR = D_1 \cup D_2 \cup ... \cup D_{i-1} \cup D_{i+1} \cup D_k$ and the test dataset $TD = D_i$. Due to the small size of the dataset, in the current study, a $3-fold$ cross-validation process is repeated multiple times with a complete data shuffle and re-partition after

Algorithm 4: Hyperparameter optimization neural network algorithm for Accurate OCD Detection (OCD Training Dataset TP, OCD Test Dataset TD)

Input: The OCD dataset is divided into training and test dataset (i.e., TP and TD). TP is utilized to train the network where as TD is used to estimate the accuracy of the trained model.
Initialization:
AF=[Logistic, Tanh, ArcTan, Softplus] //Activation Function list
SS=[ρ_1, ρ_2, ..., ρ_m] // Step size list
EL=[EP_1,EP_1, ..., EP_n] //Number of Epoch list
$Accuracy = 0$, $Model = (5, 3)$
Procedure:
for f *in* AF **do**
  for ρ *in* SS **do**
    for EP *in* EL **do**
      for $l = 0$ *to* 2 **do**
        if $l == 0$ **then**
          $M = (5, 3)$
          W=Algorithm 3(TP, M, f, ρ, EP)
          $newAccuracy = ModelAccuracyTest(M,W,TD)$
          if $newAccuracy > Accuracy$ **then**
            $Accuracy = newAccuracy$,
            $Model = M$, Preserve W.
        if $l == 1$ **then**
          for $i = 3$ *to* 15 **do**
            $M = (5, i, 3)$
            W=Algorithm 3(TP, M, f, ρ, EP)
            $newAccuracy = ModelAccuracyTest(M,W,TD)$
            if $newAccuracy > Accuracy$ **then**
              $Accuracy = newAccuracy$,
              $Model = M$, Preserve W.
        else
          for $j = 3$ *to* 15 **do**
            for $i = 3$ *to* 15 **do**
              $M = (5, i, j, 3)$
              W=Algorithm 3(TP, M, f, ρ, EP)
              $newAccuracy = ModelAccuracyTest(M,W,TD)$
              if $newAccuracy > Accuracy$ **then**
                $Accuracy = newAccuracy$,
                $Model = M$, Preserve W.
Return(Model M, and Weight Set W)

Table 3 OCD classification approaches feature study.

Approach	Working Principles and reasons for the performance
KNN	This approach classifies the new samples based on the local majority class of the new OSB sample.
LR	The class label prediction is performed based on the probability of the sample belonging to a particular class. The higher probable class is selected.
LDA	This approach works well if the features follow normal distribution.
ANN	ANN is quite suitable to handle linearly nonseparable classes. However, the technique is not performing its best if the approach's hyperparameters are not suitably selected.
HONN	ANN HONN tries to improve the ANN approach by proposing a parameter selection algorithm.
SVMR	SVMR is a support vector machine-based approach that uses the radial basis function as a kernel function to handle the nonlinearity in the class distribution.
SVML	SVML is a support vector machine-based approach that uses the linear kernel function to handle the nonlinearity in the class distribution.
RF	RF aggregates the output of multiple decision tree classifiers to take a final call on classifi-cation results.

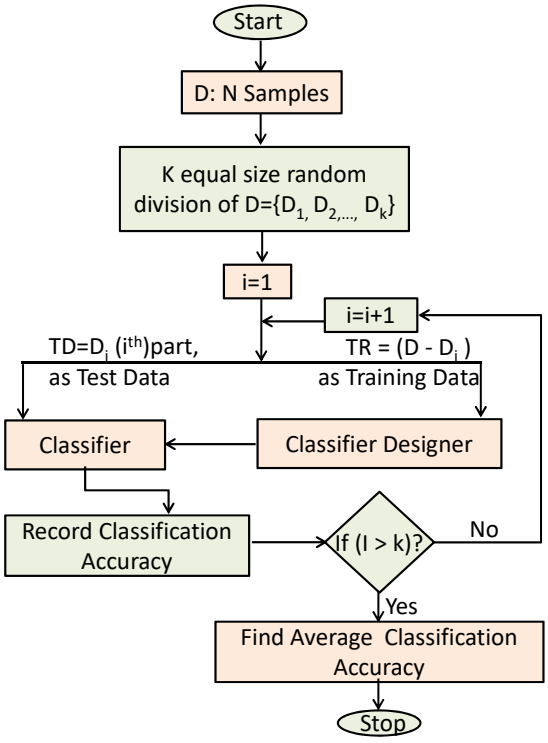

Fig. 5 Accu-Help: Experimental Setup Procedure

each $3-fold$ cross-validation to achieve $10-fold$ cross-validation. The experiments are performed on a system having `Intel(R) Core(TM) i5-3210m, 4GB RAM, 2.0 GHz Processor` and `Windows-11 as OS`. `R- 4. 2. 1` is used for programming the approaches. The $3-fold$ experimental procedure is presented in Fig. 5.

