ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 13 (2023) No. 2, pp. 241-257
DOI:10.33168/JSMS.2023.0217

Improved Adam-based Feedforward Deep Neural Network Model for Personalized Asthma Predictions

Radiah Haque [1], Sin-Ban Ho [1], Ian Chai [1] , Adina Abdullah[2]

[1] Faculty of Computing and Informatics, Multimedia University, 63100 Cyberjaya, Malaysia
[2] Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia
sbho@mmu.edu.my

Abstract. A Feedforward Deep Neural Network (FDNN) model contains densely connected layers where backpropagation is applied to calculate the loss function gradients. Optimising the network weights is important to minimise the loss value, hence decreasing the prediction errors and increasing the accuracy rate. Optimisers are used to update the weight values or the learning rate for each weight. Recent studies show that, although Adaptive Moment Estimation (Adam) produces better results in terms of optimising the parameters of the FDNN model, it might lead to poor generalisation performance. Therefore, in this paper, an improved Adam-based FDNN model was built for personalised predictions of asthma. Data transformation techniques (standardisation and normalisation) and regularisation techniques (dropout and max-norm constraint) were applied. Several experimental models were trained, and their prediction performance were compared. The empirical findings reveal that the best prediction results with low loss value can be obtained when the model is trained with standardised inputs and normalised outputs. Moreover, applying dropout (p=0.1) with max-norm (c=3) minimises the generalisation error of the model effectively. The results also show that the improved Adam-based FDNN model (with 2 hidden layers and 50 hidden nodes) produces better performance results with lower prediction loss (Mean Absolute Error (MAE)=0.0409, Mean Squared Error (MSE)=0.0038, and Root Mean Squared Error (RMSE)=0.0618) and higher accuracy rate (96%) than Stochastic Gradient Descent (SGD), Root Mean Squared Propagation (RMSProp), and Adaptive Gradient Descent (AdaGrad). Consequently, the proposed model can be used for personalised asthma predictions based on demography and weather.

Keywords: Feedforward Deep Neural Network, Adam Optimiser, Data Transformation, Regularisation.

1. Introduction

The Deep Neural Network (DNN) algorithm, which is a type of Artificial Neural Networks (ANN), is known as a biologically inspired simulation based on human brain neurons. DNN is commonly used for prediction and pattern recognition (Pandey & Janghel, 2019). Typically, a DNN model consists of an input layer, which contains the input values obtained from a given dataset, and an output layer, which contains the target values. In addition, the DNN model includes multiple hidden layers and several hidden nodes in each hidden layer (Mukherjee et al., 2021). Generally, to evaluate the prediction performance of the DNN model, a loss function is used, which calculates the errors between the actual output values and predicted values (Zhang et al., 2021). In particular, the loss function measures how well the model performed, where a low loss value indicates a good model that provides accurate forecasting results.

A feedforward network is a type of DNN architecture in which connections between the nodes do not form a cycle (Bisandu et al., 2021). The Feedforward DNN (FDNN) model contains densely connected layers wherein information from the input layer moves in a forward direction through the hidden layers to reach the output layer (Zhang et al., 2021). During the initial stage of the FDNN model training, a set of random weight values are assigned to the received inputs in each hidden node. Subsequently, the sum of the weighted input values is passed as the node element through an activation function in the hidden layers of the model (Apicella et al., 2021). In this case, it is important to update and fine tune the weight values in order to identify the optimal configuration of the FDNN model, which in turn helps to minimise the loss function and find the lowest loss value possible (Zhou et al., 2021). As such, updating the network weights, or the step size of each weight (learning rate), is considered an optimisation problem.

Backpropagation is a common algorithm applied primarily to train feedforward models. It is a differentiation method for computing the gradients of the loss function with regards to the model internal weights (Lyu et al., 2022). In backpropagation, the gradient errors are calculated iteratively backwards starting from the output layer, as seen in Figure 1. The main goal of the backpropagation algorithm is to identify how the network weights should be updated to minimise the loss function (Wright et al., 2022). In this case, the FDNN model applies an optimisation algorithm (commonly known as an optimiser), which automatically updates the parameters during the backpropagation process by fine tuning the weights or learning rates (Guo et al., 2021). Consequently, values that minimise the loss function are used for model testing and validation, thus enhancing further the prediction performance of the FDNN model.

Traditionally, the gradient descent optimiser is applied during the backpropagation process to optimise the network weights (Song et al., 2020). Nonetheless, it has been observed that learning rate optimisers, such as Adaptive Moment Estimation (Adam) algorithm, produce better results in terms of optimising the parameters of the FDNN model and locating the global minima (i.e., a point where the loss value is globally minimum) (Yi et al., 2020). In this case, the optimiser updates the parameters in the current iteration so that the subsequent iteration reduces the loss value, hence navigating down the gradient of the loss function during the backpropagation process (Zhang et al., 2021). Although the Adam optimiser is widely used for training deep learning models, the Adam-based FDNN model sometimes produces a high generalisation error (i.e., validation loss) (Zhou et al., 2020), especially when conducting regression analysis.

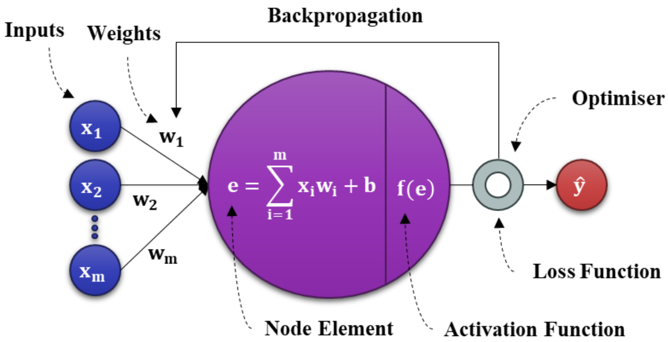


Fig. 1: Basic process of FDNN with backpropagation.

