




Article

Automated Risk Prediction of Post-Stroke Adverse Mental Outcomes Using Deep Learning Methods and Sequential Data

Chien Wei Oei ^{1,2}, Eddie Yin Kwee Ng ^{2,*} , Matthew Hok Shan Ng ³, Yam Meng Chan ⁴ ,
Vinithasree Subbhuraam ⁵ , Lai Gwen Chan ^{6,7} and U. Rajendra Acharya ⁸

¹ Management Information Department, Office of Clinical Epidemiology, Analytics and kNowledge (OCEAN), Tan Tock Seng Hospital, Singapore 308433, Singapore; chienwei001@e.ntu.edu.sg

² School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore

³ Rehabilitation Research Institute of Singapore, Nanyang Technological University, Singapore 308232, Singapore

⁴ Department of General Surgery, Vascular Surgery Service, Tan Tock Seng Hospital, Singapore 308433, Singapore

⁵ The Digital Health Hub, Austin, TX 78944, USA

⁶ Department of Psychiatry, Tan Tock Seng Hospital, Singapore 308433, Singapore

⁷ Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore

⁸ School of Mathematics, Physics and Computing, University of Southern Queensland, Brisbane, QLD 4305, Australia

* Correspondence: mykng@ntu.edu.sg

Abstract: Depression and anxiety are common comorbidities of stroke. Research has shown that about 30% of stroke survivors develop depression and about 20% develop anxiety. Stroke survivors with such adverse mental outcomes are often attributed to poorer health outcomes, such as higher mortality rates. The objective of this study is to use deep learning (DL) methods to predict the risk of a stroke survivor experiencing post-stroke depression and/or post-stroke anxiety, which is collectively known as post-stroke adverse mental outcomes (PSAMO). This study studied 179 patients with stroke, who were further classified into PSAMO versus no PSAMO group based on the results of validated depression and anxiety questionnaires, which are the industry's gold standard. This study collected demographic and sociological data, quality of life scores, stroke-related information, medical and medication history, and comorbidities. In addition, sequential data such as daily lab results taken seven consecutive days after admission are also collected. The combination of using DL algorithms, such as multi-layer perceptron (MLP) and long short-term memory (LSTM), which can process complex patterns in the data, and the inclusion of new data types, such as sequential data, helped to improve model performance. Accurate prediction of PSAMO helps clinicians make early intervention care plans and potentially reduce the incidence of PSAMO.

Keywords: artificial intelligence; deep learning; machine learning; neural network; post-stroke anxiety; post-stroke depression



Academic Editors: Yunfeng Wu and Pietro Aricò

Received: 5 March 2025

Revised: 25 April 2025

Accepted: 13 May 2025

Published: 14 May 2025

Citation: Oei, C.W.; Ng, E.Y.K.; Ng, M.H.S.; Chan, Y.M.; Subbhuraam, V.; Chan, L.G.; Acharya, U.R. Automated Risk Prediction of Post-Stroke Adverse Mental Outcomes Using Deep Learning Methods and Sequential Data. *Bioengineering* **2025**, *12*, 517. <https://doi.org/10.3390/bioengineering12050517>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

One of the main contributors to mortality and morbidity of the world population is stroke. In an article published by Global Stroke Factsheet in 2022, one in four people are estimated to have a stroke in a lifetime. The lifetime risk of developing stroke has increased by 50% over the last 17 years [1]. Based on reports from the Singapore Stroke

Registry in 2020, the number of stroke episodes increased from 5890 in 2010 to 8846 in 2020, a significant rise in the past decade [2].

The prognosis of stroke varies widely and is influenced by several factors, including the type of stroke, severity, location of brain injury, patient age, pre-existing comorbidities, and the timeliness and quality of medical intervention.

Stroke can be classified into two types: ischemic and hemorrhagic. The prognosis of ischemic and hemorrhagic strokes differs primarily in the pattern of recovery, risks of complications, and rehabilitation needs. In ischemic stroke, prognosis is shaped by the extent of brain tissue damage caused by the lack of blood flow and how quickly circulation is restored. Hemorrhagic stroke prognosis, on the other hand, is influenced by the severity and location of bleeding, as well as the resulting pressure on brain structures. Recovery is often less predictable due to the potential for acute complications [3].

Pre-existing comorbidities, such as diabetes mellitus (DM), can significantly impact the prognosis of stroke by increasing the risk of complications. DM increases the risk of stroke by damaging blood vessels and promoting clot formation. In the context of stroke, diabetic patients often experience larger areas of brain injury and slower functional recovery. Their prognosis tends to be more complicated due to poor wound healing, increased risk of infections, and fluctuating blood sugar levels that can affect brain recovery. Additionally, they are more prone to recurrent strokes and long-term cognitive decline, making post-stroke rehabilitation more prolonged and challenging [4].

While some individuals experience significant recovery, especially with early rehabilitation, others may suffer from long-term impairments in mobility, speech, cognition, or emotional regulation [2].

Many stroke survivors are burdened by adverse outcomes such as depression and anxiety. Literature has termed it as post-stroke depression (PSD) and post-stroke anxiety (PSA), collectively known as post-stroke adverse mental outcomes (PSAMO). It is estimated that 20% of stroke survivors develop anxiety in their lifetime [5,6] and about 30% experience depression [5,7]. Stroke survivors with PSAMO experience negative health outcomes, such as higher mortality rate [8,9] and poorer functional disability [10,11]. The presence of untreated post-stroke anxiety or depression can place significant emotional and practical strain on family members and caregivers, potentially leading to burnout, reduced quality of care, and interpersonal conflict within the support system.

Studies have also shown that there are shared factors that contribute to the incidence of PSA and PSD, such as left hemisphere lesions and the presence of cognitive impairment [12]. Early treatment using antidepressants and psychotherapy in the early stages of stroke can be helpful to manage the onset of PSAMO [13–15]. Therefore, a prediction model is important to accurately predict the risk of PSAMO.

In addition, artificial intelligence (AI) has also gained momentum and has its fair share of success, specifically in the field of healthcare. Deep learning (DL) algorithms have proven to be successful and adopted to develop predictive models for disease prediction [16,17], especially in healthcare, where unstructured data types such as text in clinical notes [18] or images in x-rays [19] and other forms of scans are commonly used in the industry.

1.2. Literature Review

Previous studies have employed artificial intelligence (AI) to predict PSA, PSD, and PSAMO. Most studies used ML algorithms to develop the model. Wang et al. employed multi-layer perceptron (MLP) for the prediction of PSA [20]. However, the evaluation metrics of Wang et al. is the Euclidean distance between the predicted and actual anxiety scores of the patient, which differs from the objective of this study being a classification

model. Oei et al. used ML algorithms to predict the risk of PSAMO and has yet to explore using deep learning (DL) algorithms and methods [21–23].

In summary, most of the studies mentioned above are mainly ML models, and they only used static data for modelling [20–23]. Previous studies have yet to analyze using different data types and complex DL algorithms to develop a predictive model for PSAMO. A summary of the studies reviewed can be found below (Table 1), with each study highlighting the model that exhibits the best performance.