7 Experimental Evaluation

A quantitative analysis of the effectiveness of the proposed approach is performed by experimenting with the proposed approach and testing with the OCD data which was originally presented in [30]. To make a comparative analysis the performance of the approach is compared with the performance of other popular and relevant approaches by considering the same dataset as input. Among the state-of-the-art supervised approaches, some of the most popular supervised approaches such as *k-nearest neighbor, logistic regression, linear discriminant analysis, and neural network* are considered. A brief description of the dataset is given in the following paragraph.

7.1 Oxidative Stress Biomarker Dataset Description

Oxidative stress biomarkers are recorded from the blood samples of healthy individuals, OCD patients, and first-degree relatives of OCD patients. The restriction imposed during individual selection for sample collection includes: age should be between 18 and 45, should not be under any medications in the last three months, should not have any illness in case of healthy and first-degree relatives, and should not have any other illness in case of OCD patients. Pregnant and lactating individuals are excluded. The oxidative stress biomarkers considered for this study are SD, GP, CAT, MAL, and SC. Standard biochemical methods are adopted to measure these biomarkers. To estimate these biomarkers, blood samples are drawn after overnight fasting. The assessment of these biomarkers is carried out in the biochemical laboratory of King George Medical University. Plasma is used as the source of enzymes. An in-depth description of the laboratory mechanism of assessment of these biomarkers is presented in [30].

The biomarkers recorded are in the form of real numbers. The range of the values varies from marker to marker. The distribution of the values in different