Therefore, in this paper, an Adam-based FDNN model is developed and applied to the asthma dataset for regression analysis. The goal is to develop an optimised predictive model that can provide accurate forecasting of asthma exacerbations based on personalised weather triggers with minimum prediction loss and generalisation error. The main objectives of this research work are:

1. To apply data transformation techniques, including standardisation and normalisation, to reduce the prediction loss and improve the performance of the of the Adam-based FDNN model.
2. To apply model regularisation techniques, including dropout regularisation and max-norm regularisation, to minimise the generalisation error of the Adam-based FDNN model.
3. To compare the performance of the improved Adam-based FDNN model with other learning rate optimisers, in terms of validation errors, explained variance, and overall prediction accuracy.

2. Background Study

2.1. Learning Rate Optimisers

Learning rate optimisers are responsible for tuning the learning rates of the FDNN model (Zhou et al., 2020). The Stochastic Gradient Descent (SGD) algorithm is regarded as an approximation of the traditional gradient descent optimiser, which iteratively updates the learning rates of the model. It has been observed that substantial problems could arise when employing SGD with FDNN, such as oscillations of the learning rates in the later training stages of the model, which often lead to increased variance issues (Sun et al., 2020). Recently, several adaptive learning rate optimisers have been proposed, such as Adaptive Gradient Descent (AdaGrad) and Root Mean Square Propagation (RMSProp) algorithms. One of the essential properties of AdaGrad is changing the learning rate for every weight value in each epoch (Defossez et al., 2020). However, in some cases, the learning rates continue to decrease rapidly, resulting in vanishing gradients that might cause the model to completely stop from further training (Sun et al., 2020). In this case, RMSProp utilises the averaged squared gradients to normalise them, thus preventing the vanishing gradient problem (Xu et al., 2021). Nevertheless, this could cause repetitive updates of the same learning rates, leading to slower training and convergence of the model.

Fortunately, Adam, which is a common adaptive learning rate optimiser, solves the gradient vanishing problem of AdaGrad using bounded gradients (Defossez et al., 2020). Unlike RMSProp, Adam is renowned for its convergence speed when training FDNN models. Moreover, the Adam optimiser works effectively with its default configuration properties and parameter values. This gives Adam a significant advantage over other optimisers, especially SGD, which often requires applying momentum to obtain better results (Yi et al., 2020). Selected studies, however, have argued that the Adam optimiser could lead to worse generalisation performance than SGD (Sun et al., 2020 and Chen et al., 2020). As a result, although the Adam-based FDNN model mostly produces good prediction results, it has the potential to generate high generalisation errors when applying the model on unseen data.

Therefore, in this work, various optimisation experiments are conducted to improve the Adam-based FDNN model for regression analysis. In particular, the impact of data transformation and regularisation techniques on the optimisation performance are investigated this paper. Consequently, the goal is to develop an optimised model that can provide predictions of asthma exacerbations based on personalised weather triggers with low errors and high accuracy.

2.2. Data Transformation Techniques

Data transformation is an important process that takes place before training the predictive model. It is a recommended pre-processing step for DNN models, where the values of the variables in the experimental dataset from different dynamic ranges are scaled and transformed into a specific range (Raju et al., 2020). The two most common data transformation techniques are: normalisation and standardisation. Normalisation (also known as Min-Max scaling) rescales the data with a distribution value between 0 and 1. Meanwhile, standardisation (also known as Z-score scaling) rescales the data by ensuring the mean and the standard deviation are 0 and 1, respectively (Henderi, 2021).

Generally, feature scaling is considered essential when developing a predictive model using a learning algorithm that applies optimisation during the backpropagation process, such as FDNN (Singh & Singh, 2020). This is because input features typically have values with varied units and different scales, which can increase the difficulty of the prediction problem being modelled. Moreover, unscaled input features can result in a slow learning process of the model. In particular, scaling the feature values has a significant positive impact on the model stability and prediction accuracy (Raju et al., 2020). Normally, for classification problems, scaling target values is not required (Singh & Singh, 2020). However, for regression analysis, applying data transformation techniques is crucial to train the model with scaled output variables, which, in particular can improve the performance of the optimiser.

Given the use of the loss function in the FDNN model to calculate the errors between the original outputs and predicted outputs and the use of backpropagation to calculate the gradient errors, the scale of the output values for training the model are an important factor. In this case, reducing the distribution scale of the output values can assist in minimising the size of the gradient errors. On the other hand, unscaled output values can result in exploding gradients causing the learning process to fail. Unfortunately, there is a lack of empirical evidence in the literature that investigates the impact of data transformation on the performance of the Adam optimiser. Therefore, in this paper, data transformation techniques are applied to train the Adam-based FDNN model for regression analysis.

2.3. Model Regularisation Methods

Typically, for a supervised prediction problem, the goal is to train the model using the training set to learn the relationships between the input and output values in the experimental dataset, and then apply this learning to predict the target outputs based on the input features using the test set (Zhang et al., 2021). For the Adam-based FDNN model, determining the generalisation performance is essential to identify how accurately the algorithm predicts the target outputs for previously unseen data. This is important because a high generalisation error often leads to overfitting (Bejani & Ghatee, 2021). Regularisation is an improvement technique that applies small modifications to the predictive model such that it generalises well on unseen data (Hu et al., 2020). As such, regularisation is widely used to prevent overfitting of the model by minimising the gap between training loss and validation loss.