Table 1. Summary of studies for automated prediction of post-stroke adverse mental outcome.

| Author | Dataset | Features | Outcome | Techniques | Best Performance |
|------------------------|-----------------------------------|----------------------------------------------------------------------------------|--------------------|-------------------------------------------------------------------------|--------------------------------------------------|
| Wang et al., 2021 [20] | 395 cases | Demographics, lab results, vascular risk factors | PSA | RF, DT, SVM, stochastic gradient descent, multi-layer perceptron | Demographics, lab results, vascular risk factors |
| Oei et al., 2023 [21] | 285 PSAMO and 1495 no PSAMO cases | Demographics, stroke-related data, surgical and medical history, etc. | PSAMO ¹ | Logistic regression, DT, GBT, RF, XGBoost, CatBoost, AdaBoost, LightGBM | GBT: AUC 0.620; Acc 0.747; F1 score 0.341 |
| Ryu et al., 2022 [22] | 31 PSD and 34 non-PSD cases | Medical history, demographics, neurological, cognitive, and functional test data | PSD | SVM, KNN, RF | SVM: AUC 0.711; Acc 0.70; Sens 0.742; Spec 0.517 |
| Fast et al., 2023 [23] | 49 PSD and 258 non-PSD cases | Demographics, clinical, serological, and MRI data | PSD ¹ | GBT, SVM | GBT: Balanced Acc 0.63; AUC 0.70 |

Acc, accuracy; AUC, area under the curve; DT, decision tree; GBT, gradient-boosted tree; KNN, k-nearest neighbor; RF, random forest; Sens, sensitivity; Spec, specificity; SVM, support vector machine. ¹ Developed models were explainable.

1.3. Motivation and Research Gap

This study has identified two research gaps. Firstly, DL methods have not been largely explored in this area, and this study will be the first to do so. AI has gained popularity by demonstrating its ability to model complex and challenging patterns. The development of AI models has been effective in applications to healthcare problems, such as prediction of readmission, etc. [24–26]. DL methods have also been applied specifically to detect and predict the onset of depression [27,28] and anxiety [29]. Thus, the first objective of this study is to leverage DL to develop a model capable of predicting the risk of PSAMO.

Secondly, most developed models use static data from a single time point. The study of stroke, depression, and anxiety is dynamic and longitudinal [30]. In our analysis of the literature, there was no model that could ingest and use other data types that complement static data for modelling.

Several studies have identified a significant association between elevated white blood cell (WBC) counts and symptoms of depression and anxiety. This relationship reflects the importance of the role of systemic inflammation in mood disorders [31,32]. Hence, the second objective of this study is to adapt and use different data types, such as longitudinal clinical variables such as daily lab results (e.g., white blood count, C-reactive protein, etc.), which are modeled using DL algorithms to capture temporal trends and interactions, hypothesising that including data from various data types leads to better model performance [33].

1.4. Main Contributions

To the best of our knowledge, our study is the first to use DL models to predict the risk of PSAMO instead of PSA and PSD separately. It is also the first to leverage the capabilities of DL to explore using different data types to develop a predictive model for prediction. The

developed model is trained on a 179-subject dataset and has one of the highest accuracies and AUROC thus far. Accurate prediction of PSAMO allows clinicians to formulate care plans at the early stages of stroke and prevent the incidence of PSAMO.

2. Methods

2.1. Data Collection and Study Design

Patients who had been admitted for ischemic or hemorrhagic stroke to a tertiary care hospital are included in the study. The study period is from 2010–2021. These patients would have completed an anxiety or depression screening assessment, assessed within 7–37 days from the time of stroke. The 30-day time window of prediction is consistent with previous literature for identifying PSAMO during stroke recovery [34–37]. Patients who expired before 37 days were excluded from the study. Given the retrospective nature of this study, missing data is a common limitation. Therefore, to preserve data integrity, patients with an excessive proportion of missing values were also excluded from the analysis. In total, there were 179 patients, with 41 of them experiencing PSAMO and 128 without. The study was approved by the ethics review board. A summarized workflow can be found below (Figure 1).



Figure 1. Flow diagram of the proposed model.

Features extracted include baseline characteristics such as social information (e.g., educational level, occupation, etc.), quality of life scores, stroke-related information (e.g., type of stroke, the side where the stroke occurred, etc.), and medical and medication history. Baseline characteristics, collected at the patient’s admission date, are information that remains constant and does not change over a long period. They will be referred to as “static data” in this study.

In addition, this study also collected lab results for 7 consecutive days daily from the patient’s admission date, which will be referred to as “sequential data” in this study. These lab results are usually taken in the day daily to monitor the patient. The motivation to add lab results to the study partially comes from a separate study done by Qiu et al., who showed that biomarkers are a good predictor of depression [38] and, hence, may be useful predictors in our study as well.

Moreover, DL holds an advantage over ML methods, as it can model both static and sequential data. One of the objectives of this study is to use DL methods to capture signals from lab results over a specific time period, which may be valuable for making PSAMO predictions. A list of features that were used to train the model can be found in Table S1 in the Supplementary Materials.

2.2. Identification of PSAMO

The identification of PSA, PSD, or PSAMO was completed through the Hospital Anxiety and Depression Scale (HADS) and the Patient Health Questionnaire (PHQ). HADS uses a 14-item questionnaire that assesses anxiety and depression symptoms in medical patients. PSA and PSD were diagnosed if a patient scored above 7 on the anxiety scale and likewise on the depression scale (82% sensitivity and 78% specificity [39]). PSAMO is diagnosed if a patient has a combined score above 10.

PHQ has two versions: PHQ-9 and PHQ-2. PHQ-9 is a 9-item questionnaire assessing if a patient has depression symptoms; a score above 8 indicates PSD (88% sensitivity and 86% specificity) [40]. PHQ-2 is a simplified version of PHQ-9, containing only 2 questions; a score above 3 denotes PSD (83% sensitivity and 92% specificity) [33,41]. If one of the above criteria is met, the stroke patient is defined to have PSAMO; otherwise, no PSAMO.

2.3. Data Preprocessing and Engineering

Features with more than 25% of missing values were excluded from modelling. As the study is retrospective in nature, missing data is inevitable, mainly due to the loss of follow-up. Multiple imputation-chained equations were used to impute the remaining features [42]. The data is split into train and test sets in a ratio of 70:30. The training set employed a 10-fold cross-validation with 4 repeats to train and validate the model. Test set acted as unseen data and was used to evaluate model performance after training. Continuous features were standardized for the training set and applied to the test set. Categorical features were embedded using embedding layers. Instead of one-hot encoding, embedding helps to represent each individual category into a dense vector, thus reducing the dimensions of the dataset.

A total of 65 features were used to train the model; 46 are static data, while the remaining 19 are sequential data. The list of features can be found in Table S1 in the Supplementary Materials.

2.4. Model Development

The model was developed using a multitude of DL architectures. Each DL architecture has its own uniqueness. DL architectures also allow the flexibility of combining data types for modelling. More details on each DL architecture are explained below. The model is trained on a laptop with a graphic processing unit (GPU) that is CUDA-enabled, running on NVIDIA GeForce RTX 4070. The model is trained on a batch size of 10.

2.4.1. Categorical Embedding

Categorical embedding is a form of representation learning that maps high-dimensional categorical variables into a lower-level representation using a dense vector [43]. It works by training a neural network model on a prediction task. During the model training process, the neural network learns the optimal vector representation for each category. These representations capture relationships and similarities between categories based on the patterns in the training data.

