Informatics in Medicine Unlocked 23 (2021) 100545

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Application of deep and machine learning techniques for multi-label classification performance on psychotic disorder diseases

Israel Elujide [a], Stephen G. Fashoto [b,*], Bunmi Fashoto [c], Elliot Mbunge [b], Sakinat O. Folorunso [d], Jeremiah O. Olamijuwon [e]

[a] Department of Computer Science and Engineering, The University of Texas at Arlington, Texas, USA
[b] Department of Computer Science, University of Eswatini, Kwaluseni, Eswatini
[c] Department of Psychology, Eswatini Christian University, Mbabane, Eswatini
[d] Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria
[e] EBlocks Software, 138 West Street, Sandton, Johannesburg, South Africa

ARTICLE INFO

Keywords:
Deep neural networks
Deep learning class imbalance
Mental disorder multi-label classification
Psychotic disorder disease

ABSTRACT

Electronic Health Records (EHRs) hold symptoms of many diverse diseases and it is imperative to build models to recognise these problems early and classify the diseases appropriately. This classification task could be presented as a single or multi-label problem. Thus, this study presents Psychotic Disorder Diseases (PDD) dataset with five labels: bipolar disorder, vascular dementia, attention-deficit/hyperactivity disorder (ADHD), insomnia, and schizophrenia as a multi-label classification problem. The study also investigates the use of deep neural network and machine learning techniques such as multilayer perceptron (MLP), support vector machine (SVM), random forest (RF) and Decision tree (DT), for identifying hidden patterns in patients' data. The study furthermore investigates the symptoms associated with certain types of psychotic diseases and addresses class imbalance from a multi-label classification perspective. The performances of these models were assessed and compared based on an accuracy metric. The result obtained revealed that deep neural network gave a superior performance of 75.17% with class imbalance accuracy, while the MLP model accuracy is 58.44%. Conversely, the best performance in the machine learning techniques was exhibited by the random forest model, using the dataset without class imbalance and its result, compared with deep learning techniques, is 64.1% and 55.87%, respectively. It was also observed that patient's age is the most contributing feature to the performance of the model while divorce is the least. Likewise, the study reveals that there is a high tendency for a patient with bipolar disorder to have insomnia; these diseases are strongly correlated with an R-value of 0.98. Our concluding remark shows that applying the deep and machine learning model to PDD dataset not only offers improved clinical classification of the diseases but also provides a framework for augmenting clinical decision systems by eliminating the class imbalance and unravelling the attributes that influence PDD in patients.

1. Introduction

In the past, psychotic disorder diseases (PDD) relied on traditional approaches which were constructed from expert opinion and enshrined in the International Classification of Diseases (ICD)-11, Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 and the National Institute of Mental Health's Research Domain Criteria (RDoC) to understand and classify mental disorder [1] yet it is increasingly becoming clear that the pathophysiology underlying psychotic disorder diseases is rather heterogeneous [2]. Presently, in the place of the traditional approaches, the statistical, machine learning, natural language processing techniques, and neuroimaging have been explored for early detection of the PDD [3,4] but the reliability of the findings is unclear, due to potential methodological issues that may have inflated the existing literature. In addition to that, currently, there are two major limitations in the existing literature that restrict the translational applicability of the findings in real-world clinical practice [5]. The applicability of machine learning based diagnostic tool for detecting patients with established psychotic disorder diseases is minimal [5] for clinical utilities, which is in contrast with the real-world clinical practice, since detecting diseases

* Corresponding author.
E-mail addresses: gbengafash2@gmail.com, sgfashoto@uniswa.sz (S.G. Fashoto).

https://doi.org/10.1016/j.imu.2021.100545
Received 21 June 2020; Received in revised form 28 February 2021; Accepted 1 March 2021
Available online 17 March 2021

at an early stage of sickness may be unpredictable and sometimes treatment is yet to be decided. Diagnosing a patient with PDD is exclusively clinically stressful [6] for physicians, psychologists, and psychiatrists because they need to know the medical and psychiatric history of the patient. The PDD coexists with non-communicable chronic diseases (NCD) [7]. The diagnosis of psychotic disorder diseases such as Schizophrenia is based on the DSM-5 and ICD, and is done clinically, since specific biomarkers that can predict the illness accurately remain unknown [8]. Emerging technologies like machine learning methods (such as pattern recognition, support vector machines, multivariate pattern analysis, Gaussian processes, logistic regression, random forest, neural networks) [4,9] and magnetic resonance imaging (MRI) [8,10] when applied to neuroimage data represent a new and promising approach that could support the diagnosis of mental disorders. The issue of PDD continues to persist and it is one of the most devastating mental illnesses globally [2,4], because it incapacitates individuals psychologically. There is presently no known cure for mental illness but timely detection and prompt intervention can aid in slowing down the illness [2,4,11]. The onset of mental illness is primarily preceded by non-severe symptoms which are a major challenge in diagnosing the early onset of mental illness or psychosis. The DSM-5 and ICD-10 systems require the elimination of medical conditions before diagnosing psychotic disorder [12]. In an attempt to overcome the devastating state of PDD, psychiatrists and psychologists in the psychiatric community teamed up with computer scientists and engineers to develop machine learning algorithms for the prediction of psychotic disorder diseases, simultaneously, based on the existing dataset [3,13,14].

The outline of the paper is as follows: Section 2 describes the types of mental disorder discussed in this study -Deep and Machine learning techniques for Multi-label Classification, deep learning approaches and Class imbalance in Machine Learning Techniques. Section 3 describes the methods adopted in this study. The PDD dataset used will be described and also represented as a multi-label problem. Also, the Deep Learning architecture and the machine learning techniques used for the multi-classification and class imbalance will be extensively described. Section 4 will discuss the results obtained, based on the measures used in the study. Lastly, section 5 will conclude the work and describe future work.

2. Mental disorder

2.1. Types of mental disorder

There are five types of mental disorders considered in this study, based on the available dataset in Nigeria to label psychotic patients; these include bipolar disorder, attention deficit hyperactive disorder, schizophrenia, vascular dementia and insomnia.

2.1.1. Bipolar disorder

Bipolar disorder is an unusual mood change which is often extreme and fluctuating, and which occurs at irregular intervals. An emotion ranges between euphonic feelings and bout of depression [15]. Bipolar disorder is sub-categorised into Bipolar I and Bipolar II. Diagnostic symptoms of bipolar I include mania episodes) to hypomanic episodes while Bipolar II diagnostic symptoms range from mild depression to moderate and extreme depressive episodes [16].

2.1.2. Attention deficit hyperactive disorder (ADHD)

The new name for Minimal Brain dysfunction (MBD) in the psychiatric space nowadays is Attention Deficit Hyperactivity Disorder (ADHD). Attention-deficit is characterised by the inability to sustain and/or maintain attention, control impulses and regulate the level of physical activities [17] while hyperactivity disorder is described as in children that are highly impulsive, acting too fast without thinking to complete a task and rebuff waiting for their turn [18]. ADHD in children (especially between 3 and 6 years old), young people and adults conditions them to perceive and react to most stimuli in their environment and this predisposes them to accidents or injury, sleep disturbance, aggression, mood swings, substance misuse, immature language, anxiety states, academic underachievement and unpopularity with peers [19]. Since ADHD symptoms differ from one person to another, its diagnosis takes time and in some cases, results in high error rates, due to different clinical examinations. Due to ADHD symptoms that differ from one person to another, ADHD can appear in three different ways among different people: inattention, hyperactivity-impulsivity, and a combination of both [20].

2.1.3. Schizophrenia

Schizophrenia is characterised by a collection of symptoms such as deterioration of mental functioning, language disturbance, disjointed speech, social withdrawal, hallucination, motor disturbance and irrational thinking [6]. Schizophrenia is a disorder that impacts every area of an individual's psychological functioning, and which is characterised by severe deviation from reality [12]. Prevalence of schizophrenia in the adult population is between 0.3 and 0.7%, though it is higher in males than female adults [21].

2.1.4. Vascular dementia

Vascular Dementia (VD) is a neuro-cognitive disorder which is characterised by a progressive cognitive decline in memory and cognitive functions. That is, such individuals decline in language, complex attention, learning and memory, social cognition and motor function [12]. Individuals with the dementia condition may become suicidal, depressed and harmful to themselves and others [12]. Prevalence is high among older adults (above 65 years) and varies between 2% and 25% in the adult population. The World Health Organisation estimates that 35.6 million people live with dementia, a number that is anticipated to triple by 2050, as 7.7 million new cases of dementia are diagnosed every year, posing a financial burden to the society [22].