There are several regularisation techniques exist for modifying the loss function or training algorithm. The most common technique for modifying the loss function is L2 regularisation, also called weight decay, which works by adding a penalty term to the loss function in order to decrease the weight values (Saud & Shakya, 2021). Meanwhile, the most common technique for modifying the training algorithm is dropout regularisation (Srivastava et al., 2014). By applying dropout, connections between the nodes in the hidden layers of the FDNN model are randomly dropped in each epoch based on the given dropout ratio, as seen in Fig. 2. As a result, training iterations are performed with a varied set of input elements of the existing nodes.

In many cases, dropout is combined with another type of regularisation for constraining the network weights, known as max-norm regularisation. It constrains the norm of the weight vector at each node in the hidden layers to be upper bounded by a specified constant, hence preventing the network weights from growing to a large value. In fact, dropout regularisation is known for showing significant improvements when applying it with max-norm regularisation, which often produces the lowest generalisation error (Soumare et al., 2021). Therefore, in this paper, dropout is used with the max-norm regularisation method to reduce the generalisation error of the Adam-based FDNN model and minimise the validation loss.

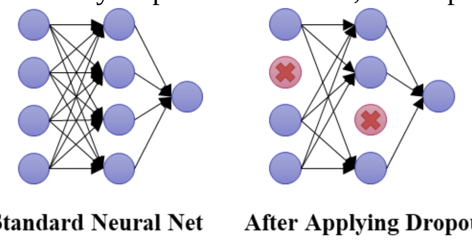


Fig. 2: FDNN model with and without dropout.

3. Methods

3.1. Asthma Dataset

The asthma dataset employed in this work was collected from an mHealth application, known as Weather Asthma (WEA) (Haque et al., 2021a, Haque et al., 2021b and Ho et al., 2020). The dataset consists of ten input variables with five weather features (temperature, humidity, air pressure, wind speed, and UV index) and five demography features (age, gender, location, outdoor job, and outdoor activities). The real-time values of the weather features in the asthma dataset were collected from the location of individual WEA application users, which represent the specific weather conditions that might exacerbate a particular user’s asthma and lead to acute attacks. Meanwhile, the demography features were collected from individual users upon their registration in the WEA application, which were used for acquiring personalised asthma predictions.

The asthma dataset also contains the Asthma Control Test (ACT) scores as the target output. Users can conduct ACTs to report their asthma severity level by answering five close-ended questions through the WEA application. Each question scores from 1 to 5, and the range of the total score for an ACT is 5 to 25. According to the Global Initiative for Asthma (GINA), a low ACT score indicates a high chance of asthma exacerbation and high ACT score indicates a low chance of asthma exacerbation (Reddel et al., 2022). As such, each record in the dataset consists of an ACT score along with the demography information of the user who submitted the ACT, and the weather information of the exact day and time when the ACT was submitted. Table 1 describes variables in the asthma dataset, which consists of 2020 records with 10 input features and 1 target output (ACT_score).

Table 1: Asthma Dataset Variables

Variable	**Description**
ACT_score	Total score of an ACT submitted by a user.
Temperature	Temperature value in degree Celsius (°C).
Humidity	Humidity percentage (%) value.
Air_pressure	Air pressure value in hectopascals (hPa).
Wind_speed	Wind speed in meters per second (m/s).

UV_index	UV index range: Low and Extreme.
Location	Location of the user (city name).
Age_group	Age group of the user (18 and below, 19-30, 31-40, 41-50, Above 50).
Gender	Gender of the user (Male, Female, Other).
Outdoor_job	Frequency of outdoor job for the user (Frequently, Occasionally, Rarely).
Outdoor_activities	Likelihood of outdoor activates for the user (Likely, Neither, Not Likely).

Afterwards, the dataset was divided into a training set and a test set. The training set consists of 1616 samples (80%) and the test set consists of 404 samples (20%). Before training the model on the asthma dataset, all the categorial variables need to be converted to numeric values. A popular way of doing this is by applying the label encoder method, which encodes the categorial labels for each variable with a number between 0 and n-1, where n is the number of labels. When a label repeats within the variable, it takes the same number as previously assigned (Zhang et al., 2021). The asthma dataset contains 6 categorial variables, and they were converted into numerical representations in this manner.

3.2. Adam-based FDNN

In this work, the Adam-based FDNN model was developed for regression analysis. To build the optimal architecture of the model, the hyperparameters of the model were tuned using Grid Search (GS). GS is a simple optimisation algorithm for tuning model hyperparameters in order to find their optimum values. The basic process of GS includes three primary steps: (1) select a set of candidate values for each hyperparameter from a predefined search space, (2) train various models separately using all possible combinations of the selected values, and (3) calculate the loss value for each model and display the hyperparameter values of the model with the lowest loss. In this case, Mean Squared Error (MSE) was applied as the loss function to train the Adam-based FDNN model. Table 2 represents the optimum hyperparameter values.