2.4.2. Multi-Layer Perceptron

Multi-layer perceptron (MLP) is a type of artificial neural network combined by multiple layers of interconnected nodes [44]. An MLP consists of three or more layers: an input layer, an output layer, and one or more hidden layers. Each layer consists of weights, biases, and non-linear functions that serve as activating nodes for the other layers. Weights and biases are trained using backpropagation, an example of supervised learning [45].

An iteration of a training loop starts with the forward pass as described below with a general formula:

$$z_k^l = \sum_j w_{kj}^l a_j^{l-1} + b_k^l$$

$$a_k^l = \sigma(z_k^l)$$

l denotes the index of the layer, k denotes the index of the neuron in layer l , w_{kj}^l refers to the weight from neuron j to k in layer l , a_j^{l-1} denotes the activation of neuron j in the previous layer, and b_k^l is the bias of neuron k in layer l . σ is the activation function.

Once the forward pass runs through the whole neural network, the loss can be calculated based on the predicted value and the actual value. The general formula for loss is shown below:

$$Loss = L(\hat{y}, y)$$

L denotes the loss function, \hat{y} is the predicted value, and y is the actual value.

Backpropagation then takes place with the intent of minimizing the loss functions with respect to its weights and biases. The gradients are computed using the chain rule as shown below:

$$\frac{\partial L}{\partial w_{kj}^l} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_k^l} \cdot \frac{\partial z_k^l}{\partial w_{kj}^l}$$

$$\frac{\partial L}{\partial b_k^l} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_k^l} \cdot \frac{\partial z_k^l}{\partial b_k^l}$$

The weights and biases are then updated using a gradient descent algorithm with a hyperparameter learning rate (α), using the formula below:

$$w_{kj}^l \leftarrow w_{kj}^l - \alpha \frac{\partial L}{\partial w_{kj}^l}$$

$$b_k^l \leftarrow b_k^l - \alpha \frac{\partial L}{\partial b_k^l}$$

The process is repeated depending on the number of iterations (epochs) until the model converges.

2.4.3. Long-Short Term Memory (LSTM)

LSTM is based on a recurrent neural network (RNN)-based architecture, possessing the ability for sequence analysis and retaining temporal signals [46]. The structure of an LSTM cell can be found in Figure 2 below.

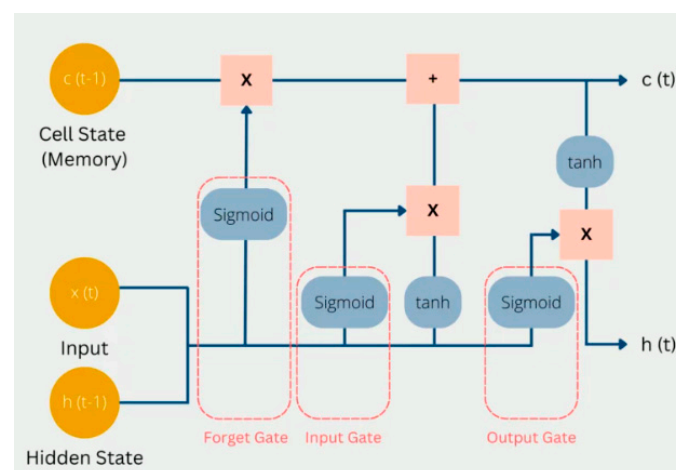


Figure 2. Illustration of an LSTM cell (extracted from: <https://databasecamp.de/en/ml/lstms>, accessed on 21 January 2025).

LSTM possesses a cell state, where it stores and carries information across different time steps. The cell state is influenced by the forget gate, which controls (using the sigmoid

activation function) what information from the cell state should be retained or discarded. The uniqueness of LSTM is its ability to retain past information with its structure [47]. The hidden state in an LSTM cell acts similarly to hidden layers in MLP, where “short-term” information from the previous calculation steps is stored.

The complexity of the LSTM cell having many “gates” enables it to selectively learn, remember, and forget information over long sequences, providing the ability to model and capture long-term dependencies in sequential data [48]. Among these, the forget gate is especially important, as it controls what past information should be discarded from the cell state. By filtering out irrelevant or outdated memory, it prevents information overload and helps the model stay focused on what matters most. This selective forgetting also stabilizes training by reducing the risk of vanishing gradients, making LSTMs more effective in learning patterns across time [49].

LSTM models are particularly well suited for sequential data due to their ability to handle variable-length sequences and capture long-term dependencies. Furthermore, their architecture is designed to retain relevant information over extended time intervals, enabling them to learn and model long-term relationships within the data. This capability is especially valuable in scenarios where earlier inputs significantly influence future outputs, thereby enhancing the model’s contextual understanding and predictive accuracy. Like MLP, LSTM also employs backpropagation for training its weights and biases [50].

2.4.4. Model Architecture

The proposed PSAMO predictive model uses a combination of MLP and LSTM layers, as shown in Figure 3 below. Static and sequential data will be ingested into the model separately. Categorical variables from the static data are first separated and embedded through the embedding layers. An embedding layer works by mapping high-dimensional, sparse input into dense, low-dimensional vectors that capture meaningful relationships. The model learns these relationships during training, allowing similar inputs to match with similar vector representations. The embedded representations are then combined with the remaining variables found in the static data and parsed through the MLP layer, also known as “Multi-layer Perceptron (static data)” in Figure 3 below. “Multi-layer Perceptron (static data)” has three linear layers and a dropout layer with 20% dropout to prevent the model from overfitting [51].

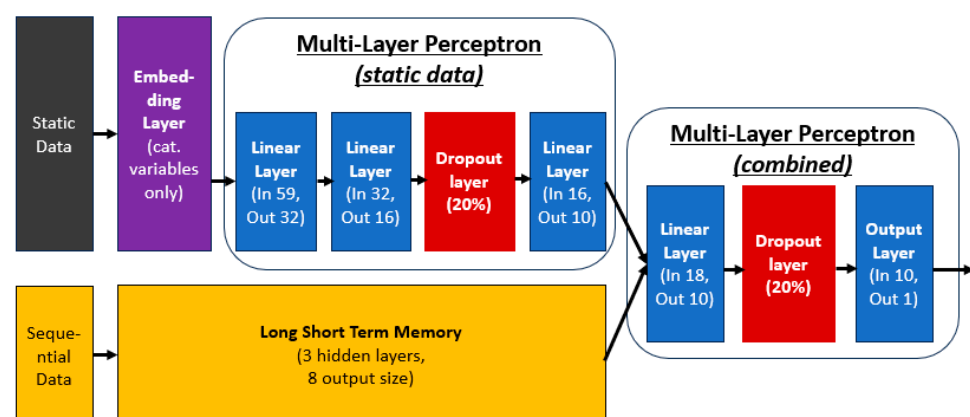


Figure 3. Model architecture—prediction of PSAMO.

Sequential data will be ingested into the model through the LSTM layer, represented by the yellow box in Figure 3 below. The strength of LSTM being able to capture temporal patterns is deployed in this context to capture temporal signals in the patient’s lab results 7 days leading to the prediction window. The LSTM layer has three hidden layers and

outputs an eight-dimensional vector representing the signals captured from the daily lab results.