2.1.5. Insomnia

Insomnia is also known as a sleep disorder and it is a chronic sleep disorder. An individual with insomnia experiences the challenge of falling asleep and is unable to maintain sleep [23]. This mental condition may be intermittently ranging from acute (a few weeks) to chronic (several months).

2.2. Deep and machine learning techniques for multi-label classification on PDD

Several models have been applied to address single-labelled classification and multi-class problems. A *single-labelled* classification bothers with learning from a set of instances that are related with a single label l from a set of disjoint labels L, $|L| > 1$. If $|L| = 2$, then this is termed *binary* classification problem. A multi-class problem is characterised by $|L| > 2$. Whereas, a *multi-label* classification is characterised by instances associated with a set of labels $Y \subseteq L$ [24]. For instance, authors in Ref. [25] presented an unsupervised deep feature learning method named 'deep patient' with a 3-layered stack of denoising autoencoders to capture hierarchical regularities and dependencies in the aggregated EHRs. These EHRs contain about 700,000 patients as samples with 78 diseases as labels from the Mount Sinai data warehouse which were split into the ratio of 89.11–10.89 train – test split. Results obtained revealed that severe diabetes, schizophrenia, and various cancers performed best.

The approach taken by Ref. [4] is a semantic and latent content analysis. It considered semantic density which was determined by the number of parts in a sentence which makes the sentence meaningful. Furthermore, it compared an individual's speech with a large database of other people's speech patterns to determine when the speech becomes abnormal. This results in the prediction of possible psychosis in an individual. Machine learning techniques in the form of a two-layer neural network were then implemented with these two linguistic indicators to predict psychosis. The authors in Ref. [3] proposed the use of machine learning evaluation on PDD like label and ranking-based perspectives. The evaluation metrics considered in the study are hamming loss (HL), one error (OE), zero-one loss (ZOL), ranking loss (RL), accuracy (Acc), average precision (AP), Micro-F1 and Macro-F1. The dataset for the study was evaluated on the Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Model Tree (LMT), and Naïve Bayes Tree (NBTree), base classifiers in the Problem Transformation (PT) and Ensemble methods. The PT approach uses Binary relevance (BR), Classifier Chains (CC), Probabilistic Classifier Chains (PCC), Pruned Set (PS), FW and RT, while the ensemble approach uses Random k-label sets (RAkEL), RAkELd, EBR, Ensemble of Classifier Chains (ECC), EFW, EPCC, Ensemble of Pruned Set (EPS), ELC and ERT. The study shows that the Label Powerset (LP) and Pruned Sets (PS) in the multi-label classification methods, with Naïve Bayes (NB) and Naïve Bayes Tree (NBTree), consistently performed best in terms of the evaluation measures on the PDD dataset.

Schizophrenia diagnosis was examined in Ref. [26] by utilising a hybrid of artificial intelligence and a knowledge base supplying an expert system to make accurate diagnoses. The knowledge base for classifying psychotic disorders was sourced from previously established works and models for multi-criteria decision analysis, while artificial intelligence was used to create production rules and probabilities. The work in Ref. [6] reviewed 35 previous studies that utilised various combinations of different machine learning techniques, such as support vector machines, multivariate pattern analysis, and random forest, in order to analyse neuro-images and determine the presence of schizophrenia in a particular subject. The use of artificial intelligence and machine learning in medicine and psychiatry were considered in Ref. [27]. This study looks at a possible future using these technologies in these industries, and in particular, the benefits and challenges.

2.2.1. Overview of deep learning approaches

The increase in psychotic disorder cases and the popularity of deep learning algorithms present unprecedented opportunities for the application of deep learning algorithms, for modelling and classifying PDD, using patients' data. Deep learning (DL) approaches achieved remarkable results in many domains, thereby revolutionising the field of machine learning through deep hierarchical feature construction in the dataset. DL is a subset of machine learning, composed of multi-layered neural networks; one or more hidden layers are connected to each other to form a network that is capable of learning complex structures with a high level of abstraction [28]. Deep learning approaches such as convolutional neural networks (CNN), recurrent neural network (RNN), deep reinforcement learning (DRL) and multilayer feedforward networks can learn the optimal representation from the raw data through consecutive nonlinear transformations, thus achieving increasingly higher levels of abstraction and complexity, as compared to machine learning algorithms [5]. The DL neural networks are inspired by the way the human brain processes information [29], therefore, its architecture is composed of an input layer, two or more hidden layers, and an output layer. The PDD data is fed into the input layer; the abstract features of the PDD data are passed into hidden layers, which process them using activation function. Features or patterns are then fed to the output layer that assigns the observations to classes. The performance of DL algorithms is improved through an iterative process of adjustment of the interconnections (weights and learning rate) between hidden artificial neurons in the hidden layer.

2.2.1.1. Multilayer feedforward networks. Several scholars, including [5, 30] applied multilayer feedforward networks to the classification and modelling of psychiatric disorders. A multilayer feedforward network is a feedforward neural network organised in a layer-wise structure where information propagates from one layer to the other in one direction, from the input layer to the output layer without feedback loop [31,32]. A multilayer feedforward network is also known as multilayer perceptron (MLP) [33]. In a multilayer feedforward network, each input variable is associated with weight values which are inputted in the input layer. Each node in the input layer consists of the artificial neuron, which sums the product of input variables ($x_1, x_2, \ldots, x_n$) and associated weights values ($w_1, w_2, \ldots, w_n$); and the sum of weighted inputs are calculated as in equation (1).

$$Y_i = \sum_{i=1}^{n} x_i \cdot w_i \tag{1}$$

then the summation output of the weighted inputs is then passed through a nonlinear activation function $f(\sum_{i=1}^{n} x_i \cdot w_i)$ o transform the pre-activation level of the neuron to an output y_i as shown in Fig. 1. The weight value determines the strength and direction (inhibitory or excitatory) of each neuron input.

The output of the input layer is then used as the input of the preceding layer, called the hidden layer. The number of hidden layers represents the depth of the network. The output of hidden layers is fed into the output layer, whose purpose is to classify labels in the context of classification. Each neuron is fully connected with all predecessor neurons in the previous layer as well as all neurons in the preceding layer in a feedforward structure, as shown in Fig. 2. Usually, MLPs are trained using backpropagation method, where the error is adjusted using associated weights values (learning rate or momentum value), from the output layer to the input layer. This process is called network training. Therefore, network training continues iteratively using a training set and training algorithm until the error has reached its minimum value.

2.2.1.2. Convolutional neural networks (CNNs). Recent studies have shown that convolutional neural networks are subsequently applied to classifying psychiatric disorders [34–36]. CNNs is a deep neural network architecture initially introduced by LeCun in 1989, designed to process visual imagery [37]. CNN can be thought of as a classifier that extracts and processes hierarchical features for imagery data. Thus, images are given as input labels and training is done automatically. In CNNs, the first layer is the input image. Instead of having input layer and output layer only, CNNs have more additional types of layers called convolutional layer, pooling layer, and fully-connected layer.

The **convolutional layer** is the core module of CNN which is responsible for convolving the input image with learnable filters and extract its features [38]. Every filter is composed of neurons that detect features for the layer inputs. A filter has spatial dimensions of width (W_i), height (H_i), and channel number (D_i). Therefore, convolution means the sum of the element by element multiplication of the neurons in each filter, with the corresponding values at the input. Therefore, a 2-dimensional feature map with parameters such as padding and stride is produced accordingly on convolution with a single filter in each layer. The strides determine the size of the feature map and weights associated with the filter determine the features. Each convolutional layer has n-filters (where n is the number of filters), with each resulting in a feature map. The output of the convolutional layer is the stacked feature maps produced by filters containing different weights, as shown in Fig. 3.

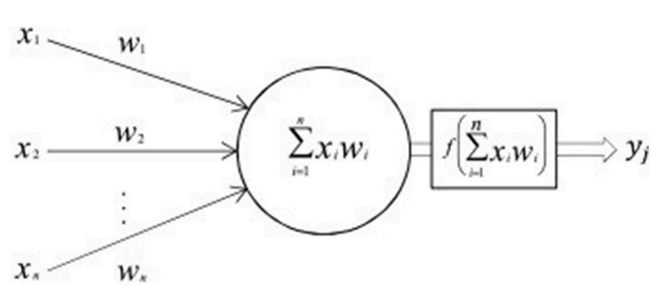

Fig. 1. Structure of an artificial neuron.