Table 2: Optimum Values of Model Hyperparameters

Hyperparameter	**Optimum Value**
No. of hidden layers	2
No. of hidden nodes	50
Bach size	10
No. of epochs	100
Weight initialiser	Normal

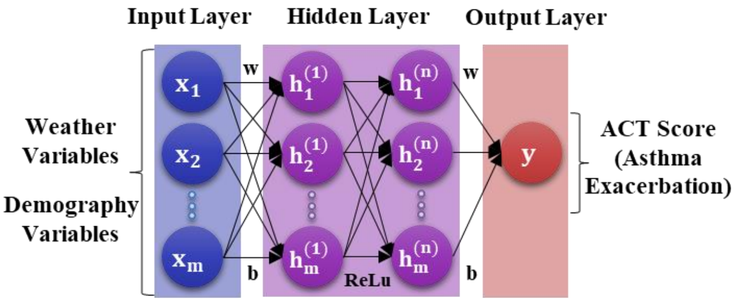


Fig. 3: Adam-based FDNN model architecture.

Figure 3 demonstrates the architecture of the Adam-based FDNN model. The Rectified Linear Unit (ReLU) was used as the activation function, which provides simple nonlinear transformations in the hidden layers of the model. Considering the element (e), ReLU is the maximum of e and 0 (Zhang et al., 2021). The following equations were used to predict the target outputs using the FDNN model:

$$w = [w_1, w_2, \dots, w_{10}]$$

$$f(e) = \begin{cases} 0 & \text{for } e < 0 \\ e & \text{for } e \geq 0 \end{cases}$$

$$e = \sum_{i=1}^{10} xw + b$$

$$\hat{y} = f(e)$$

where x is the input value, $\hat{y}$ is the target output, w is the weight value, b is the bias, e is the node element, and f(e) is the activation function.

3.3. Evaluation Metrics

To evaluate the prediction performance of the Adam-based FDNN model for regression analysis, error metrics were used as the evaluation metrics. In this case, MSE was applied to calculate the training and validation loss of each experimental model. Two more error metrics were used, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). MAE is the basic measure of errors applied to calculate the absolute difference between the original target outputs and the predicted outputs. MSE is used to calculate the squared difference between the original target outputs and the predicted outputs. The key difference between MAE and MSE is that the latter can measure the outliers in the dataset. RMSE is the square root of MSE, which is measured with the same unit as the output values. The following equations were used to calculate the prediction errors of the Adam-based FDNN model:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

where n is the number of records in the dataset, y is the actual output values, and $\hat{y}$ is the predicted values.

For further performance analysis, Explained Variance Score (EVS) was calculated. The EVS indicates how well the model captured the variance in the data that exists in the nonlinear dataset (Zhang et al., 2021). A high EVS indicates a good prediction performance. The following equation was used to calculate the EVS of the Adam-based FDNN model:

$$EVS = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}}$$

where Var is the variance (i.e., the square of the standard deviation), y is the original output values, and $\hat{y}$ is the predicted values.

The overall accuracy (ACC) was also measured to identify how accurately the model predicted the target outputs. For regression analysis, ACC can be calculated using the Mean Absolute Percentage Error (MAPE) (Zhang et al., 2021). The following equations were used to calculate the MAPE and ACC rate of the Adam-based FDNN model:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$\mathrm{ACC} = 100 \times (1 - \mathrm{MAPE})$$

where n is the number of records in the dataset, y is the actual output values, and ŷ is the predicted values.

4. Results and Discussion

All the experiments in this work have been deployed using Python 3 and Anaconda on Jupyter notebook. In this work, several experimental models were constructed to investigate the impact of data transformation and regularisation techniques on the performance of the Adam-based FDNN model, in terms of prediction loss and generalisation error. The experiments were applied in three stages to achieve the objective of this research work. The following sections demonstrate the results of the predictive modelling in each stage.

4.1. Standardisation/Normalisation Results

In the first experimentation stage, the first objective of this research work was addressed, which is to apply data transformation techniques to reduce the prediction loss and improve the performance of the Adam-based FDNN model. In this case, both standardisation and normalisation techniques were applied to rescale the variables in the asthma dataset. The goal is to identify which technique, or combination of techniques, provides the best prediction performance of the Adam-based FDNN model. As such, 6 predictive models were constructed. The first model was trained using standardised input variables (Stand_In). The second model was trained using normalised input variables (Norm_In). The output variables in both of these models were not scaled. The third model was trained using standardised input variables and standardised output variables (Stand_In_Stand_Out). The fourth model was trained using normalised input variables and normalised output variables (Norm_In_Norm_Out). The fifth model was trained using standardised input variables and normalised output variables (Stand_In_Norm_Out). The sixth model was trained using normalised input variables and standardised output variables (Norm_In_Stand_Out).

Table 3 summarises the performance of all 6 models, and Figure 4 demonstrates the training loss and validation loss. It can be seen that the fourth model, which was trained using standardised input variables and normalised output variables, achieved the best prediction results, with MAE = 0.0441, MSE = 0.0054, RMSE = 0.0733, EVS = 0.94, and ACC = 95%. Meanwhile, the worst prediction results were produced by the first and second models, which were trained using unscaled output variables. Therefore, for regression analysis, it is important to scale the output values to reduce the prediction loss and improve the performance of the Adam-based FDNN model. Moreover, the findings reveal that different data transformation techniques can be used to scale the input and output variables to train the same model. Subsequently, to perform the experiments in the second stage, all the input variables in the asthma dataset were standardised using the StandardScaler() function, and the output values were normalised using the MinMaxScaler() function from scikit-learn library.