The proposed deep learning architecture leverages a hybrid design that effectively integrates both static and sequential data through a multi-component structure. The “Multi-layer Perceptron (static data)” module is dedicated to modelling static features, enabling the network to capture complex, non-temporal relationships and interactions. In parallel, the LSTM layer is employed to model sequential data, allowing the architecture to retain temporal dependencies and contextual patterns over time. These two representations are subsequently fused, enabling the model to learn higher-order interactions between static and temporal information. This is represented by “Multi-layer Perceptron (combined)” in Figure 3. This approach enhances the model’s capacity to learn from heterogeneous inputs, ultimately improving predictive performance and generalizability across tasks that involve both temporal dynamics and fixed attributes.

The output layer will be a sigmoid function, predicting the probability of a stroke patient experiencing PSAMO. The output layer will be a sigmoid function, predicting the probability of a stroke patient experiencing PSAMO.

2.4.5. Model Initialization

The weights and biases of MLPs are initialized using “Kaiming initialization” [52], which accounts for the non-linearity of activation functions, such as rectified linear unit (ReLU), which are used in the model. Bias is initialized with 0. The weights are initialized with a normal distribution of

$$w_l \sim N\left(0, \frac{2}{n_l}\right)$$

n_l refers to the number of input neurons at layer l .

The weights and biases of LSTMs are initialized using “Xavier initialization” or “Glorot initialization” [53]. Bias is initialized to be 0.001. The weights are initialized with a uniform distribution of

$$w_{ij} \sim U\left(-\frac{\sqrt{6}}{\sqrt{S_{in} + S_{out}}}, \frac{\sqrt{6}}{\sqrt{S_{in} + S_{out}}}\right)$$

S_{in} is the size of the previous layer, and S_{out} is the size of the current layer.

Initializing the weights and biases of the neural network helps in faster convergence and avoids vanishing and exploding gradients. If the weights are too small, the gradients may vanish, making learning slow. On the other hand, if the weights are too large, the gradients may explode, leading to numerical instability. Proper initialization helps maintain a balance and prevents the model from reaching a local minima or maxima at the early stages of training [54,55].

2.4.6. Model Training

The model is trained using mini-batch gradient descent (MBGD) [56], employing a learning rate of 0.001 and 2000 epochs. MBGD has been proven to have faster convergence [57] and efficient use of memory, as it processes only a small batch of training data at a given time and produces better generalization, which can help the model avoid local minima [58]. The formula for MBGD that updates the weights can be found below:

$$w_{i+1} = w_i - \alpha \cdot \nabla_{w_i} L\left(x^{i:i+b}, y^{i:i+b}; w_i\right)$$

w_i represents the weights of the layers in the model in iteration i . α represents the learning rate, and b denotes the size of a single batch of data. The objective of MBGD is to minimize the loss function $L(..)$ with respect to w .

The training pipeline also included an early stopping criterion of 50 epochs of minimal (0.001) or no improvement of the loss in the validation set to prevent overfitting.

As the model is a classification model, binary cross entropy loss is used as a loss function [59,60]. The formula for binary cross entropy loss can be found below:

$$-\frac{1}{N}\sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$

N refers to total samples, y_i represents the actual class of sample i , and $\log(p(y_i))$ is the probability of being classified into that class.

Binary cross entropy loss penalizes incorrect labelling of the data class by a model if deviations in probability occur in classifying the labels. The objective is to minimize binary cross entropy loss through backpropagation, where low loss values equate to high accuracy.

2.5. Model Evaluation

The model was evaluated using standard classification metrics, accuracy, and F1 score. Area under receiver operating characteristics (AUROC) curve was also reported to measure the degree of discrimination between the two groups [61].

2.6. Packages Used

The preprocessing and modelling done in this study were implemented using Python 3 [62–66]. Each package used and their functions are listed in Appendix A.

3. Results

The classification results are shown in Table 2 below. The train set is reported with a confidence interval from the 10-fold cross-validation, while the test set is the performance of the model on the unseen data. It is evident that the proposed model in this study (MLP with LSTM) obtained the highest accuracy and AUROC of 0.792 and 0.789, respectively, which is better than baseline traditional ML algorithms. The F1 score has also improved to 0.353, improving classification accuracy for the minority class as well. The confusion matrix of the classification can be found in Table 3 below.

Table 2. Model performance in prediction of PSAMO.

| Developed Models | Train Set | | | Test Set | | |
|--------------------------------------------------------------------------------------------------|------------------------|------------------------|------------------------|----------|-------|----------|
| | Accuracy | AUROC | F1 Score | Accuracy | AUROC | F1 Score |
| Gradient-Boosted Trees (Oei et al., 2023) [21] (Best Model using Classical ML Approach) | 0.973 (0.958–0.982) | 0.946 (0.932–0.957) | 0.950 (0.924–0.964) | 0.747 | 0.620 | 0.341 |
| MLP + LSTM (Using both static and sequential data) | 0.823 (0.721–0.852) | 0.752 (0.621–0.784) | 0.586 (0.328–0.622) | 0.796 | 0.789 | 0.353 |

Table 3. Confusion matrix obtained from the test set.

| Actual label | Non-PSAMO PSAMO | Predicted Label | |
|--------------|--------------------|----------------------|--------------------|
| | | Non-PSAMO | PSAMO |
| | | 40 cases 11 cases | 0 cases 3 cases |

The difference between the proposed model in this study and the traditional ML models is (1) the utilization of DL architectures and (2) the addition of sequential data (daily lab results) for model training, yielding better performance.

4. Discussion

4.1. Model Performance

Table 2 above shows the improvement in model performance by using DL methods (MLP and LSTM) compared to the ML model (gradient-boosted trees) developed by Oei et al. Accuracy, AUROC, and F1 score have improved to 0.792, 0.789, and 0.353, respectively. Benchmarking against the models published to date, our model attained higher accuracy and AUROC (see Table 1).

Table 4 below provides a summary and compares the proposed method in this study versus other studies. The improvement in performance can be attributed to two reasons.

Table 4. Summary of a comparison of methods for automated prediction of post-stroke adverse mental outcome.

| Author | Data Type | | Artificial Intelligence Methods | | Outcome | Best Performance |
|------------------------|-----------|------------|---------------------------------|---------------|--------------------|--------------------------------------------------|
| | Static | Sequential | Machine Learning | Deep Learning | | |
| Wang et al., 2021 [20] | ✓ | | ✓ | | PSA | 18.625 Euclidean distance between anxiety scores |
| Oei at al., 2023 [21] | ✓ | | ✓ | | PSAMO ¹ | AUC 0.620; Acc 0.747; F1 score 0.341 |
| Ryu et al., 2022 [22] | ✓ | | ✓ | | PSD | AUC 0.711; Acc 0.70; Sens 0.742; Spec 0.517 |
| Fast et al., 2023 [23] | ✓ | | ✓ | | PSD ¹ | Balanced Acc 0.63; AUC 0.70 |
| Current study | ✓ | ✓ | | ✓ | PSAMO | AUC 0.789; Acc 0.796; F1-score 0.353 |

Sens, sensitivity; Spec, specificity; SVM, support vector machine. ¹ Developed models were explainable.

First is the utilization of DL algorithms for modelling. As shown in Table 4, most literature has utilized only classical ML methods, such as support vector machines, which may not effectively capture the complexity of the underlying patterns [67]. The nature of DL algorithms being able to effectively capture complex non-linear relationships may be one of the factors that contribute to the improvement of the prediction [68–70]. However, MLP itself does not inherently capture sequential dependencies, limiting its ability to model sequential or time-series data [71]. Therefore, there is a need for LSTM architectures to be incorporated to model the sequential lab results data that were collected, as proposed in this work.