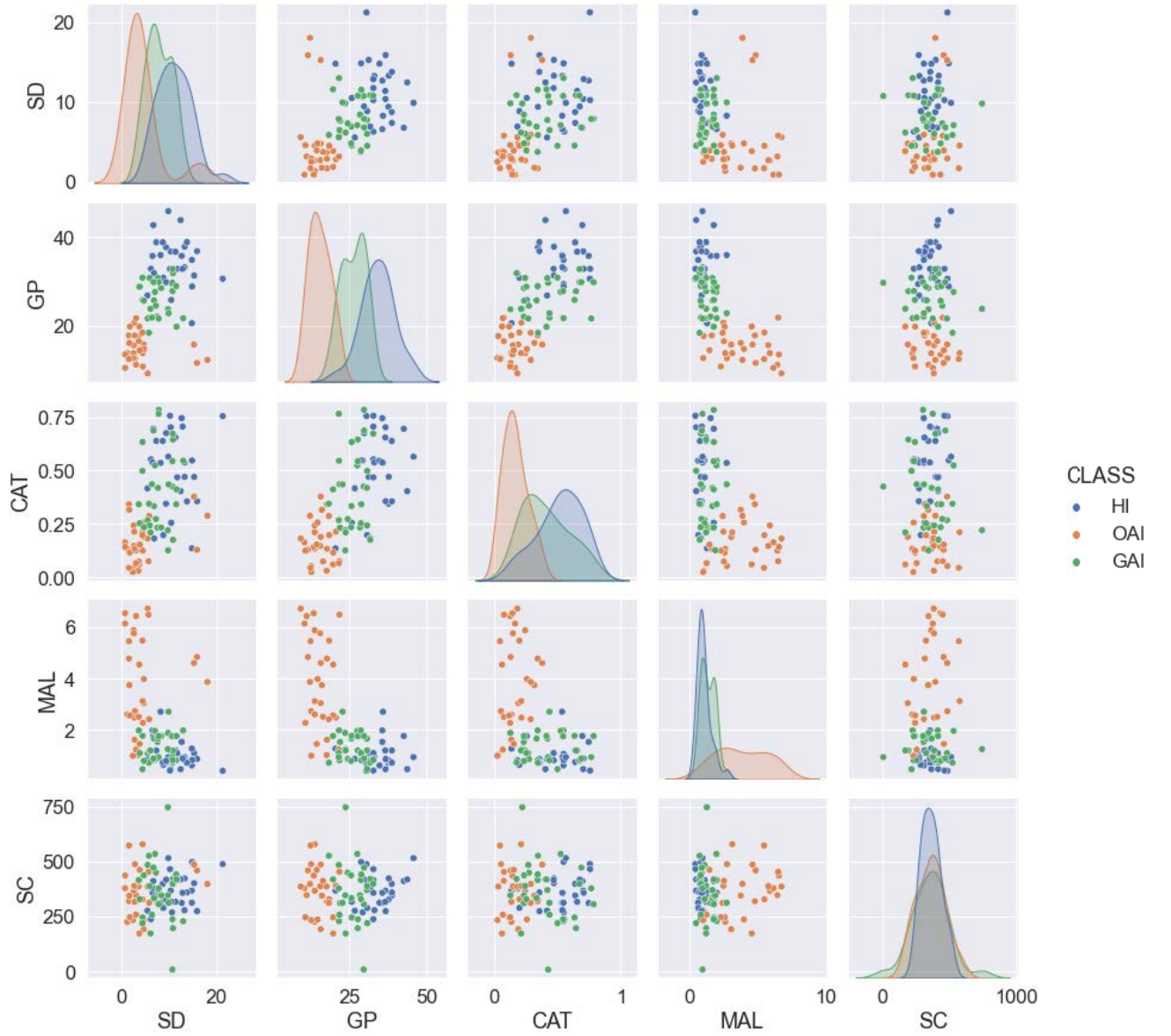

Fig. 6 Class density distribution over biomarkers and all biomarker pair scatter plots

metrics and all biomarker pair wise scatter plot is presented in Fig. 6.

To project all the values of the markers into a common range a normalization procedure is followed. Let $\beta = \{\beta_1, \beta_2, ..., \beta_p\}$ represents the set of biomarkers where p is the number of biomarkers.

Let $\beta_i = \{\beta_i^1, \beta_i^2, ..., \beta_i^N\}$ be the collection of i^{th} marker for all N samples. Let β_i^s and β_i^l represent the smallest and largest value among the values stored in β_i respectively. To normalize β_i in the range $[n_1, n_2]$, we adopt the min-max normalization method. The normalized value $\overline{\beta_i^j} = n_1 + (n_2 - n_1)\left(\frac{\beta_i^j - \beta_i^s}{\beta_i^l - \beta_i^s}\right)$.

7.2 Quantitative Analysis

To access the effectiveness of the model, a test dataset $TD = \{X_i, y_i\}_{i=1}^{M}$ is used which consists of M number of samples and these samples are not used for the training of the model.

$X_i =< X_i^{SD}, X_i^{GP}, X_i^{CAT}, X_i^{MAL}, X_i^{SC} >$ represents the five oxidative stress biomarkers of the i^{th} sample of the test dataset and y_i is the class label of X_i. The test samples (X_i) are presented to the model to get the predicted output $\hat{y}_i$ of a model and $\hat{y}_i$ is compared with y_i to get the test accuracy of the model. The probable outcomes of this comparison is given in equation 23 where TP_c represents the true positive for class c, TN_c represents the true negative for class c,

Table 4 Classification accuracy metrics.

Accuracy Metric	Estimation Method
Overall Accuracy	$\left(\sum_{c=1}^{C} \frac{TP_c+TN_c}{TP_c+FN_c+FP_c+TN_c}\right)/3$
Precision	$\frac{\sum_{c=1}^{C} TP_c}{\sum_{c=1}^{C}(TP_c+FP_c)}$
Recall	$\frac{\sum_{c=1}^{C} TP_c}{\sum_{c=1}^{C}(TP_c+FN_c)}$
F1-Score	$2*\left(\frac{Precision*Recall}{Precision+Recall}\right)$

FP_c represents the false positive for class c, and FN_c represents the true positive for class c,

$$Comp(\hat{y}_i, y_i) \in \begin{cases} TP_c & if\ \hat{y}_i = y_i = c\ (class\ label\ ) \\ TN_c & if\ \hat{y}_i\ and\ y_i \neq c \\ FP_c & if\ \hat{y}_i = c\ and\ y_i \neq c \\ FN_c & if\ \hat{y}_i \neq c\ and\ y_i = c \end{cases} \tag{23}$$

The output of the *Comparison* function for all the samples can be represented in a confusion matrix. The confusion matrix CM for a 3 class classification problem can be represented as a (3×3) matrix, where $CM[i, j]$ represents the number of samples predicted as class i and the actual class of the sample is j. $\therefore TP_1 = CM[1,1]$, $TP_2 = CM[2,2]$, $TP_3 = CM[3,3]$, $TN_1 = CM[2,2] + CM[2,3] + CM[3,2] + CM[3,3]$, $TN_2 = CM[1,1] + CM[1,3] + CM[3,1] + CM[3,3]$, $TN_3 = CM[1,1] + CM[1,2] + CM[2,1] + CM[2,2]$, $PF_1 = CM[1,2]+CM[1,3]$, $PF_2 = CM[2,1]+CM[2,3]$, $PF_3 = CM[3,1]+CM[3,2]$, $FN_1 = CM[2,1]+CM[3,1]$, $FN_2 = CM[1,2]+CM[3,2]$, and $FN_3 = CM[1,3]+CM[2,3]$.