Table 3: Standardisation and Normalisation Results

Model	MAE	MSE	RMSE	EVS	ACC
Stand_In	1.0810	2.2263	1.4921	0.90	93%
Norm_In	1.8500	5.8222	2.4129	0.75	87%
Stand_In_Stand_Out	0.1611	0.0741	0.2721	0.93	95%
Norm_In_Norm_Out	0.0527	0.0061	0.0783	0.92	94%
Stand_In_Norm_Out	**0.0441**	**0.0054**	**0.0733**	**0.94**	**95%**
Norm_In_Stand_Out	0.2234	0.1019	0.3192	0.90	93%

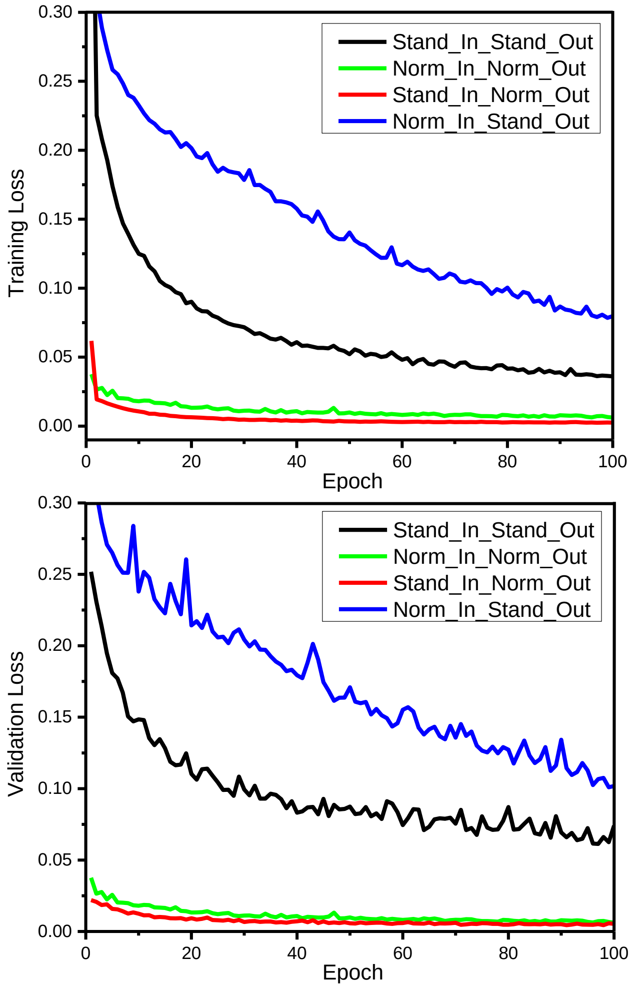


Fig. 4: Training vs. validation loss.

4.2. Regularisation Results

In the second experimentation stage, the second objective of this research work was addressed, which is to apply model regularisation techniques, including dropout regularisation and max-norm regularisation, to minimise the generalisation error of the Adam-based FDNN model. In this case, dropout was applied to the hidden layers of the network with dropout value p = {0.1, 0.2, 0.3, 0.4, 0.5}. The p value represents the probability ratio for dropping the nodes in the hidden layers. For example, p = 0.1 denotes that 10% of the hidden nodes (i.e., 5 nodes) were dropped in each training epoch randomly. The goal is to identify which dropout ratio is required to minimise the generalisation error of the Adam-based FDNN model. As such, 5 predictive models were trained using dropout ratio 10%, 20%, 30%, 40%, and 50%. Moreover, max-norm regularisation was applied in the hidden layers of each model with a constant value c = 3. The c value denotes that the network weights in the hidden nodes were constrained to have a norm less than or equal to 3.

Table 4: Dropout Regularisation Results

Model	**MAE**	**MSE**	**RMSE**	**EVS**	**ACC**
Dropout_01	**0.0409**	**0.0038**	**0.0618**	**0.95**	**96%**
Dropout_02	0.0449	0.0051	0.0711	0.94	95%
Dropout_03	0.0512	0.0054	0.0735	0.93	94%
Dropout_04	0.0601	0.0072	0.0837	0.92	93%
Dropout_05	0.0785	0.0103	0.1016	0.87	91%

Table 4 summarises the performance of all 5 models, and Figure 5 demonstrates the training loss and validation loss of each predictive model. It can be seen that, by applying dropout with p = 0.1, the Adam-based FDNN model achieved the best prediction results, with MAE = 0.0409, MSE = 0.0038, RMSE = 0.0618, EVS = 0.95, and ACC = 96%. For further analysis on the generalisation performance, the percent difference between the training loss and validation loss was calculated for each predictive model, as seen in Table 5. For comparison purpose, the prediction results of the best performing model from the first experimentation stage that did not apply dropout were also added in Table 4, Table 5, and Figure 5.

It can be recognised that, without applying dropout regularisation, the model produced the best training loss value. However, the validation loss was significantly high. In fact, compared to the other models, it generated the highest difference ratio between training loss and validation loss, around 70%. Meanwhile, the lowest difference ratio was observed when dropout was applied with p = 0.5. Nevertheless, the model produced worse prediction results than all other models. This is probably because the architecture of the Adam-based FDNN model used in this work is relatively small with only 2 hidden layers. In effect, applying dropout with p = 0.5 forces the model to drop 50% of its hidden nodes along with their respective input elements, which restricts its learning ability. However, since dropout was not applied during the testing process, all the input elements containing new information were passed to the output layer, leading to a high validation loss. The next lowest difference ratio with around 30% was generated by the first model with p = 0.1, which also produced the lowest generalisation error, thus improving the generalisation performance of the Adam-based FDNN model.