Second, adding sequential lab results data, along with the LSTM architecture, improves model performance. Using a combination of different data types for prediction has proven successful in healthcare, as done by Zhang et al. [33]. Our study suggests potential signals found in the daily lab results that might contribute to the prediction of PSAMO, leading to the prediction window. The robustness of DL architectures enables different data types to be harmonized and modelled together, facilitating the integration of diverse information sources into a unified framework for more comprehensive and accurate insights.

4.2. Related Works

This study demonstrated a modest improvement in predictive performance, as reflected in both AUC and accuracy metrics, compared to existing work in the field (Table 1). For instance, Ryu et al. reported an AUC of 0.711, whereas the present study achieved an AUC of 0.791. Additionally, the current study employed a substantially larger sample

size ($n = 179$ vs. $n = 65$), enhancing the generalizability of its findings. Notably, while Ryu et al. utilized traditional machine learning algorithms, this study leveraged deep learning approaches capable of modeling sequential data, an aspect that may contribute to improved predictive capability. It is important to note, however, that Ryu et al. work focused exclusively on PSD, whereas the present study addressed a broader spectrum of PSAMO. As such, a direct one-to-one comparison between the two studies is limited in scope [22].

In comparison to the study conducted by Oei et al. [21], the present study incorporates a more comprehensive set of features and a richer dataset. The inclusion of sequential laboratory results and the transition from traditional ML models to DL architectures may have contributed to the observed improvement in predictive performance. This is evidenced by an increase in AUC from 0.620 (Oei et al.) to 0.789 in the current study. Nevertheless, it is important to acknowledge that Oei et al. study employed a significantly larger sample size ($n = 1790$ vs. $n = 179$), which enhances the external validity and generalizability of their findings [21].

Recent advancements in the application of “dynamic displacement” data to machine learning models have demonstrated considerable promise in a range of domains. Several studies have highlighted the value of time-evolving physiological signals (e.g., biomarker fluctuations) in enhancing predictive accuracy [72–74]. In the context of healthcare, where baseline values and deviations from baseline are clinically meaningful, dynamic modeling approaches have proven effective in capturing subtle physiological changes that may precede critical adverse events [75,76]. By accounting for patterns such as sustained elevations, abrupt inflections, and rates of change in laboratory parameters like WBC count or CRP, machine learning models can potentially produce better results. LSTM networks, like the one used in this study, are well suited to capture these evolving dynamics, thereby facilitating earlier clinical intervention and more precise risk stratification.

Using this study as an example, an acute elevation in WBC count can be interpreted as a clinically meaningful deviation from baseline, representing a form of temporal displacement that the model is trained to detect and learn from. The LSTM model is designed to implicitly learn such non-linear and time-dependent patterns, enabling it to distinguish between transient fluctuations and sustained trends. Future extensions of this work may explore the explicit engineering of displacement-related features, including the rate of change, cumulative deviation from baseline, and other time-sensitive transformations. Incorporating these elements has the potential to enhance both the predictive performance.

4.3. Advantages and Limitations

PSAMO is a complex and multi-disciplinary condition; indicators and signals that predict the onset of it may come from various data types. This study was also the first to combine data types for modelling the prediction of PSAMO. The inclusion of sequential time series data in the model has improved model prediction. The developed model in this study performs better than most models in predicting PSAMO. Past studies have also suggested that regular screening for depression helps to prevent the incidence of PSAMO. Therefore, improved accuracy of such prediction can yield better triaging of stroke patients, leading to better allocation of healthcare resources and better management of such conditions upstream [77,78].

Having a manually human-administered PHQ and HADS test is also subject to human errors [79], translation errors [80], and biases [81]. Developing a predictive model helps mitigate these issues by introducing a systematic and automated approach and reducing reliance on subjective interpretation [82].

However, the nature of using complex DL architecture has made the model unexplainable. The model was also tested on a smaller sample size due to missing data, and further validations with different sites are necessary for external validity. Future expansion of this study can include data from other data types, such as images or text, for a more comprehensive approach, as stroke, depression, and anxiety are multi-disciplinary conditions.

Though the model has proven that adding sequential data improves model performance, it is operationally intensive for data or results to be taken daily. Clinicians on the ground should carry out careful considerations to weigh between operational needs and model performance. Additionally, data quality is essential for the successful implementation of this method. As the approach is sensitive to missing or inaccurate data, the validity of the results may be affected by inconsistencies or gaps within the dataset.

A small sample size poses significant limitations, particularly in the context of complex operations like stroke patient management and hospital workflows. The diversity of stroke cases, varying recovery trajectories, and individualized treatment protocols make it challenging to capture a representative dataset. Moreover, the dynamic nature of hospital operations, influenced by fluctuating patient loads and resource constraints, introduces variability that a small sample size might fail to reflect accurately. This limitation can result in models or analyses that restrict generalizability.

4.4. Potential Implementation

The implementation of a predictive model such as what was mentioned in this paper offers several significant clinical and operational benefits. Early identification of patients at high risk allows for timely psychological assessment and intervention, which can prevent the escalation of symptoms and improve overall mental health outcomes. From a healthcare systems perspective, it enables more efficient allocation of mental health resources and supports personalized treatment planning. Additionally, early intervention can alleviate the emotional and logistical burden on caregivers, contributing to a more sustainable and supportive recovery environment. Overall, such a model promotes a holistic approach to stroke recovery by integrating mental health into routine clinical care.

However, to ensure effective implementation, several components must be addressed. First, clinical validation and external testing of the model are essential to confirm its predictive accuracy and generalizability across diverse patient populations. Without robust validation, there is a risk of overfitting or misclassification, which could lead to inappropriate care decisions. Second, training and awareness among clinical staff must accompany implementation so that providers understand the model's purpose, limitations, and how to use its outputs responsibly [83]. The objective of the model should allow clinicians to work in tandem with the model and not replace clinical decisions.

Additionally, ongoing monitoring and recalibration of the model are critical to maintain its performance over time [84]. As patient demographics drift, care practices, or mental health trends evolve, the model must be updated to reflect these changes. Creating feedback loops where clinicians and researchers can evaluate predictions against real outcomes will strengthen the model's reliability and support continuous improvement. These steps are vital for transforming a predictive model from a theoretical tool into a practical, effective part of stroke care that contributes to better long-term recovery outcomes.

5. Conclusions

In conclusion, this study has shown that using DL architectures and the addition of sequential time series data has improved model performance in predicting PSAMO with 79.2% accuracy. The improvement of model performance may lead to better predictions of PSAMO, better triaging of healthcare resources, and potentially reducing the incidence

of PSAMO. Early detection and intervention can mitigate the emotional, physical, and logistical demands placed on caregivers, thereby enhancing their well-being and sustaining their capacity to provide long-term support. This study has yet to explore using other data types (e.g., text or images) for prediction, which can potentially be studied in the future and may improve model performance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/bioengineering12050517/s1>, Table S1: List of features collected and used for modelling.