Based on the values of TP_c, TN_c, FP_c , and FN_c of all the classes, one can estimate classification accuracy metrics such as *Overall Accuracy, Precision, Recall, and F1-Score*. The formulas used to estimate these metrics are given in Table 4.

7.3 Comparative Analysis

For comparative analysis, the experiments are performed twofold. In the first fold, manually different variants of neural networks are selected and experimented with for OCD classification and the outcomes are compared with the outcomes of the proposed model. In the second fold of the experiment, some of the popular classification approaches (such as k-Nearest Neighbor, Logistic Regression, Linear Discriminant Analysis, Support Vector Machine with Radial Basis Function kernel function, Support Vector Machine with Linear kernel function, Random Forest, and Neural Network) are selected and experimented with for OCD classification and the outcomes are compared with the outcomes of the proposed model.

Table 5 Classification accuracy achieved by the neural network models: Mean values of all accuracy measures over 10-fold validation (HLN: Hidden Layer Numbers, and HLNN: Hidden Layer computational unit Numbers)

Accuracy measure	HLN	HLNN	Accuracy
Overall Accuracy	1	6	0.833333
Precision			0.839907
Recall			0.83482
F1 score			0.836035
Overall Accuracy	1	10	0.777778
Precision			0.768943
Recall			0.771247
F1 score			0.770093
Overall Accuracy	1	15	0.811111
Precision			0.825838
Recall			0.820978
F1 score			0.823345
Overall Accuracy	2	5,5	0.755556
Precision			0.769639
Recall			0.747397
F1 score			0.75834
Overall Accuracy	2	10,8	0.788889
Precision			0.800385
Recall			0.773107
F1 score			0.786435
Overall Accuracy	2	15,10	0.733333
Precision			0.74236
Recall			0.738571
F1 score			0.740436
Overall Accuracy	3	5,5,4	0.7
Precision			0.705804
Recall			0.686802
F1 score			0.695818
Overall Accuracy	3	10,8,5	0.655556
Precision			0.668515
Recall			0.665826
F1 score			0.666983
Overall Accuracy	3	15,10,8	0.722222
Precision			0.738492
Recall			0.731519
F1 score			0.734933

Nine different neural network architectures are selected for OCD classification. The models which are chosen are defined by $M1(5,\ 6,\ 3)$, $M2\ (5,\ 10,\ 3)$, $M3(5,\ 15,\ 3)$, $M4\ (5,\ 5,\ 5, 3)$, $M5\ (5,\ 10,\ 8,\ 3)$, $M6\ (5,\ 15,\ 10,\ 3)$, $M7\ (5,\ 5,\ 5,\ 4,\ 3)$, $M8\ (5,\ 10,\ 8,\ 5,\ 3)$, and $M9\ (5,\ 15,\ 10,\ 8,\ 3)$. Logistic function is used as the activation function for all the computational units. The step size or learning rate is set to 0.005 and the maximum number of epoch is taken as 10000. All these models goes through a $10 - fold$ cross-validation process. The *Overall Accuracy, Error Rate, Precision, Recall, Micro Averaging F1-Score, Macro Averaging F1-Score* obtained by different models are presented in Table 5. For visual analysis, bar plots are used for each