Table 5: Generalisation Error Results

Model	**Train_Loss**	**Val_Loss**	**Percent_Diff**
No_Dropout	**0.0026**	0.0054	70%
Dropout_01	0.0028	**0.0038**	30%
Dropout_02	0.0031	0.0051	47%
Dropout_03	0.0036	0.0054	40%
Dropout_04	0.0050	0.0072	36%
Dropout_05	0.0091	0.0103	**12%**

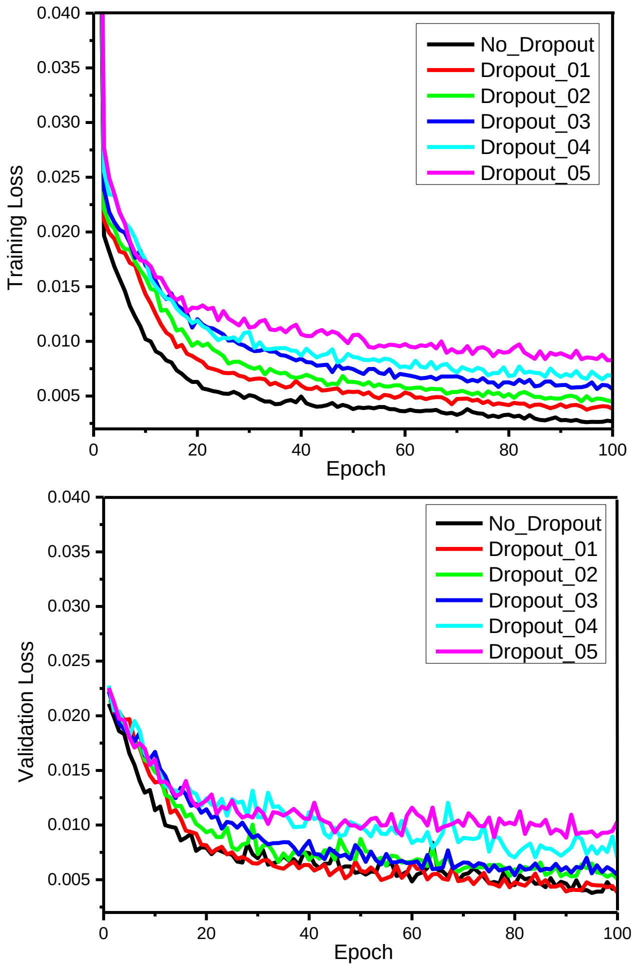


Fig. 5: Training vs. validation loss.

High generalisation error is a common problem that often occurs while training deep learning-based models. In this regard, previous work has suggested applying other normalisation and regularisation techniques, such as Batch Normalisation (BN) (Roburin et al., 2020) and scaling of weights using L2 regularisation (Loshchilov & Hutter, 2019), to improve the training and performance of the DNN models. In this work, both BN and L2 regularisation were applied (with and without dropout) on the Adam-based FDNN model. As such, 4 more predictive models were constructed. The first model was trained with a BN layer that was added after the activation layer (Batch_Norm). In the second model, L2 regularisation was used (L2_Norm). The third model combined both dropout and BN (Dropout+BN), and the fourth model was trained using dropout and L2 regularisation (Dropout+L2).

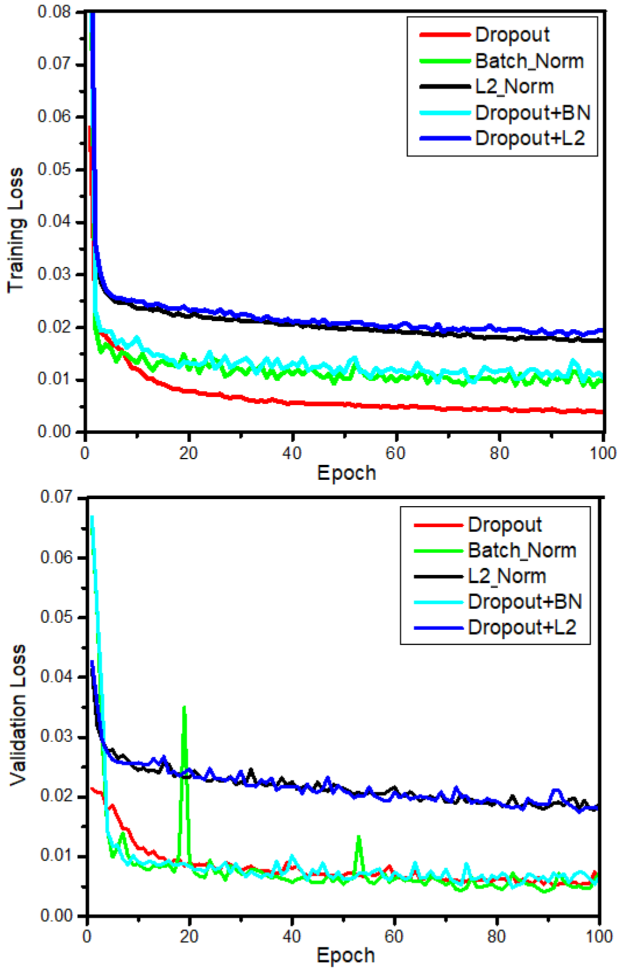


Fig. 6: Training vs. validation loss.