Author Contributions: Conceptualization, C.W.O. and E.Y.K.N.; Methodology and formal analysis, C.W.O., L.G.C. and U.R.A.; Data curation, M.H.S.N. and L.G.C.; Writing—original draft preparation, C.W.O.; Writing—review and editing, V.S., E.Y.K.N., L.G.C. and U.R.A.; supervision, E.Y.K.N., L.G.C. and U.R.A.; project administration, C.W.O., E.Y.K.N., M.H.S.N., L.G.C. and Y.M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study is approved by the Ethics Committee of National Healthcare Group, Domain Specific Review Board (DSRB) (DSRB Ref No. 2022/00289, Approved on 20 January 2022).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is unavailable due to privacy restrictions.

Acknowledgments: The authors would like to thank Tan Tock Seng Hospital, Singapore, for allowing this study to be done.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-------|---------------------------------------|
| HADS | Hospital Anxiety and Depression Scale |
| PHQ | Patient Health Questionnaire |
| PSAMO | Post-Stroke Adverse Mental Outcomes |
| PSA | Post-Stroke Anxiety |
| PSD | Post-Stroke Depression |

Appendix A

Appendix A lists the key Python packages used in this analysis and research.

Table A1. List of Python packages used.

| No. | Package Name | Functions |
|-----|--------------|-------------------------------------------------------------------------|
| 1 | Pandas | Data preprocessing and engineering |
| 2 | Numpy | |
| 3 | fancyimpute | Implementation of MICE imputation |
| 4 | scikit-learn | Implementation of 10-fold cross-validation and model diagnostic metrics |
| 5 | Matplotlib | Display of graphs and charts |
| 6 | Pytorch | Training of deep learning models |

References

- Feigin, V.L.; Brainin, M.; Norrving, B.; Martins, S.; Sacco, R.L.; Hacke, W.; Fisher, M.; Pandian, J.; Lindsay, P. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *Int. J. Stroke Off. J. Int. Stroke Soc.* **2022**, *17*, 18–29. [\[CrossRef\]](#)
- Stroke—National Registry of Diseases Office. Available online: <https://nrdo.gov.sg/publications/stroke> (accessed on 29 February 2024).
- Miller, A.H.; Raison, C.L. The role of inflammation in depression: From evolutionary imperative to modern treatment target. *Nat. Rev. Immunol.* **2016**, *16*, 22–34. [\[CrossRef\]](#) [\[PubMed\]](#)
- American Heart Association; American Stroke Association. Diabetes and Stroke: A Scientific Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke* **2014**, *45*, 3035–3058.
- Schöttke, H.; Giabbiconi, C.-M. Post-stroke depression and post-stroke anxiety: Prevalence and predictors. *Int. Psychogeriatr.* **2015**, *27*, 1805–1812. [\[CrossRef\]](#) [\[PubMed\]](#)
- Burton, C.A.C.; Murray, J.; Holmes, J.; Astin, F.; Greenwood, D.; Knapp, P. Frequency of Anxiety after Stroke: A Systematic Review and Meta-Analysis of Observational Studies. *Int. J. Stroke* **2013**, *8*, 545–559. [\[CrossRef\]](#)
- Ellis, C.; Zhao, Y.; Egede, L.E. Depression and increased risk of death in adults with stroke. *J. Psychosom. Res.* **2010**, *68*, 545–551. [\[CrossRef\]](#)
- de Mello, R.F.; Santos, I.d.S.; Alencar, A.P.; Benseñor, I.M.; Lotufo, P.A.; Goulart, A.C. Major Depression as a Predictor of Poor Long-Term Survival in a Brazilian Stroke Cohort (Study of Stroke Mortality and Morbidity in Adults) EMMA study. *J. Stroke Cerebrovasc. Dis.* **2016**, *25*, 618–625. [\[CrossRef\]](#)
- Cai, W.; Mueller, C.; Li, Y.-J.; Shen, W.-D.; Stewart, R. Post stroke depression and risk of stroke recurrence and mortality: A systematic review and meta-analysis. *Ageing Res. Rev.* **2019**, *50*, 102–109. [\[CrossRef\]](#) [\[PubMed\]](#)
- Astuti, P.; Kusnanto, K.; Dwi Novitasari, F. Depression and functional disability in stroke patients. *J. Public Health Res.* **2020**, *9*, 1835. [\[CrossRef\]](#)
- Lee, E.-H.; Kim, J.-W.; Kang, H.-J.; Kim, S.-W.; Kim, J.-T.; Park, M.-S.; Cho, K.-H.; Kim, J.-M. Association between Anxiety and Functional Outcomes in Patients with Stroke: A 1-Year Longitudinal Study. *Psychiatry Investig.* **2019**, *16*, 919–925. [\[CrossRef\]](#)
- Starkstein, S.E.; Cohen, B.S.; Fedoroff, P.; Parikh, R.M.; Price, T.R.; Robinson, R.G. Relationship between anxiety disorders and depressive disorders in patients with cerebrovascular injury. *Arch. Gen. Psychiatry* **1990**, *47*, 246–251. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fang, Y.; Mpofu, E.; Athanasou, J. Reducing depressive or anxiety symptoms in post-stroke patients: Pilot trial of a constructive integrative psychosocial intervention. *Int. J. Health Sci.* **2017**, *11*, 53–58.
- Woranush, W.; Moskopp, M.L.; Sedghi, A.; Stuckart, I.; Noll, T.; Barlinn, K.; Siepmann, T. Preventive Approaches for Post-Stroke Depression: Where Do We Stand? A Systematic Review. *Neuropsychiatry Dis. Treat.* **2021**, *17*, 3359–3377. [\[CrossRef\]](#)
- Mikami, K.; Jorge, R.E.; Moser, D.J.; Arndt, S.; Jang, M.; Solodkin, A.; Small, S.L.; Fonzetti, P.; Hegel, M.T.; Robinson, R.G. Prevention of Post-Stroke Generalized Anxiety Disorder, Using Escitalopram or Problem-Solving Therapy. *J. Neuropsychiatry Clin. Neurosci.* **2014**, *26*, 323–328. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yu, Z.; Wang, K.; Wan, Z.; Xie, S.; Lv, Z. Popular deep learning algorithms for disease prediction: A review. *Cluster Comput.* **2023**, *26*, 1231–1251. [\[CrossRef\]](#)
- Byeon, H.; Gc, P.; Hannan, S.A.; Alghayadh, F.Y.; Soomar, A.M.; Soni, M.; Bhatt, M.W. Deep neural network model for enhancing disease prediction using auto encoder based broad learning Byeon. *SLAS Technol.* **2024**, *29*, 100145. [\[CrossRef\]](#)
- Khalifa, M.; Albadawy, M. Artificial Intelligence for Clinical Prediction: Exploring Key Domains and Essential Functions. *Comput. Methods Programs Biomed. Update* **2024**, *5*, 100148. [\[CrossRef\]](#)
- Çallı, E.; Sogancioglu, E.; van Ginneken, B.; van Leeuwen, K.G.; Murphy, K. Deep learning for chest X-ray analysis: A survey. *Med. Image Anal.* **2021**, *72*, 102125. [\[CrossRef\]](#)
- Wang, J.; Zhao, D.; Lin, M.; Huang, X.; Shang, X. Post-Stroke Anxiety Analysis via Machine Learning Methods. *Front. Aging Neurosci.* **2021**, *13*, 657937. Available online: <https://www.frontiersin.org/articles/10.3389/fnagi.2021.657937> (accessed on 13 July 2023). [\[CrossRef\]](#)
- Oei, C.W.; Ng, E.Y.K.; Ng, M.H.S.; Tan, R.-S.; Chan, Y.M.; Chan, L.G.; Acharya, U.R. Explainable Risk Prediction of Post-Stroke Adverse Mental Outcomes Using Machine Learning Techniques in a Population of 1780 Patients. *Sensors* **2023**, *23*, 7946. [\[CrossRef\]](#)
- Ryu, Y.H.; Kim, S.Y.; Kim, T.U.; Lee, S.J.; Park, S.J.; Jung, H.-Y.; Hyun, J.K. Prediction of Poststroke Depression Based on the Outcomes of Machine Learning Algorithms. *J. Clin. Med.* **2022**, *11*, 2264. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fast, L.; Temuulen, U.; Villringer, K.; Kufner, A.; Ali, H.F.; Siebert, E.; Huo, S.; Piper, S.K.; Sperber, P.S.; Liman, T.; et al. Machine Learning-Based Prediction of Clinical Outcomes After First-Ever Ischemic Stroke. *Front. Neurol.* **2023**, *14*, 1114360. Available online: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1114360> (accessed on 13 July 2023). [\[CrossRef\]](#) [\[PubMed\]](#)
- Gao, R.; Cheng, W.X.; Suganthan, P.N.; Yuen, K.F. Inpatient Discharges Forecasting for Singapore Hospitals by Machine Learning. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4966–4975. [\[CrossRef\]](#)
- Davis, S.; Zhang, J.; Lee, I.; Rezaei, M.; Greiner, R.; McAlister, F.A.; Padwal, R. Effective hospital readmission prediction models using machine-learned features. *BMC Health Serv. Res.* **2022**, *22*, 1415. [\[CrossRef\]](#)