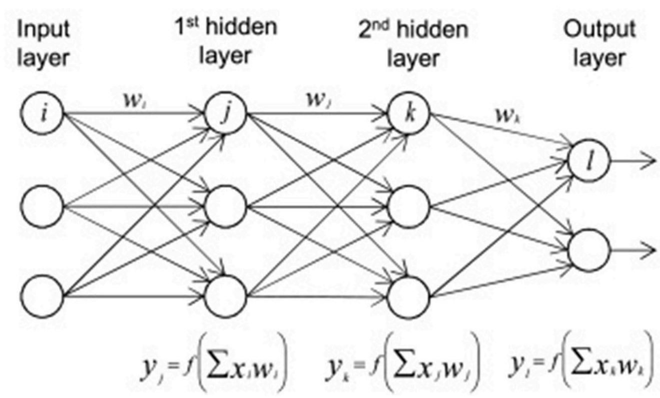

Fig. 2. Multilayer feedforward neural network.

The **pooling layer** is a layer that is periodically added between two successive convolutional layers, in order to reduce redundant representations from the predecessor layers and hence, it controls overfitting [39]. According to Ref. [38], average pooling and max pooling are typical pooling operations of convolutional neural networks. The max-pooling is more suitable when the pooled features are very sparse, whereas average pooling allows these networks to act on different frequencies at each layer while down sampling the images to increase invariance and reduce redundancies [40]. The pooling layer simply reduces the number of neurons of the previous convolutional layer which is located in a small rectangular receptive field.

The **fully-connected layer**, just like hidden neurons in MLP, all neurons have full connections to all activations in the previous layer. Thus, all neurons neither contain spatial information nor do they have a feedback loop. The main goal of the fully-connected layer is to reshape and organise feature vector results from the succeeding convolutional layer and pooling layer.

2.2.1.3. Recurrent neural network (RNN). The recurrent neural network was discovered by John Hopfield in 1982. It was initially used to discover patterns by traversing input labels, both forward and backward, by introducing loops in the network [41]. Recurrent neural networks differ from the multilayer feedforward network architecture in the sense that there is at least one feedback connection that allows the information from past inputs to affect the current output. This means there are connections between neurons form directed cycles or feedback cycles to exhibit dynamic temporal behaviour, especially in the hidden layer. RNNs use internal memory to process inputs, unlike MLP. This is a distinct characteristic of recurrent neural networks. However, RNNs are difficult to train, due to the well-known gradient vanishing and exploding problems and it is hard to learn long-term patterns [42].

2.3. Class imbalance in machine learning techniques

Datasets in real-world problems are typically imbalanced, which may cause some classes to have much more instances than others [43]. This means one class, that is of more interest, either a positive or a minority class, is insufficiently represented. This has significant detrimental effects on training classifiers, such as convergence during the training phase and generalisation of the classifier on the test set [44]. Thus, class imbalance poses several difficulties in training classifiers, including imbalanced class distribution and class overlapping [45]. Therefore, class imbalance should be carefully handled when training a classifier, especially when there is a huge imbalance in the dataset. One of the solutions to counter class imbalance problem is to change class distributions, by resampling [46]. This technique involves under-sampling, over-sampling and advanced sampling. In this paper, we applied the over-sampling method, focusing on synthetic minority oversampling technique (SMOTE), as proposed by Ref. [47]. SMOTE has been used in many fields, including medicine and bio-informatics. It generates synthetic minority examples to over-sample the minority class. SMOTE generates synthetic samples without taking into consideration neighbour examples in machine learning [48]. To address the imbalanced classification problem, SMOTE uses the following equation.

$$D_{new} = D_i + rand * (D_{knn} - D_i) \quad i = 1, 2, 3, \ldots N \tag{2}$$

where D_{new} is the synthetic sample, D_i are minority samples, D_{knn} a sample of k-nearest neighbour from minority samples and *rand* is a random number between 0 and 1 [49].

Modified SMOTE algorithm is implemented as follows.

1 Determine T(i - n) (target variables) and one-hot encode t_i (where $0, 1 \in t_i$)
2 Concatenate all T(i - n), such that $D = concat(t_1, t_2, \ldots t_n)$
3 Determine both D_i (feature vector) and D_{knn} (*k*-nearest neighbour from minority samples).
4 Output the difference between the feature vector and the *k*-nearest neighbour from minority samples.
5 Multiply output by *rand* (a random number between 0 and 1).
6 Add the output to the feature vector D_i to select a new point on the line segment between feature vectors.
7 Repeat steps from 3 to 6 to identify new feature vectors explanation.

We used one-hot encoding on all target variables (where each target

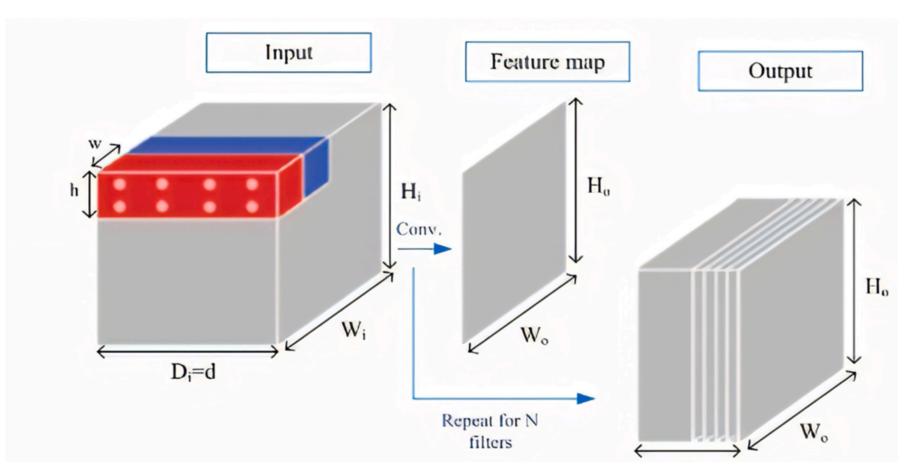

Fig. 3. Convolutional neural network (CNN) architecture [37].

variable is binary). Then, we concatenated the target variables into one variable. We then applied the conventional multiclass SMOTE on the algorithm. Therefore, the study aims to determine the classification performance of the deep learning model on a multiclass dataset of the PDD, based on internal and statistical validation for performance measures and significance test check, respectively. The statistical technique was used to produce statistical significance result that the deep learning model could generate. This study analysed five PDD, Bipolar disorder, Vascular Dementia, Insomnia, Schizophrenia and Attention Deficit Hyperactive Disorder (ADHD) dataset to classify the different types of psychotic diseases for a patient. A patient can be diagnosed with one or more of the diseases. In our previous study [3], set of introductory experiment was implemented focusing on multi-label classification, using the Ensemble Machine Learning. There was a comparison between four multi-label classification algorithms, 15 MLC methods and 10 evaluation measures. In this study, we employed the use of deep neural network and machine learning techniques such as multilayer perceptron (MLP), random forest (RF), decision tree (DT) and support vector machine (SVM) from class imbalance and balanced class perspective. We compared the results of the machine learning techniques with deep learning, with and without class imbalance.

3. Materials and methods

3.1. Data collection

The data were obtained from Yaba Psychiatry Hospital, Yaba, Lagos State, Nigeria by Adejumo et al. [50]. It contained medical records of 500 psychotic patients, 16 variables (11 independent and 5 dependent variables). The information spans a period of five years (Jan. 2010–Dec. 2014). A deep learning neural network was employed to cater for edge-cases that could not be addressed by machine learning algorithms. Machine learning techniques were employed to eliminate the class imbalance in the dataset using the Synthetic Minority Oversampling Technique (SMOTE). The categorical feature vectors from the experiment are transformed into binary using a one-hot-vector encoding technique. The deep learning neural network is designed to be a 3-layer deep architecture as presented in Fig. 4. The activation functions used in the architecture were rectified linear units (RELU) and sigmoid for the connected layers and output layer respectively. The loss function for training is a binary cross-entropy and evaluation metrics accuracy. The first layer in Fig. 4a and Fig. 4b are the input layers, where the input data is read into the network model. The fifth layers are the output layer used for the classification (single and multi-label). The second, third, and fourth layers are called hidden layers. The hidden layers are the layers between the input layers and the output layers. The number of hidden layers represents the depth of the network.

3.2. Parameters for deep learning

In deep learning, settings of parameters are one of the crucial tasks to be performed. The parameters considered in deep and machine learning are model parameters and hyper-parameters. Model parameters concerned with the internal parameters used by the network, such as the weights, neurons, while Hyper-parameters deal with external parameter settings such as the learning rate, momentum, epochs, activation function, dropout, number of hidden layers, optimizer, to mention a few. Hyper-parameters are used to decide the structure, functions, accuracy, and validity of the network. The commonly used methods to determine the optimal settings of fine-tuning hyper parameters are trial and error, grid search (brute force), Bayesian optimisation, and random search [6, 51]. The grid-based search approach was used in this study, due to its simplicity to implement. The following are the hyper-parameters used in this study.