Table 6: Model Regularisation Results

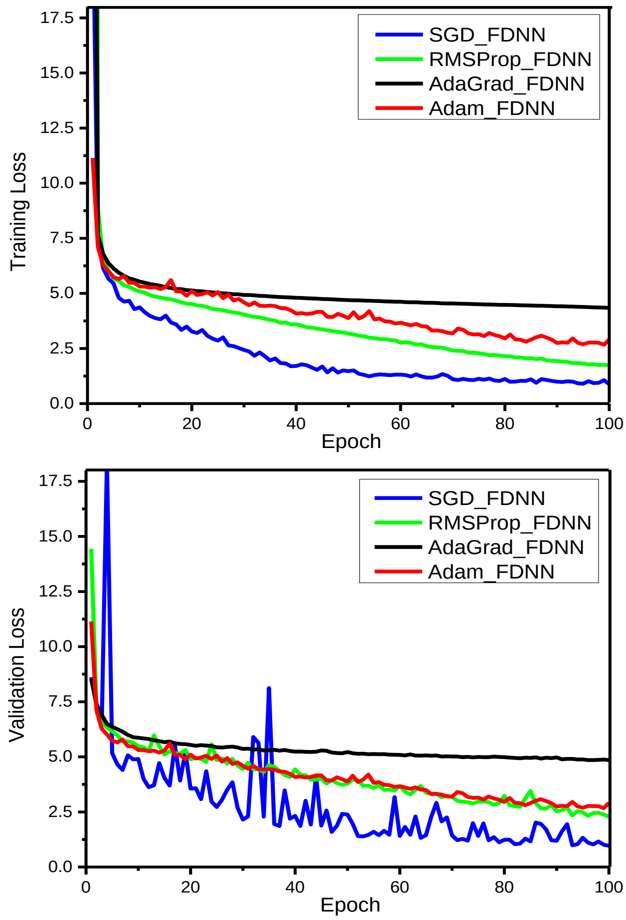
Model	MAE	MSE	RMSE	EVS	ACC
Dropout	**0.0409**	**0.0038**	**0.0618**	**0.95**	**96%**
Batch_Norm	0.0644	0.0075	0.0869	0.90	92%
L2_Norm	0.0986	0.0154	0.1241	0.82	88%
Dropout+BN	0.0505	0.0067	0.0816	0.91	92%
Dropout+L2	0.0927	0.0141	0.1188	0.82	89%

Table 6 summarises the performance of all 4 models, and Figure 6 demonstrates the training loss and validation loss. For comparison purpose, the prediction results of the best performing model with dropout (p = 0.1) were also added in Table 6 and Figure 6. In this case, it can be seen that the Adam-based FDNN model performed best when only dropout was applied, followed by BN. However, high variance was observed when the model was trained using BN. In this case, although applying dropout with BN reduced variance, it did not enhance the generalisation performance of the Adam-based FDNN model. This issue was previously examined in a study (Li et al., 2019) where it was revealed that dropout and BN can lead to a worse performance when they are combined together. This is because dropout is applied during training only, hence the variance of a specific node is shifted during testing. On the other hand, BN maintains its statistical variance in the test phase. Meanwhile, Table 6 shows that L2 produced the highest generalisation error and lowest accuracy rate with around 88%. In this regard, a study (Loshchilov & Hutter, 2019) argued that, unlike SGD, adaptive learning algorithms (such as Adam) do not apply weight decay when updating model parameters, thus making L2 regularisation less effective in Adam-based models.

4.3. Comparative Analysis and Discussion

In the third experimentation stage, the final research objective of this research work was addressed, which is to compare the performance of the improved Adam-based FDNN model with other learning rate optimisers. In this case, the FDNN model was trained using SGD, RMSProp, and AdaGrad. The goal is to find out if standardisation, normalisation, and regularisation also improve the performance of these optimisers, as it was observed with the Adam optimiser during the previous experimentation stages. As such, 6 predictive models were constructed. The first 3 models (SGD_FDNN, RMSProp_FDNN, and AdaGrad_FDNN) were trained using the SGD, RMSProp, and AdaGrad optimisers, respectively. Subsequently, each optimiser was used again to train the subsequent 3 models (SGD_FDNN_StNrRg, RMSProp_FDNN_StNrRg, and AdaGrad_FDNN_StNrRg), with standardised inputs, normalised outputs, and dropout with p = 0.1.

Table 7 summarises the performance of all 6 models, and Figure 7 demonstrates the training loss and validation loss of each predictive model. For the purpose of comparison, the prediction results of the original Adam-based FDNN model (Adam_FDNN) and the improved model (Adam_FDNN_StNrRg) obtained from the previous experimentation stage were also added in Table 7 and Figure 7. It can be seen that, when data transformation and regularisation techniques were not applied, SGD provided better prediction results than the other optimisers, with MAE = 0.6640, MSE = 0.9652, RMSE = 0.9825, EVS = 0.94, and ACC = 95%.