26. Nithya, B.; Ilango, V. Predictive analytics in health care using machine learning tools and techniques. In Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 15–16 June 2017; pp. 492–499.
27. Vandana; Marriwala, N.; Chaudhary, D. A hybrid model for depression detection using deep learning. *Meas. Sens.* **2023**, *25*, 100587. [\[CrossRef\]](#)
28. Uddin, M.Z.; Dysthe, K.K.; Følstad, A.; Brandtzaeg, P.B. Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Comput. Appl.* **2022**, *34*, 721–744. [\[CrossRef\]](#)
29. Daza, A.; Saboya, N.; Necochea-Chamorro, J.I.; Zavaleta Ramos, K.; Vásquez Valencia, Y. del R. Systematic review of machine learning techniques to predict anxiety and stress in college students. *Inform. Med. Unlocked* **2023**, *43*, 101391. [\[CrossRef\]](#)
30. Aström, M.; Adolfsson, R.; Asplund, K. Major depression in stroke patients. A 3-year longitudinal study. *Stroke* **1993**, *24*, 976–982. [\[CrossRef\]](#)
31. Shafiee, M.; Tayefi, M.; Hassanian, S.M.; Ghaneifar, Z.; Arabzadeh, S.; Ferns, G.A. Depression and anxiety symptoms are associated with white blood cell count and red cell distribution width: A sex-stratified analysis in a population-based study. *Brain Behav. Immun.* **2017**, *61*, 198–203. [\[CrossRef\]](#)
32. Sealock, J.M.; Lee, Y.H.; Moscati, A.; Smoller, J.W.; Davis, L.K. Use of the PsycheMERGE network to investigate the association between depression polygenic scores and white blood cell count. *JAMA Psychiatry* **2021**, *78*, 1365–1374. [\[CrossRef\]](#)
33. Zhang, D.; Yin, C.; Zeng, J.; Yuan, X.; Zhang, P. Combining structured and unstructured data for predictive models: A deep learning approach. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 280. [\[CrossRef\]](#)
34. Chen, Y.-M.; Chen, P.-C.; Lin, W.-C.; Hung, K.-C.; Chen, Y.-C.B.; Hung, C.-F.; Wang, L.-J.; Wu, C.-N.; Hsu, C.-W.; Kao, H.-Y. Predicting New-Onset Post-Stroke Depression from Real-World Data Using Machine Learning Algorithm. *Front. Psychiatry* **2023**, *14*, 1195586. Available online: <https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1195586> (accessed on 20 June 2023). [\[CrossRef\]](#) [\[PubMed\]](#)
35. Zhanina, M.Y.; Druzhkova, T.A.; Yakovlev, A.A.; Vladimirova, E.E.; Freiman, S.V.; Eremina, N.N.; Guekht, A.B.; Gulyaeva, N.V. Development of Post-Stroke Cognitive and Depressive Disturbances: Associations with Neurohumoral Indices. *Curr. Issues Mol. Biol.* **2022**, *44*, 6290–6305. [\[CrossRef\]](#)
36. Li, J.; Oakley, L.D.; Brown, R.L.; Li, Y.; Luo, Y. Properties of the Early Symptom Measurement of Post-Stroke Depression: Concurrent Criterion Validity and Cutoff Scores. *J. Nurs. Res. JNR* **2020**, *28*, e107. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Khazaal, W.; Taliani, M.; Boutros, C.; Abou-Abbas, L.; Hosseini, H.; Salameh, P.; Sadier, N.S. Psychological Complications at 3 Months Following Stroke: Prevalence and Correlates Among Stroke Survivors in Lebanon. *Front. Psychol.* **2021**, *12*, 663267. [\[CrossRef\]](#)
38. Qiu, X.; Wang, H.; Lan, Y.; Miao, J.; Pan, C.; Sun, W.; Li, G.; Wang, Y.; Zhao, X.; Zhu, Z.; et al. Blood biomarkers of post-stroke depression after minor stroke at three months in males and females. *BMC Psychiatry* **2022**, *22*, 162. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Wu, Y.; Levis, B.; Sun, Y.; He, C.; Krishnan, A.; Neupane, D.; Bhandari, P.M.; Negeri, Z.; Benedetti, A.; Thombs, B.D. Accuracy of the Hospital Anxiety and Depression Scale Depression subscale (HADS-D) to screen for major depression: Systematic review and individual participant data meta-analysis. *BMJ* **2021**, *373*, n972. [\[CrossRef\]](#)
40. Urtasun, M.; Daray, F.M.; Teti, G.L.; Coppolillo, F.; Herlax, G.; Saba, G.; Rubinstein, A.; Araya, R.; Irazola, V. Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina. *BMC Psychiatry* **2019**, *19*, 291. [\[CrossRef\]](#)
41. Kroenke, K.; Spitzer, R.L.; Williams, J.B.W. The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Med. Care* **2003**, *41*, 1284–1292. [\[CrossRef\]](#)
42. Jakobsen, J.C.; Gluud, C.; Wetterslev, J.; Winkel, P. When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Med. Res. Methodol.* **2017**, *17*, 162. [\[CrossRef\]](#)
43. Guo, C.; Berkahn, F. Entity Embeddings of Categorical Variables 2016. *arXiv* **2016**, arXiv:1604.06737. [\[CrossRef\]](#)
44. Popescu, M.-C.; Balas, V.; Perescu-Popescu, L.; Mastorakis, N. Multilayer perceptron and neural networks. *WSEAS Trans. Circuits Syst.* **2009**, *8*, 579–588.
45. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [\[CrossRef\]](#)
46. Ghojogh, B.; Ghodsi, A. Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey. *arXiv* **2023**, arXiv:2304.11461. Available online: <http://arxiv.org/abs/2304.11461> (accessed on 4 March 2024).
47. Long Short-Term Memory Networks (LSTM)- Simply Explained! | Data Basecamp. Available online: <https://databasecamp.de/en/ml/lstms> (accessed on 4 March 2024).
48. Nath Saha, B.; Senapati, A. Long Short Term Memory (LSTM) based Deep Learning for Sentiment Analysis of English and Spanish Data. In Proceedings of the 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2–4 July 2020; pp. 442–446.
49. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Staudemeyer, R.C.; Morris, E.R. Understanding LSTM—A tutorial into Long Short-Term Memory Recurrent Neural Networks. *arXiv* **2019**, arXiv:1909.09586. [\[CrossRef\]](#)

51. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 929–1958.
52. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv* **2015**, arXiv:1502.01852. [[CrossRef](#)]
53. Glorot, X.; Bengio, Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; JMLR Workshop and Conference Proceedings; pp. 249–256. Available online: <https://proceedings.mlr.press/v9/glorot10a.html> (accessed on 24 February 2024).
54. Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the Importance of Initialization and Momentum in Deep Learning. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; PMLR; pp. 1139–1147. Available online: <https://proceedings.mlr.press/v28/sutskever13.html> (accessed on 4 March 2024).
55. Kumar, S.K. On weight initialization in deep neural networks. *arXiv* **2017**, arXiv:1704.08863. [[CrossRef](#)]
56. Qian, X.; Klabjan, D. The Impact of the Mini-batch Size on the Variance of Gradients in Stochastic Gradient Descent. *arXiv* **2020**, arXiv:2004.13146. [[CrossRef](#)]
57. Khirirat, S.; Feyzmahdavian, H.R.; Johansson, M. Mini-batch gradient descent: Faster convergence under data sparsity. In Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC), Melbourne, Australia, 12–15 December 2017; pp. 2880–2887. [[CrossRef](#)]
58. Li, M.; Zhang, T.; Chen, Y.; Smola, A.J. Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; ACM: New York, NY, USA, 2014; pp. 661–670.
59. Wali, R. Xtreme Margin: A Tunable Loss Function for Binary Classification. *Mach. Learn.* **2022**. [[CrossRef](#)]
60. Zhang, Z.; Sabuncu, M.R. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. *arXiv* **2018**, arXiv:1805.07836. [[CrossRef](#)]
61. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
62. McKinney, W. Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 51–56. [[CrossRef](#)]
63. Array Programming with NumPy | Nature. Available online: <https://www.nature.com/articles/s41586-020-2649-2> (accessed on 20 June 2023).
64. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
65. Droettboom, M.; Hunter, J.; Firing, E.; Caswell, T.A.; Dale, D.; Lee, J.-J.; Elson, P.; McDougall, D.; Straw, A.; Root, B.; et al. *Matplotlib*, version 1.4.0; Zenodo: Geneva, Switzerland, 2014. [[CrossRef](#)]
66. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, arXiv:1912.01703. [[CrossRef](#)]
67. Raudys, Š. How good are support vector machines? *Neural Netw.* **2000**, *13*, 17–19. [[CrossRef](#)]
68. Dargan, S.; Kumar, M.; Ayyagari, M.R.; Kumar, G. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Arch. Comput. Methods Eng.* **2020**, *27*, 1071–1092. [[CrossRef](#)]
69. Lee, J.; Woo, J.; Kang, A.R.; Jeong, Y.-S.; Jung, W.; Lee, M.; Kim, S.H. Comparative Analysis on Machine Learning and Deep Learning to Predict Post-Induction Hypotension. *Sensors* **2020**, *20*, 4575. [[CrossRef](#)]
70. Badawy, M.; Ramadan, N.; Hefny, H.A. Healthcare predictive analytics using machine learning and deep learning techniques: A survey. *J. Electr. Syst. Inf. Technol.* **2023**, *10*, 40. [[CrossRef](#)]
71. Djerioui, M.; Brik, Y.; Ladjal, M.; Attallah, B. Heart Disease prediction using MLP and LSTM models. In Proceedings of the 2020 International Conference on Electrical Engineering (ICEE), Istanbul, Turkey, 25–27 September 2020; pp. 1–5.
72. Brain Sciences | Free Full-Text | Predictive Biomarkers of Treatment Response in Major Depressive Disorder. Available online: <https://www.mdpi.com/2076-3425/13/11/1570> (accessed on 7 March 2024).
73. Nathan, R.; Monk, C.T.; Arlinghaus, R.; Adam, T.; Alós, J.; Assaf, M.; Baktoft, H.; Beardsworth, C.E.; Bertram, M.G.; Bijleveld, A.I.; et al. Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* **2022**, *375*, eabg1780. [[CrossRef](#)]
74. Meyer, P.G.; Cherstvy, A.G.; Seckler, H.; Hering, R.; Blaum, N.; Jeltsch, F.; Metzler, R. Directedness, correlations, and daily cycles in springbok motion: From data via stochastic models to movement prediction. *Phys. Rev. Res.* **2023**, *5*, 043129. [[CrossRef](#)]
75. Cabalar, I.; Le, T.H.; Silber, A.; O'Hara, M.; Abdallah, B.; Parikh, M.; Busch, R. The Role of Blood Testing in Prevention, Diagnosis, and Management of Chronic Disease. *Am. J. Med. Sci.* **2024**, *368*, 274–286. [[CrossRef](#)] [[PubMed](#)]
76. Glasziou, P.; Irwig, L.; Mant, D. Monitoring in Chronic Disease: A Rational Approach. *BMJ* **2005**, *330*, 901–904. [[CrossRef](#)]

77. Sheng, J.Q.; Hu, P.J.-H.; Liu, X.; Huang, T.-S.; Chen, Y.H. Predictive Analytics for Care and Management of Patients with Acute Diseases: Deep Learning–Based Method to Predict Crucial Complication Phenotypes. *J. Med. Internet Res.* **2021**, *23*, e18372. [\[CrossRef\]](#)
78. Dev, S.; Wang, H.; Nwosu, C.S.; Jain, N.; Veeravalli, B.; John, D. A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthc. Anal.* **2022**, *2*, 100032. [\[CrossRef\]](#)
79. Eack, S.M.; Greeno, C.G.; Lee, B.-J. Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression: Many Cases Are Undetected. *Res. Soc. Work Pract.* **2006**, *16*, 625–631. [\[CrossRef\]](#)
80. Maters, G.A.; Sanderman, R.; Kim, A.Y.; Coyne, J.C. Problems in Cross-Cultural Use of the Hospital Anxiety and Depression Scale: “No Butterflies in the Desert”. *PLoS ONE* **2013**, *8*, e70975. [\[CrossRef\]](#)
81. Moulton, C.D.; Hopkins, C.W.P.; Mohamedali, Z.; Powell, N. Out of Sight, Out of Mind: The Limitations of the Hospital Anxiety and Depression Scale in Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **2019**, *25*, e100. [\[CrossRef\]](#)
82. Pinto-Meza, A.; Serrano-Blanco, A.; Peñarrubia, M.T.; Blanco, E.; Haro, J.M. Assessing Depression in Primary Care with the PHQ-9: Can It Be Carried Out over the Telephone? *J. Gen. Intern. Med.* **2005**, *20*, 738–742. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Lu, J.; Liu, A.; Dong, F.; Gu, F.; Gama, J.; Zhang, G. Learning under Concept Drift: A Review. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 2346–2363. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.