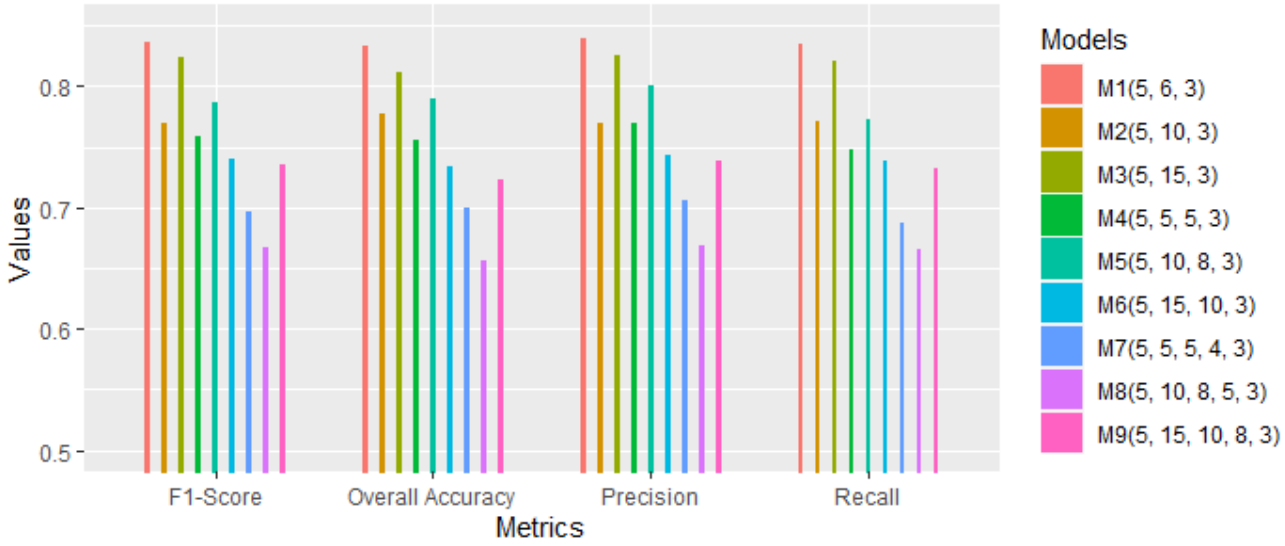

Fig. 7 Classification accuracy comparison bar plots for all network models

Table 6 Classification accuracy achieved by ANN, KNN, LR, LDA, and HONN : Mean values of all accuracy measures over 10-fold validation

Metric	ANN	KNN	LR	LDA	HONN	SVMR	SVML	RF
Overall Accuracy	0.833333	0.785741	0.777789	0.821431	0.861111	0.848511	0.821421	0.821433
Precision	0.839907	0.814132	0.768943	0.833333	0.8650794	0.841111	0.821111	0.821433
Recall	0.83482	0.786613	0.771247	0.828347	0.8630752	0.843821	0.823521	0.821411
F1-Score	0.836035	0.800102	0.770093	0.830827	0.8640761	0.843821	0.821421	0.831131

accuracy measure. The *x-axis* of these plots is marked with the metrics, the *y-axis* is marked with the accuracy level, and different colours are used to represent different models. The bar plots are presented in Fig. 7. From Table 5 and Fig. 7 it can be observed that the model $M1$ performs batter compared to others with an *Overall Accuracy* of 0.833333, *Precision* of 0.839907, *Recall* of 0.83482 and *F1-Score* of 0.836035.

Further, experiments are performed using *KNN, LR, LDA, SVMR, SVML, RF, Neural Network (Model M1) (ANN), and HONN* for OCD classification. The hyperparameters of these models are tuned using hit-and-trial method. $10 - fold$ cross-validation process is fol-lowed to estimate performance accuracy. For a com-parative analysis, the results obtained are presented in Table 6. Accuracy measure wise bar plots are presented in Fig. 8 for a visual comparative analysis. From Table 6 and Fig. 8 it can be observed that *HONN* performs batter compared to others with an *Overall Accuracy* of 0.861111, *Precision* of 0.8650794, *Recall* of 0.8630752 and *F1-Score* of 0.8640761.

8 Conclusion and Future Directions of Research

This article proposes an H-CPS called Accu-Help for OCD detection. Data collection, class label integration, machine learning model design, and online OCD classification process can be managed and monitored using Accu-Help. The core component of the H-CPS is a machine-learning model for OCD prediction. In this machine learning model, a hyperparameter-optimized neural network (HONN) approach is proposed to classify oxidative stress biomarkers into one category from HI, GAI, and OAI. The OCD detection accuracy by the network is mainly dependent on the hyperparameters set for the network. Deciding the best set of hyperparameters is a real challenge. This challenge can be reduced by adopting the HONN method. For comparative study *k-Nearest Neighbor, Logistic Regression, Linear Discriminant Analysis, and Artificial Neural Network* are used. From experimental result analysis, it is observed that HONN yields better accuracy in the test dataset with respect to all the classification measures. The OCD detection accuracy achieved by HONN is $86 \pm 2\%$. As GAI identification is possible with this approach, appropriate preventive actions can be recommended well in advance.