Hidden layers: The more the hidden layers of neurons, the better the accuracy. Three hidden layers were used in this study.

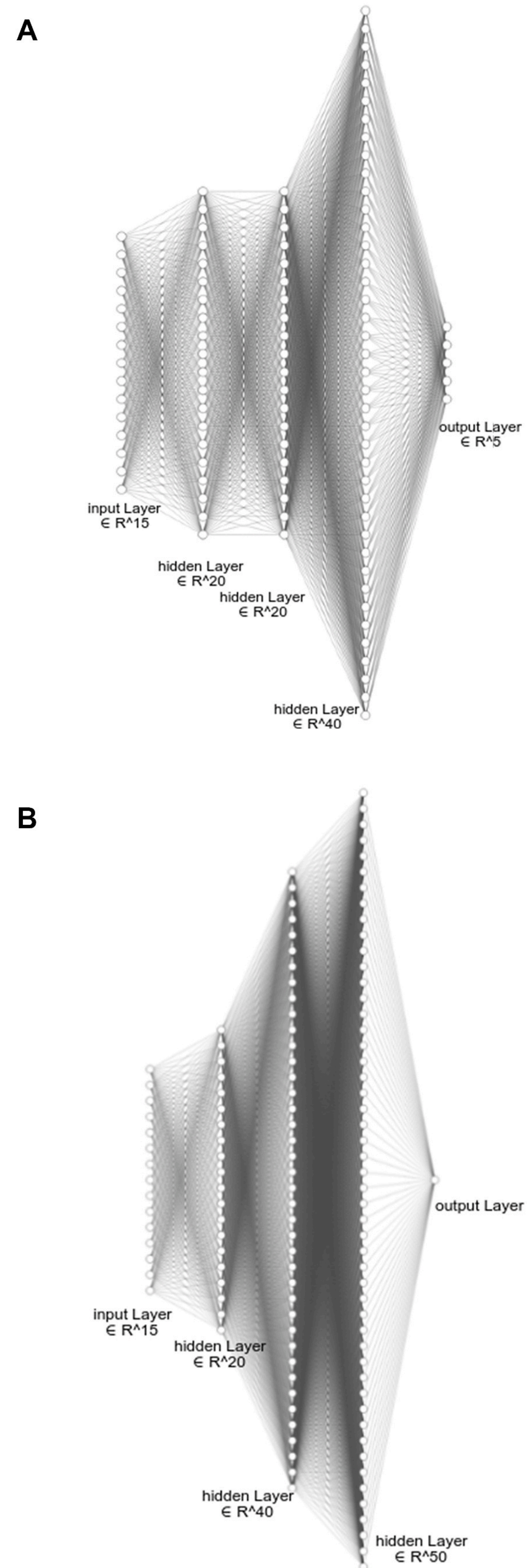

Fig. 4. Deep learning network classification architecture.

Drop-out: This is one of the examples of regularisation methods used to prevent overfitting. Drop-out is chosen over L1 and L2 regularisation in this study because it does not rely on the modification of the cost function, rather it modifies the network.

Activation functions: These are mathematical equations mostly used to determine which function is to process the inputs that are fed into each neuron. Activation function can also be called the Transfer function. It is considered in this study to help the network to converge and learn faster during training. There are two types of activation functions and they are linear and non-linear activation functions. The most commonly used is the non-linear activation function.

The non-linear activation functions commonly used are as follows.

1. Sigmoid: The output values of this function as represented in equation (3) is between 0 and 1. The sigmoid function is differentiable but it does not allow jumping in the output values and the gradient is smooth. The sigmoid function can also be referred to as the logistic function. They can be used for feedforward neural network in deep and machine learning [52].

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

2. Hyperbolic tangent: The output values of this function as represented in equation (4) is between −1 and 1. Its major problem is that the gradient disappears. It is zero-centric in nature.

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{4}$$

3. Rectifier Linear Unit (ReLU): The output values of this function as represented in equation (5) is between 0 and infinity. It is computationally efficient and solves the problem of gradient that disappears in hyperbolic tangent function. This function can be used for the hidden and output layer in deep learning. Its main problem is the conversion of negative values to zero. Its main advantage over sigmoid function is that it can handle backpropagation. Equation (5) shows the mathematical equation of ReLU.

$$R(x) = max(0, x) \tag{5}$$

$if\ x < 0\ then\ R(x) = 0$ and if $x \geq 0\ then\ R(x) = x$

4. *Softmax*: The output values of the function are between 0 and 1 like the sigmoid function. It can only be used for the output layer in deep learning.

Sigmoid and ReLU is the activation function used in this study.

Weight Initialization: The initial weights must be set for the first forward pass of a network. This can be achieved by setting the weights to zero or by randomizing the weights.

3.3. Experiments

The deep learning model was trained using Keras functional API, running on top of TensorFlow in Google Colaboratory online platform with Python 3.6 notebook. The architectural setup presented in Fig. 4a and b are 3-layer deep fully-connected network with the RELU activation function and architectural layer are 15–20–20-40–1 and 15-20–20-40–5 respectively. The training data is split into a 30% validation set running for 40 epochs with an early stop monitor on validation loss. The optimizer is Adam with a learning rate of 0.01. The initial deep learning network was used to jointly classify all the five target variables from the trained model as shown in Fig. 4b. Another network was also designed to classify each element of the five target variables, as shown in Fig. 4a.

The study was implemented on Google Colaboratory notebook with Python 3 runtime. The deep learning training and validation was done using the run-time engine with the following configurations.

(continued on next column)

(continued)

Configurations for deep learning training.	
Operating System	"Ubuntu 18.04.3 LTS"
CPU	
Architecture:	x86_64
CPU op-mode(s):	32-bit, 64-bit
Byte Order:	Little Endian
CPU(s):	2
Model name:	Intel(R) Xeon(R) CPU @ 2.30 GHz
RAM:	13 G
DISK:	73 G

3.3.1. Machine learning techniques

The machine learning techniques model was run on Python 3.7 on Windows 10 operating system (win10) and RAM 32 GB. The libraries include pandas, numpy, matplotlib and sklearn.

3.4. Deep learning performance evaluation measure

The DNN is an example of deep learning used in this study and the performance metrics used are accuracy, recall, precision, and F-score. The confusion matrix also called contingency table is a combination of rows and columns in a tabular form in four quadrants for presenting the classification results of the classifier [37]. It assists in the computation of the performance evaluation metrics after making the formula available.

True Positives (TP): means patients are correctly classified that they are sufferings from PDDTrue Negatives (TN): means patients are correctly classified that they are not suffering from PDDFalse Positives (FP): means patients are incorrectly classified that they are sufferings from PDDFalse Negatives (FN): means patients are incorrectly classified that they are not suffering from PDD.

(a) Accuracy is the sum of number of true positives and true negatives (number of correct predictions) divided by the total number of predictions made. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
(b) Precision can also be called Positive predictive value (PPV) and it is the number of true positives divided by the sum of true positives and false positives number. $Precision = \frac{TP}{TP+FP}$
(c) Recall is the number of True Positives divided by sum of the number of True Positives and False Negatives. Another name for Recall is sensitivity or True Positive Rate. $Recall = \frac{TP}{TP+FN}$
(d) F-Score is the combination of the features of recall (sensitivity) and precision. $Fscore = \frac{2*recall*precision}{recall+precision}$

4. Results

The multi-label classification results for the loss and accuracy obtained while training the deep learning model are presented in Fig. 5 and Fig. 6 respectively. A portion of the training data (30%) is held back to validate the performance of the model during training. The result shows that the model could not learn new information after 36 iterations, by tracking the changes in the validation loss.

The dataset is split into 70% training data and 30% validation dataset. The results illustrate the comparison between the model and validation accuracy. The evaluation for single-label classification is performed by keeping the feature variables the same, but changing the target to represent a symptom. The accuracy results of the classification models with imbalanced and balanced dataset are presented in Table 1 and Table 2, respectively. From the result, the trained model using

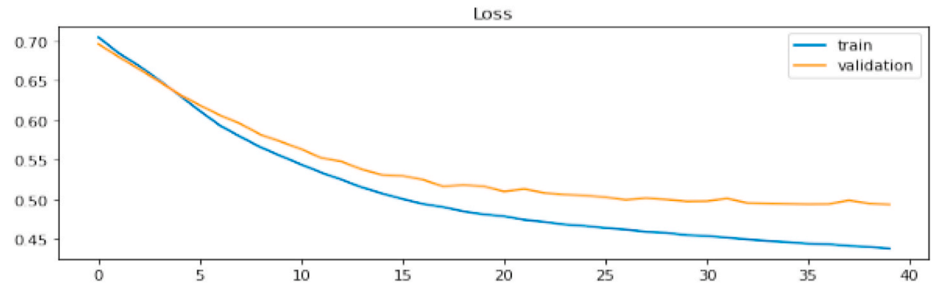

Fig. 5. Loss of multi-label classification.