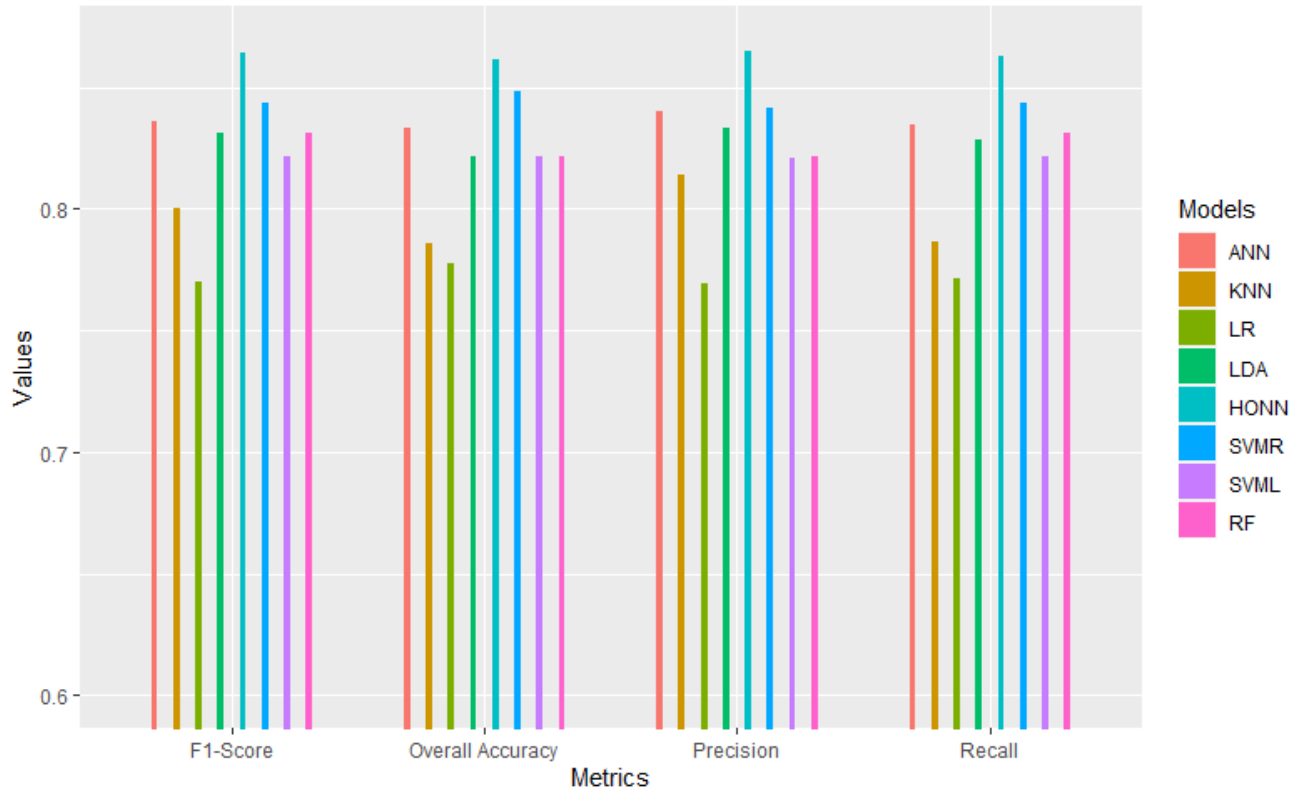

Fig. 8 Classification accuracy comparison bar plots for ANN, KNN, LR, LDA, and HONN:: Mean values of all accuracy measures over 10-fold validation

The future research directions include the expansion of the scope of the Accu-Help and the design of a more accurate machine-learning model for OCD detection. Even though the proposed machine learning model performs well compared to popular classification models, the approach is not scalable to use in real life because this approach is computationally costly. In the future, we will try to propose a computationally less costly and scalable model with better OCD detection accuracy.

Acknowledgment

Funding: This study was funded by Odisha Higher Education Programme for Excellence and Equity (OHEPEE) World Bank (6770/GMU).

An earlier version of this paper is made available as a preprint [35].