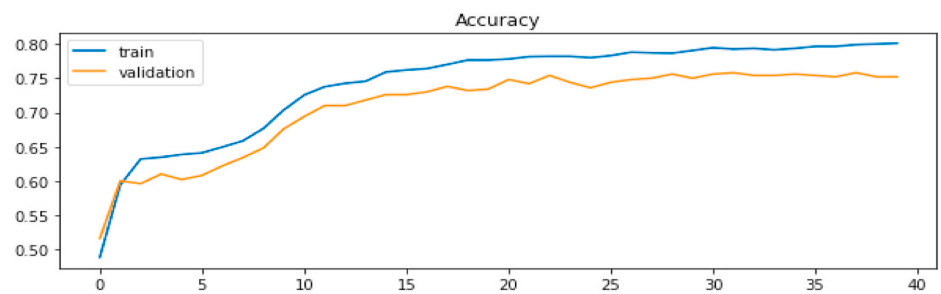

Fig. 6. Accuracy of multi-label classification.

Table 1
Accuracy Results of Single and Multi-Label Classification Model on dataset with class imbalance.

Deep Learning	Model (%)	Validation (%)
Multi-output (Multi label)	0.7786	0.7517
Insomnia	0.7929	0.7417
Schizophrenia	0.9250	0.9000
Vascular dementia	0.8536	0.8010
ADHD	0.7786	0.6500
Bipolar disorder	0.8143	0.7667

Table 2
Accuracy Results of Single and Multi-Label Classification Model on dataset without class imbalance.

Deep Learning	Model (%)	Validation (%)
Multi-output (Multi label)	0.5823	0.5587
Insomnia	0.6526	0.5137
Schizophrenia	0.7015	0.6627
Vascular dementia	0.6358	0.5451
ADHD	0.7051	0.6078
Bipolar disorder	0.6863	0.5294

schizophrenia as a target, produced the best performance of 90% on the validation dataset with class imbalance. However, using ADHD as a target yielded an accuracy of 77.86% and it also produced the worst result of 65% classification accuracy on test data. In this section, we present the results from the experiments carried out in this study (see Table 1 and Table 2).

4.1. ROC curve for PDD classification and Random Forest Confusion Matrix

To validate the results obtained from accuracy, we investigated the model further, using receiver operating characteristics (ROC) and confusion matrix. The ROC curve for the single-label deep learning classification is presented in Fig. 7. It shows the relationship between the rate of true positives and false positives. It is worthy of note that the model for schizophrenia yields better performance with more area under the curve (AUC) of 93%. This indicates that the model has more true positives and fewer false positives. On the other hand, the worst performance is exhibited by the model for bipolar disorder with 71% AUC.

For the confusion matrix, all the values in the diagonal top left to bottom right are correctly classified data samples, as presented in Fig. 8. Now, in our validation set, if you get the sum across columns (left to right), you would get the total samples present in that class. For example, in row one, total 0 samples is $23 + 1 = 24$. This means in the actual validation set, we have 24 samples belonging to the "0" class, of which our model predicted 23 correctly (95.8%). For observe class "11″, the model didn't perform well; it showed 28.6% accuracy by classifying only 6 samples correctly and misclassifying most data samples into different classes. Here, the model is overfitted on this class. This shows the data samples in the training belonging to class "11″ is not enough.

4.2. Discussion of results

We discovered from the statistical analysis reported by Ref. [50] that 40.2% tested positive to bipolar disorder, 40.6% to insomnia, 75% to schizophrenia, 43.6% to ADHD and 69.2% to vascular dementia. This implies that there is class imbalance in the dataset that needs to be balanced, in order to avoid bias in the classification accuracy, a problem which was not also addressed by Ref. [3]. In order to deal with the data imbalance in this study, we generated synthetic samples from existing samples using Synthetic Minority Oversampling Technique (SMOTE). We had a challenge at first. The problem was a multi-label classification task and not the conventional multi-classification problem, where there is one target variable but more than 2 classes. Thus, the multiclass SMOTE approach would not work directly. We thus converted our multi-label, firstly, to a multi-class problem. Our target variables are in this order: Insomnia, schizophrenia, vascular dementia, ADHD and Bipolar disorder. Each column contains either Positive or Negative value. The Negative or N values are encoded as 0 and the Positive or P values are encoded as 1. The top five data samples are thus encoded as:

Insomnia	schizophrenia	vascular dementia	ADHD	bipolar disorder
0	1	1	1	0
1	1	1	0	1

(continued on next page)

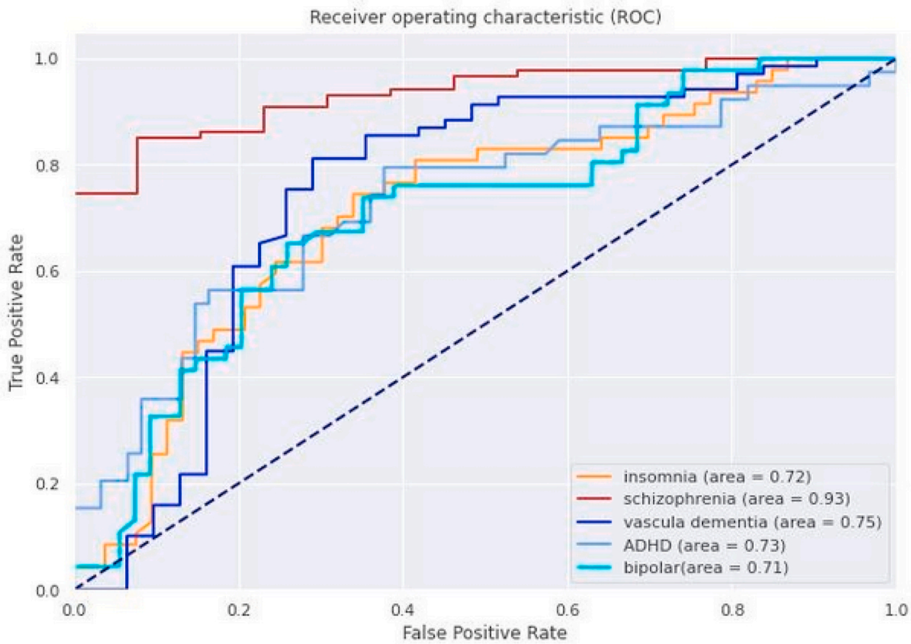

Fig. 7. ROC curve for classification of PDD.

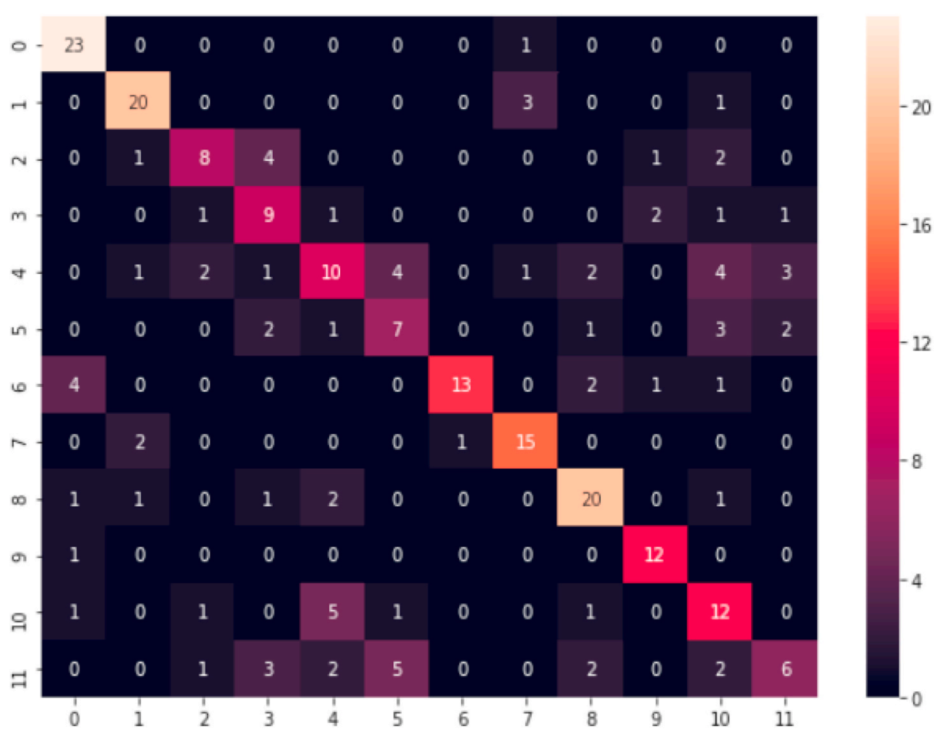

Fig. 8. Random forest confusion matrix.

(*continued*)

1	1	0	0	1
1	1	1	1	1
0	1	1	0	0

The entire target variables which look like the above five are then combined together. The combination is shown in the target column in the data frame.

Insomnia	schizophrenia	vascular dementia	ADHD	bipolar disorder	**target**
0	1	1	1	0	01110
1	1	1	0	1	11101
1	1	0	0	1	11001
1	1	1	1	1	11111
0	1	1	0	0	01100

The target column now forms our multiclass variable. The values in this column were found to contain 19 different combinations. When the distribution was checked, we found that some of the combinations occurred only once, some two times, some 5 times and some more than 10 times. The different combinations found are 01110, 11100, 11101, 11001, 11111, 01100, 01000, 01111, 10001, 10011, 00000, 10111, 11011, 00100, 00110, 01010, 10101, 00010 and 10100. In the SMOTE recommendation, it is recommended that for good performance, we have at least 6 neighbours. As such, we removed all combinations with less than 6 occurrences. The removed combinations are 11100, 01111, 10011, 00110, 00010, 10100, 10111. After removing combinations less than 6 (which were 7 in number), the following are the remaining combinations 01110, 11101, 11001, 11111, 01100, 01000, 10001, 00000, 11011, 00100, 01010, 10101. We then apply SMOTE on the data, such that every sample had a total of 101 samples each. We further run an analysis on the dataset to compute the feature importance, and the plot is as shown in Fig. 9. Age is the top feature contributing to the classifications of the PDD as derived from the random forest model, followed by occupation and divorce is the least as shown in Fig. 9. We further train a Multilayer Perceptron (MLP), Support Vector Machines (SVM), Random Forest (RF) and Decision Tree (DT) on the training data (which is 80% of the whole data). The four algorithms are evaluated on the test set (which is 20% of the sampled data). The accuracy and balanced accuracy are as presented in Table 3.

Our results on machine learning techniques were compared with [3, 50] in terms of accuracy, balanced accuracy and correlation matrix on attributes influencing PDD. Our proposed study performed better with MLP, compared to SVM, RF and decision tree, based on accuracy, while the RF outperform SVM, MLP and DT are based on balanced accuracy, as presented in Table 3. The MLP and RF also outperformed the best ensemble machine learning employed in Ref. [3], based on accuracy and balanced accuracy, respectively, as presented in Table 3.

4.3. Evaluation of techniques based on accuracy and balanced accuracy

Different validations accuracy were obtained as presented in Tables 4 and 5 based on optimal accuracy on the dataset, with class imbalance and optimally balanced accuracy on the dataset with a balanced class from different studies. This was achieved using different techniques on the same dataset. The proposed deep neural network-multilayer perceptron (DNN-MLP) and machine learning - multilayer perceptron (ML-MLP) outperformed the other approaches in Table 4.

The proposed and research carried using a deep neural network and machine learning with RF with balanced class are as shown in Table 5. After balancing the dataset, we discovered that the Machine learning-random forest (ML-RF) outperform the Machine learning-multilayer perceptron (ML-MLP) while ML-RF classification performance is less than deep neural network.

Table 3
Comparison of machine learning techniques results.

Algorithm	Accuracy (%)	Balanced Accuracy
MLP	0.5843621399176955	0.5853072853072853
SVM	0.4691358024691358	0.4982219169719169
RF	0.4691358024691358	0.6406784188034188
DT	0.4691358024691358	0.4982219169719169

Table 4
Accuracy on multi-label classification with class imbalance.

Techniques	Optimal Accuracy (%)	Author
LC-Naïve Bayes tree (Multi-label)	56.56	[3]
LC-Naïve bayes (Multi-label)	56.56	[3]
Proposed (Deep learning)	**75.17**	this study
Proposed (ML-MLP)	**58.44**	this study

Table 5
Balanced accuracy on multi-label classification without class imbalance.

Techniques	Optimal Accuracy (%)	Author
Proposed (Deep learning)	55.87	this study
Proposed (ML-RF)	**64.07**	this study

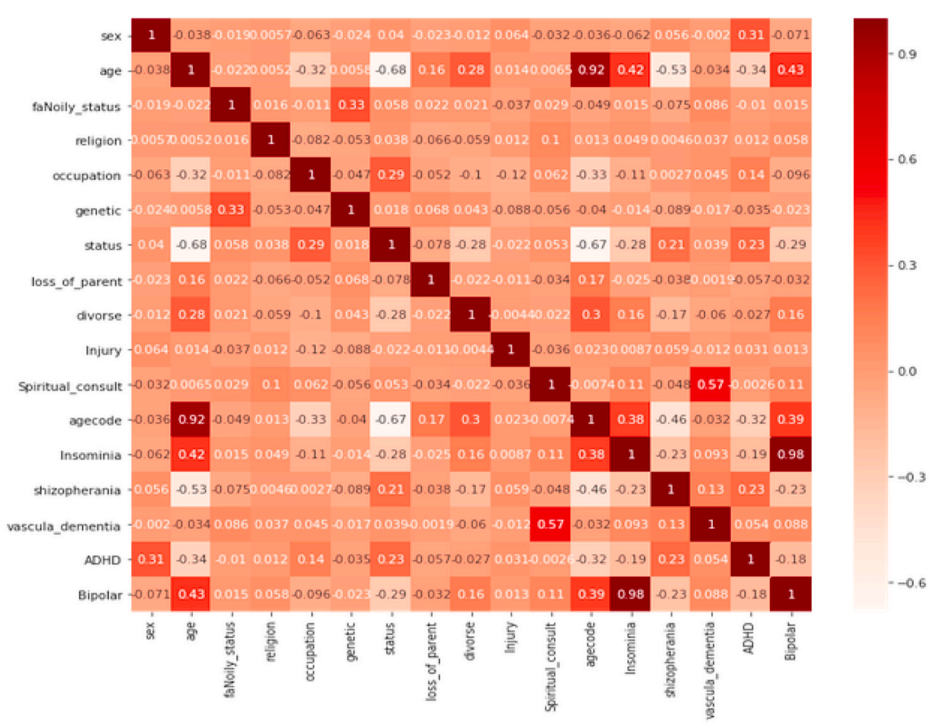

Fig. 9. Correlation matrix on features that influenced PDD.

4.4. Evaluation of techniques based on features influencing PDD

The features that influence the PDD by Ref. [50] and the proposed study are based on a statistical approach, using χ^2 statistics and machine learning approach, using a pairwise correlation function are as presented in Table 6.

In this study, the result shows that there is a high tendency for a patient with Bipolar disorder to have insomnia. The claim is supported by Ref. [53] because the features influencing them are almost the same, except for occupation, in Bipolar disorder. Our findings in Table 6 is supported by Refs. [50,53]. Bipolar disorder and insomnia are strongly correlated with R-value of 0.98 as shown in Fig. 10. Bipolar disorder and insomnia are most prevalent in old adults also. Fig. 10 also shows that marital stress can also predispose married couples to ADHD, while insomnia and bipolar are higher in single patients.

4.5. Is deep learning better than machine learning on PDD?

In spite of the accomplishment of the deep learning technique, its ascendency cannot be demonstrated in all instances in real-life scenarios. Most researchers have concluded that deep learning technique will always outperform machine learning technique [54,55]. This is not always the case in all circumstances as established in this study. Deep learning outperforms machine learning, considering a dataset that contains both imaging and non-imaging raw data as established in some systematic literature review [5]. We established that deep learning outperforms machine learning on PDD dataset with class imbalance from multi-classification accuracy perspective whereas on PDD dataset with balanced class, machine learning outperforms the deep learning. Many studies also supported the fact that machine learning will outperform deep learning on the dataset with a small sample size as corroborated in this study. The deep neural network algorithm performance in this study on the psychotic disorder dataset, using two split (train and validate) produced good results which when used for three splits (train, validate and test) of the same algorithm produced poor results because the sample size was small. Secondly, the accuracy of the multi-classification, based on the deep learning-multilayer perceptron algorithm was high on the class imbalance dataset due to bias towards the majority class, but it was low on the balanced class.