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. I. L. Olokodana, S. P. Mohanty, E. Kougianos, and R. S. Sherratt, "EZcap: A novel wearable for real-time automated seizure detection from EEG signals," *IEEE Transactions on Consumer Electronics*, vol. 67, no. 2, pp. 166–175, 2021.
2. L. Rachakonda, A. K. Bapatla, S. P. Mohanty, and E. Kougianos, "BACTmobile: A Smart Blood Alcohol Concentration Tracking Mechanism for Smart Vehicles in Healthcare CPS Framework," *SN Computer Science*, vol. 3, no. 3, pp. 1–24, 2022.
3. S. U. Amin, M. S. Hossain, G. Muhammad, M. Alhussein, and M. A. Rahman, "Cognitive smart healthcare for pathology detection and monitoring," *IEEE Access*, vol. 7, pp. 10 745–10 753, 2019.
4. T. Vos, A. A. Abajobir, K. H. Abate, C. Abbafati, K. M. Abbas, F. Abd-Allah, R. S. Abdulkader, A. M. Abdulle, T. A. Abebo, S. F. Abera *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *The Lancet*, vol. 390, no. 10100, pp. 1211–1259, 2017.
5. S. Mas, P. Gasso, A. Morer, A. Calvo, N. Bargallo, A. Lafuente, and L. Lazaro, "Integrating genetic, neuropsychological and neuroimaging data to model early-onset obses-