5. Conclusion

Diagnosing mental illness is increasingly becoming more complex because of confusing symptoms. Also, some patients do not clearly articulate their mental health state and diseases' symptoms, especially PDD. A proper diagnosing tool is necessary to assist medical practitioners to distinguish and properly diagnose psychotic disorders. Therefore, this study presented PDD as a multi-label classification problem to investigate the use of deep neural network and machine learning architectures and techniques - multilayer perceptron (MLP), support vector machine (SVM), random forest (RF) and Decision tree (DT). This was done for the purpose of identifying hidden patterns in patients' data and for classifying Psychotic Disorder Diseases like Bipolar disorder, Vascular Dementia, Insomnia, Schizophrenia and Attention Deficit Hyperactive Disorder (ADHD) in patients, for the use of clinicians. The architectures are treasured tools as correlations in the data, through which iterative optimisation techniques can be concluded. The outcomes obtained showed that Deep Learning Neural Networks give good classification performance results, based on accuracy, true positive rate and false positive rate and AUC, compared to the machine learning approach adopted by Ref. [56]. The proposed approaches in this study and the use of ensemble machine learning on the same dataset in the previous studies are both pointing to schizophrenia as the one with the best performance in terms of accuracy, and to ADHD as the least performance. The limitation is that the use of temporal validation performs badly, compared to the train/test split both from deep learning perspective on the same dataset. These results show that applying the deep learning model to PDD can derive patient representations that offer improved clinical predictions and augment machine learning framework for making clinical decisions. The deep neural network algorithm performance on the psychotic disorder dataset, using two split (train and validate) produced strongly good results which when used for three splits (train, validate and test) of the same algorithm produced poor results because the sample size is small.

The result obtained revealed that deep neural network gave a superior performance of 75.17% with class imbalance accuracy while the MLP model accuracy is 58.44%. Conversely, the best performance in the machine learning techniques are exhibited by the random forest model, using the dataset without class imbalance and its result, compared with deep learning techniques is 64.1% and 55.87%, respectively. It was also observed that patients' age is the most contributing feature to the performance of the model while divorce is the least. Likewise, the study reveals that there is a high tendency for a patient with bipolar disorder to have insomnia; these diseases are strongly correlated with an R-value of 0.98. Bipolar disorder and insomnia are most prevalent in old adults also. Our concluding remark shows that applying the deep learning model to PDD data not only offers improved clinical classification of the diseases but also provides a framework for augmenting clinical decision systems, by eliminating the class imbalance and unravelling the attributes that influence PDD in patients. This study also supports other researchers to be assertive that deep learning does not outperform machine learning techniques in all real-life scenarios [57]. In the future,

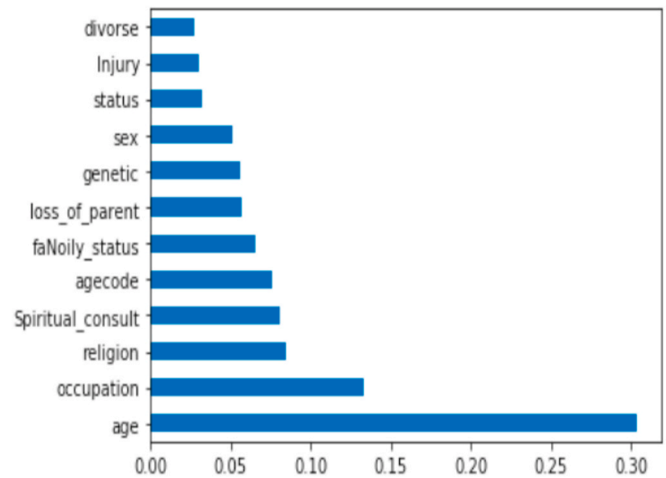

Fig. 10. Feature importance in PDD.

Table 6
Attributes influencing PDD.

Techniques	Features	Author
Insomnia	age, occupation, marital status, divorce, spiritual consult	This study
Schizophrenia	age, marital status, divorce	This study
Vascular dementia	spiritual consult	This study
ADHD	sex, age, occupation, marital status	This study
Bipolar disorder	age, marital status, divorce, spiritual consult	This study
Insomnia	age, occupation, status, divorce, spiritual consult	[50]
Schizophrenia	age, occupation, religion, status, hereditary, divorce	[50]
Vascular dementia	history, spiritual consult	[50]
ADHD	sex, age, occupation, religion, marital status	[50]
Bipolar disorder	age, occupation, marital status, divorce, spiritual consult	[50]

the authors intend to develop a soft-computing model that can handle early diagnosis of the multi-classification psychotic diseases analysis matrix confusion.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to appreciate the Yaba Psychiatry hospital, Yaba, Lagos state, Nigeria team and Adejumo, A. O., Ikoba, N. A., Suleiman, E. A., Okagbue, H. I., Oguntunde, P. E., Odetunmibi, O. A., Job, O for making the dataset available for research purpose.

References

[1] Lee Anna Clark, Cuthbert Bruce, Lewis-Fernández Roberto, Narrow William E, Reed Geoffrey M. Three approaches to understanding and classifying mental disorder: Icd-11, dsm-5, and the national institute of mental health's research domain criteria (rdoc). Psychol Sci Publ Interest 2017;18(2):72–145.
[2] Bzdok Danilo, Meyer-Lindenberg Andreas. Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatr: Cognit Neurosci Neuroimag 2018;3(3):223–30.
[3] Folorunso SO, Fashoto SG, Olaomi J, Fashoto OY. A multi-label learning model for psychotic diseases in Nigeria. Inf Med Unlocked 2020;19:100326.
[4] Rezaii Neguine, Walker Elaine, Wolff Phillip. A machine learning approach to predicting psychosis using semantic density and latent content analysis. npj Schizophrenia 2019;5(1):1–12.
[5] Vieira Sandra, Pinaya Walter HL, Mechelli Andrea. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci Biobehav Rev 2017;74:58–75.
[6] de Filippis Renato, Carbone Elvira Anna, Gaetano Raffaele, Bruni Antonella, Pugliese Valentina, Segura-Garcia Cristina, De Fazio Pasquale. Machine learning techniques in a structural and functional mri diagnostic approach in schizophrenia: a systematic review. Neuropsychiatric Dis Treat 2019;15:1605.
[7] Stein Dan J, Benjet Corina, Gureje Oye, Lund Crick, Scott Kate M, Poznyak Vladimir, van Ommeren Mark. Integrating mental health with other non-communicable diseases. Bmj 2019;364:l295.
[8] Mikolas Pavol, Hlinka Jaroslav, Skoch Antonin, Pitra Zbynek, Frodl Thomas, Filip Spaniel, Hajek Tomas. Machine learning classification of first-episode schizophrenia spectrum disorders and controls using whole brain white matter fractional anisotropy. BMC Psychiatr 2018;18(1):97.
[9] Stamate Daniel, Katrinecz Andrea, Stahl Daniel, Simone JW Verhagen, Delespaul Philippe AEG, Jim van Os, Guloksuz Sinan. Identifying psychosis spectrum disorder from experience sampling data using machine learning approaches. Schizophr Res 2019;209:156–63.
[10] Vieira Sandra, Gong Qi-yong, Walter HL Pinaya, Scarpazza Cristina, Tognin Stefania, Crespo-Facorro Benedicto, Tordesillas-Gutierrez Diana, Ortiz-García Victor, Setien-Suero Esther, Scheepers Floortje E, et al. Using machine learning and structural neuroimaging to detect first episode psychosis: reconsidering the evidence. Schizophr Bull 2020;46(1):17–26.
[11] Larson Molly K, Walker Elaine F, Compton Michael T. Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders. Expert Rev Neurother 2010;10(8):1347–59.
[12] Botha K, Moletsane M. Western and african aetiological models. In: Abnormal psychology. A South African perspective; 2012. p. 80–99.
[13] Morel Didier, Yu Kalvin C, Liu-Ferrara Ann, Caceres-Suriel Ambiorix J, Kurtz Stephan G, Tabak Ying P. Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. Int J Med Inf 2020:104136.
[14] Scarpazza Cristina, Baecker Lea, Vieira Sandra, Mechelli Andrea. Applications of machine learning to brain disorders. In: Machine learning. Elsevier; 2020. p. 45–65.
[15] Agbu Jane-Frances. "Language and thoughts". In: Agiobu Kemmer I, editor. Essentials of psychology. Springfield Books; 2004.
[16] Grunze Heinz. Bipolar disorder. In: Neurobiology of brain disorders. Elsevier; 2015. p. 655–73.
[17] Anastopoulos Arthur D, Guevremont David C, Shelton Terri L, George J DuPaul. Parenting stress among families of children with attention deficit hyperactivity disorder. J Abnorm Child Psychol 1992;20(5):503–20.
[18] Green Christopher, Chee Kit. Understanding adhd. *A paren's guide to attention deficit hyperactivity disorder in children*. London: Vermilion; 1997.
[19] Joseph Biederman, Mick Eric, Faraone Stephen V. Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type. Am J Psychiatr 2000;157(5):816–8.
[20] Sartipi Shadi, Kalbkhani Hashem, Ghasemzadeh Peyman, Shayesteh Mahrokh G. Stockwell transform of time-series of fmri data for diagnoses of attention deficit hyperactive disorder. Appl Soft Comput 2020;86:105905.
[21] Saha Sukanta, Chant David, Joy Welham, McGrath John. A systematic review of the prevalence of schizophrenia. PLoS Med 2005;2(5).
[22] Iadecola Costantino. The pathobiology of vascular dementia. Neuron 2013;80(4): 844–66.
[23] Hublin CG, Partinen Markku M. The extent and impact of insomnia as a public health problem. J Clin Psychiatry Prim Care Companion 2002;4(8–12).
[24] Tsoumakas Grigorios, Katakis Ioannis. Multi-label classification: an overview. Int J Data Warehous Min 2007;3(3):1–13.
[25] Miotto Riccardo, Li Li, Kidd Brian A, T Dudley Joel. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016;6(1):1–10.
[26] Nunes Luciano Comin, Rogério Pinheiro Plácido, Pequeno Cavalcante Tarcísio, Dantas Pinheiro Mirian Calíope. Handling diagnosis of schizophrenia by a hybrid method. Comput Math. Methods Med. 2015;2015:13. https://doi.org/10.1155/2015/987298. 987298.
[27] Kalanderian Hripsime, Nasrallah Henry A. Artificial intelligence in psychiatry. Curr Psychiatr 2019;18(8):33–8.
[28] Elliot Mbunge, Makuyana Ralph, Chirara Nation, Antony Chingosho. Fraud detection in e-transactions using deep neural networks-a case of financial institutions in Zimbabwe. Int J Sci Res 2017;6(9):1036–40.
[29] Chu Lei, Qiu Robert, Liu Haichun, Ling Zenan, Zhang Tianhong, Wang Jijun. Individual recognition in schizophrenia using deep learning methods with random forest and voting classifiers: insights from resting state eeg streams. 2017. arXiv preprint arXiv:1707.03467.
[30] Bashyal Shishir. Classification of psychiatric disorders using artificial neural network. In: International symposium on neural networks. Springer; 2005. p. 796–800.
[31] Gbenga Fashoto Stephen, Akinnuwesi Boluwaji, Owolabi Olumide, Adelekan David. Decision support model for supplier selection in healthcare service delivery using analytical hierarchy process and artificial neural network. Afr J Bus Manag 2016;10(9):209.
[32] Gnana Sheela K, Deepa Subramaniam N. Review on methods to fix number of hidden neurons in neural networks. Math Probl Eng 2013:2013.
[33] Gardner Matt W, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ 1998;32(14–15):2627–36.
[34] Campese Stefano, Lauriola Ivano, Scarpazza Cristina, Sartori Giuseppe, Aiolli Fabio. Psychiatric disorders classification with 3d convolutional neural networks. In: INNS big data and deep learning conference. Springer; 2019. p. 48–57.
[35] Durstewitz Daniel. Georgia Koppe, and Andreas Meyer-Lindenberg. Deep neural networks in psychiatry. Mol Psychiatr 2019;24(11):1583–98.
[36] Lanillos Pablo, Oliva Daniel, Philippsen Anja, Yamashita Yuichi, Nagai Yukie, Cheng Gordon. A review on neural network models of schizophrenia and autism spectrum disorder. Neural Network 2020;122:338–63.
[37] Balas Valentina Emilia, Roy Sanjiban Sekhar, Sharma Dharmendra, Samui Pijush. Handbook of deep learning applications, vol 136. Springer; 2019.
[38] Jing Hua, Zhong Zichun. Spectral geometry of shapes. Academic Press; 2020.
[39] Gu Jiuxiang, Wang Zhenhua, Kuen Jason, Ma Lianyang, Shahroudy Amir, Shuai Bing, Liu Ting, Wang Xingxing, Wang Gang, Cai Jianfei, et al. Recent advances in convolutional neural networks. Pattern Recogn 2018;77:354–77.
[40] Hadji Isma, Wildes Richard P. What do we understand about convolutional networks?. 2018. arXiv preprint arXiv:1803.08834.
[41] Miikkulainen Risto, Liang Jason, Meyerson Elliot, Rawal Aditya, Fink Daniel, Olivier Francon, Raju Bala, Shahrzad Hormoz, Navruzyan Arshak, Duffy Nigel, et al. Evolving deep neural networks. In: Artificial intelligence in the age of neural networks and brain computing. Elsevier; 2019. p. 293–312.
[42] Li Shuai, Li Wanqing, Cook Chris, Zhu Ce, Gao Yanbo. Independently recurrent neural network (indrnn): building a longer and deeper rnn. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 5457–66.
[43] Buda Mateusz, Maki Atsuto, Mazurowski Maciej A. A systematic study of the class imbalance problem in convolutional neural networks. Neural Network 2018;106: 249–59.
[44] Liu Xu-Ying, Wu Jianxin, Zhou Zhi-Hua. Exploratory undersampling for class-imbalance learning. IEEE Trans Syst Man Cybern B Cybern 2008;39(2):539–50.
[45] Ali Aida, Shamsuddin Siti Mariyam, Ralescu Anca L, et al. Classification with class imbalance problem: a review. Int J Adv Soft Comput Appl 2015;7(3):176–204.
[46] Guo Xinjian, Yin Yilong, Dong Cailing, Yang Gongping, Zhou Guangtong. On the class imbalance problem. In: 2008 Fourth international conference on natural computation, vol. 4. IEEE; 2008. p. 192–201.
[47] Nitesh V Chawla, Bowyer Kevin W, Hall Lawrence O, Kegelmeyer W Philip. Smote: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–57.
[48] Pruengkarn Ratchakoon, Wong Kok Wai, Fung Chun Che. Imbalanced data classification using complementary fuzzy support vector machine techniques and smote. In: 2017 IEEE international conference on systems, man, and cybernetics (SMC). IEEE; 2017. p. 978–83.
[49] Abdelrahman I Saad, Yasser MK Omar, Maghraby Fahima A. Predicting drug interaction with adenosine receptors using machine learning and smote techniques. IEEE Access 2019;7:146953–63.
[50] O Adejumo Adebowale, Nehemiah A Ikoba, Suleiman Esivue A, Okagbue Hilary I, Oguntunde Pelumi E, Odetunmibi Oluwole A, Job Obalowu. Quantitative exploration of factors influencing psychotic disorder ailments in Nigeria. Data in Brief 2017;14:175–85.

[51] James Bergstra, Bengio Yoshua. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13(Feb):281–305.
[52] Vabalas Andrius, Gowen Emma, Poliakoff Ellen, Alexander J Casson. Machine learning algorithm validation with a limited sample size. PloS One 2019;14(11):e0224365.
[53] Kaplan Katherine A, Harvey Allison G. Behavioral treatment of insomnia in bipolar disorder. Am J Psychiatr 2013;170(7):716–20.
[54] Sergey M Plis, Devon R Hjelm, Salakhutdinov Ruslan, Allen Elena A, Bockholt Henry J, Long Jeffrey D, Johnson Hans J, Paulsen Jane S, Turner Jessica A, Calhoun Vince D. Deep learning for neuroimaging: a validation study. Front Neurosci 2014;8:229.
[55] Han Xiaobing, Zhong Yanfei, He Lifang, Philip S Yu, Zhang Liangpei. The unsupervised hierarchical convolutional sparse auto-encoder for neuroimaging data classification. In: International conference on brain informatics and health. Springer; 2015. p. 156–66.
[56] Akinnuwesi Boluwaji A, Fashoto Stephen G, Andile S Metfula, Akinnuwesi Adetutu N. Experimental application of machine learning on financial inclusion data for governance in eswatini. In: Conference on e-Business, e-Services and e-Society. Springer; 2020. p. 414–25.
[57] Stephen Fashoto, Elliot Mbunge, Gabriel Ogunleye, Van den Burg Johan. Implementation of machine learning for predicting maize crop yields using multiple linear regression and backward elimination. Malay J Comput 2021;6(1):679–97.