sive compulsive disorder severity," *PLoS One*, vol. 11, no. 4, p. 153846, 2016.
6. S. Aydin, N. Arica, E. Ergul, and O. Tan, "Classification of obsessive compulsive disorder by EEG complexity and hemispheric dependency measurements," *International journal of neural systems*, vol. 25, no. 03, p. 1550010, 2015.
7. T. T. Erguzel, S. Ozekes, G. H. Sayar, O. Tan, and N. Tarhan, "A hybrid artificial intelligence method to classify trichotillomania and obsessive compulsive disorder," *Neurocomputing*, vol. 161, pp. 220–228, 2015.
8. P. Desarkar, V. K. Sinha, K. Jagadheesan, and S. H. Nizamie, "A high resolution quantitative EEG power analysis of obsessive-compulsive disorder," *German Journal of Psychiatry*, vol. 10, no. 2, pp. 29–35, 2007.
9. J. Kopřivová, M. Congedo, J. Horáček, J. Praško, M. Raszka, M. Brunovskỳ, B. Kohútová, and C. Höschl, "EEG source analysis in obsessive–compulsive disorder," *Clinical Neurophysiology*, vol. 122, no. 9, pp. 1735–1743, 2011.
10. G. Salomoni, M. Grassi, P. Mosini, P. Riva, P. Cavedini, and L. Bellodi, "Artificial neural network model for the prediction of obsessive-compulsive disorder treatment response," *Journal of clinical psychopharmacology*, vol. 29, no. 4, pp. 343–349, 2009.
11. M. Q. Hoexter, E. C. Miguel, J. B. Diniz, R. G. Shavitt, G. F. Busatto, and J. R. Sato, "Predicting obsessive–compulsive disorder severity combining neuroimaging and machine learning methods," *Journal of affective disorders*, vol. 150, no. 3, pp. 1213–1216, 2013.
12. M. Weygandt, C. R. Blecker, A. Schäfer, K. Hackmack, J.-D. Haynes, D. Vaitl, R. Stark, and A. Schienle, "fMRI pattern recognition in obsessive–compulsive disorder," *Neuroimage*, vol. 60, no. 2, pp. 1186–1193, 2012.
13. T. L. Gehris, R. G. Kathol, D. W. Black, and R. Noyes Jr, "Urinary free cortisol levels in obsessive-compulsive disorder," *Psychiatry research*, vol. 32, no. 2, pp. 151–158, 1990.
14. M. Kuloglu, M. Atmaca, E. Tezcan, Ö. Gecici, H. Tunckol, and B. Ustundag, "Antioxidant enzyme activities and malondialdehyde levels in patients with obsessive-compulsive disorder," *Neuropsychobiology*, vol. 46, no. 1, pp. 27–32, 2002.
15. S. Ersan, S. Bakir, E. E. Ersan, and O. Dogan, "Examination of free radical metabolism and antioxidant defence system elements in patients with obsessive–compulsive disorder," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 30, no. 6, pp. 1039–1042, 2006.
16. A. Behl, G. Swami, S. Sircar, M. Bhatia, and B. Banerjee, "Relationship of possible stress-related biochemical markers to oxidative/antioxidative status in obsessive-compulsive disorder," *Neuropsychobiology*, vol. 61, no. 4, pp. 210–214, 2010.
17. A. Shrivastava, S. K. Kar, E. Sharma, A. A. Mahdi, and P. K. Dalal, "A study of oxidative stress biomarkers in obsessive compulsive disorder," *Journal of Obsessive-Compulsive and Related Disorders*, vol. 15, pp. 52–56, 2017.
18. M. Nouman, S. Y. Khoo, M. P. Mahmud, and A. Z. Kouzani, "Recent Advances in Contactless Sensing Technologies for Mental Health Monitoring," *IEEE Internet of Things Journal*, vol. 9, pp. 274 – 297, 2022.
19. P. Jain, A. M. Joshi, and S. P. Mohanty, "iGLU: An intelligent device for accurate noninvasive blood glucose-level monitoring in smart healthcare," *IEEE Consumer Electronics Magazine*, vol. 9, no. 1, pp. 35–42, 2019.
20. L. Catarinucci, D. De Donno, L. Mainetti, L. Palano, L. Patrono, M. L. Stefanizzi, and L. Tarricone, "An IoT-aware architecture for smart healthcare systems," *IEEE internet of things journal*, vol. 2, no. 6, pp. 515–526, 2015.
21. A. K. Tripathy, A. G. Mohapatra, S. P. Mohanty, E. Kougianos, A. M. Joshi, and G. Das, "EasyBand: a wearable for safety-aware mobility during pandemic outbreak," *IEEE Consumer Electronics Magazine*, vol. 9, no. 5, pp. 57–61, 2020.
22. K. D. Askland, S. Garnaat, N. J. Sibrava, C. L. Boisseau, D. Strong, M. Mancebo, B. Greenberg, S. Rasmussen, and J. Eisen, "Prediction of remission in obsessive compulsive disorder using a novel machine learning strategy," *International journal of methods in psychiatric research*, vol. 24, no. 2, pp. 156–169, 2015.
23. F. Lenhard, S. Sauer, E. Andersson, K. N. Månsson, D. Mataix-Cols, C. Rück, and E. Serlachius, "Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach," *International Journal of Methods in Psychiatric Research*, vol. 27, no. 1, p. e1576, 2018.
24. P. Piaggi, D. Menicucci, C. Gentili, G. Handjaras, A. Gemignani, and A. Landi, "Singular spectrum analysis and adaptive filtering enhance the functional connectivity analysis of resting state fMRI data," *International journal of neural systems*, vol. 24, no. 03, p. 1450010, 2014.
25. D. Rangaprakash, X. Hu, and G. Deshpande, "Phase synchronization in brain networks derived from correlation between probabilities of recurrences in functional MRI data," *International journal of neural systems*, vol. 23, no. 02, p. 1350003, 2013.
26. J. Cheng, P. Li, Y. Tang, C. Zhang, L. Lin, J. Gao, and Z. Wang, "Transcranial direct current stimulation improve symptoms and modulates cortical inhibition in obsessive-compulsive disorder: A TMS-EEG study," *Journal of Affective Disorders*, vol. 298, pp. 558–564, 2022.
27. M. A. Özçoban, O. Tan, S. Aydin, and A. Akan, "Decreased global field synchronization of multichannel frontal eeg measurements in obsessive-compulsive disorders," *Medical & Biological Engineering & Computing*, vol. 56, no. 2, pp. 331–338, 2018.
28. M. Kluge, P. Schüssler, H. E. Künzel, M. Dresler, A. Yassouridis, and A. Steiger, "Increased nocturnal secretion of ACTH and cortisol in obsessive compulsive disorder," *Journal of psychiatric research*, vol. 41, no. 11, pp. 928–933, 2007.
29. M. H. Shohag, M. A. Ullah, M. A. Azad, M. S. Islam, S. Qusar, S. F. Shahid, and A. Hasnat, "Serum Antioxidant Vitamins and Malondialdehyde Levels in Patients with Obsessive-Compulsive Disorder," *German Journal of Psychiatry*, vol. 15, 2012.
30. S. K. Kar, I. Choudhury *et al.*, "An empirical review on oxidative stress markers and their relevance in obsessive-compulsive disorder," *International Journal of Nutrition, Pharmacology, Neurological Diseases*, vol. 6, no. 4, p. 139, 2016.
31. A. Field, "Logistic regression," *Discovering statistics using SPSS*, vol. 264, p. 315, 2009.
32. A. J. Izenman, "Linear discriminant analysis," in *Modern multivariate statistical techniques*. Springer, 2013, pp. 237–280.
33. O. Kramer, "K-nearest neighbors," in *Dimensionality reduction with unsupervised nearest neighbors*. Springer, 2013, pp. 13–23.
34. G. Bebis and M. Georgiopoulos, "Feed-forward neural networks," *IEEE Potentials*, vol. 13, no. 4, pp. 27–31, 1994.
35. K. Patel, A. K. Tripathy, L. N. Padhy, S. K. Kar, S. K. Padhy, and S. P. Mohanty, "Accu-help: A machine learning based smart healthcare framework for accurate detection of obsessive compulsive disorder," 2022.