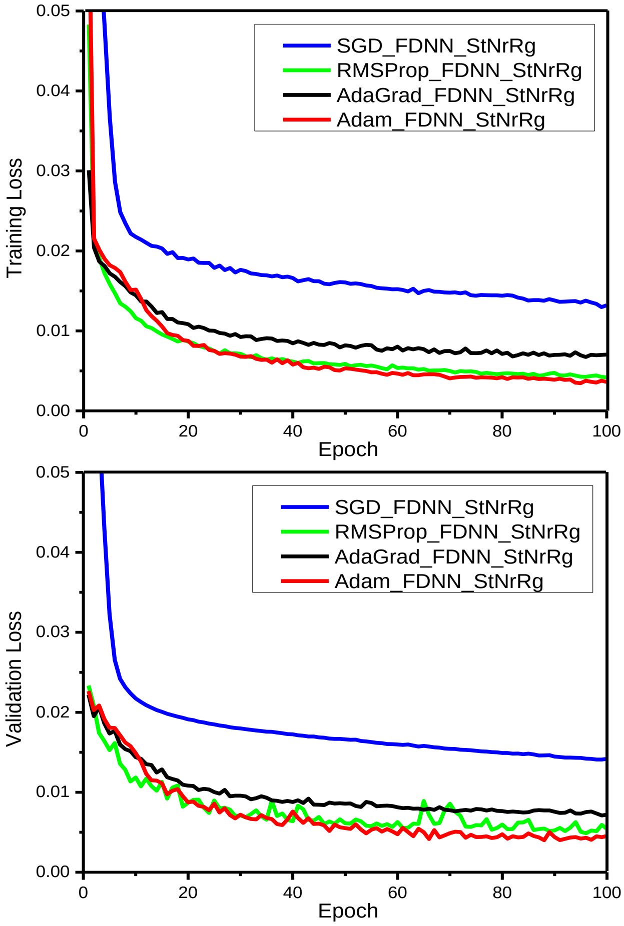


Fig. 7: Training vs. validation loss.

Table 7: Model Optimisation Results

Model	**MAE**	**MSE**	**RMSE**	**EVS**	**ACC**
SGD_FDNN	0.6640	0.9652	0.9825	0.94	95%
RMSProp_FDNN	1.0772	2.2833	1.5111	0.90	93%
AdaGrad_FDNN	1.6924	4.8536	2.2031	0.79	88%
Adam_FDNN	1.2072	2.9093	1.7057	0.88	92%
SGD_FDNN_StNrRg	0.0906	0.0142	0.1191	0.82	89%
RMSProp_FDNN_StNrRg	0.0518	0.0054	0.0735	0.94	95%
AdaGrad_FDNN_StNrRg	0.0615	0.0072	0.0848	0.91	93%
Adam_FDNN_StNrRg	**0.0409**	**0.0038**	**0.0618**	**0.95**	**96%**

However, high variance was observed with this model, which denotes that although it was able to represent the training set accurately, overfitting might appear when applying the model on unseen data. Meanwhile, when standardisation, normalisation, and regularisation techniques were applied, SGD produced the worst performance results. Noticeably, the improved Adam-based FDNN model has the best performance results with lowest loss and highest accuracy. This indicates that, with its default parameter values, the Adam optimiser outperforms all the other learning rate optimisers when trained with scaled data and with dropout and max-norm regularisation.

5. Conclusion

In this paper, an FDNN model was developed for personalised predictions of asthma. Typically, FDNN applies backpropagation to calculate the gradients of the loss function with regards to the internal weights of the model. Optimising the network weights is important to minimise the loss value, hence decreasing prediction errors. Therefore, optimisers are applied to update the weight values or the learning rate for each weight. SGD, AdaGrad, RMSProp, and Adam are common learning rate optimisers used for training the FDNN model. It was recognised that, although Adam produces good results in terms of optimising the model parameters, it might lead to poor generalisation performance. Therefore, in this paper, an Adam-based FDNN model was built for regression analysis using the asthma dataset. To improve the performance of the model, data transformation and regularisation techniques were applied. Several experimental models were trained, and their prediction results were compared. The empirical findings reveal that the best prediction performance with the lowest loss can be obtained when the model is trained with standardised inputs and normalised outputs. Moreover, applying dropout (p=0.1) with max-norm (c=3) minimises the generalisation error of the model effectively. The experimental work in this paper provides further evidence that the Adam optimiser produces better generalisation results when a dropout layer is added in the model than BN and L2 regularisation. The results also show that the improved Adam-based FDNN model provides the best performance results with the lowest validation loss and highest accuracy than other learning rate optimisers. Consequently, the proposed model can be applied to accurately predict asthma exacerbations based on personalised weather triggers with low errors.

Acknowledgments

The authors are grateful for this research project was funded by the Fundamental Research Grant Scheme (FRGS/1/2019/SS06/MMU/02/4).

References

Apicella, A., Donnarumma, F., Isgrò, F., & Prevete, R. (2021). A survey on modern trainable activation functions. In *Neural Networks*,138. https://doi.org/10.1016/j.neunet.2021.01.026

Bala Bisandu, D., Salih Homaid, M., Moulitsas, I., & Filippone, S. (2021). A Deep Feedforward Neural Network and Shallow Architectures Effectiveness Comparison: Flight Delays Classification Perspective. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3505711.3505712

Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, 54(8). https://doi.org/10.1007/s10462-021-09975-1

Chen, J., Zhou, D., Tang, Y., Yang, Z., Cao, Y., & Gu, Q. (2020). Closing the generalization gap of adaptive gradient methods in training deep neural networks. *IJCAI International Joint Conference on Artificial Intelligence*, *2021-January*. https://doi.org/10.24963/ijcai.2020/452

Defossez, A., Bottou, L., Bach, F., & Usunier, N. (2020). On the convergence of adam and adagrad. *ArXiv*.

Guo, Z., Xu, Y., Yin, W., Jin, R., & Yang, T. (2021). *A Novel Convergence Analysis for Algorithms of the Adam Family*. 3. http://arxiv.org/abs/2112.03459

Haque, R., Ho, S.-B., Chai, I., & Abdullah, A. (2021). Optimised deep neural network model to predict asthma exacerbation based on personalised weather triggers. *F1000Research*, 10. https://doi.org/10.12688/f1000research.73026.1

Haque, R., Ho, S.-B., Chai, I., Teoh, C.-W., Abdullah, A., Tan, C.-H., & Dollmat, K. S. (2021).

Intelligent Health Informatics with Personalisation in Weather-Based Healthcare Using Machine Learning. In *Lecture Notes on Data Engineering and Communications Technologies*,72. https://doi.org/10.1007/978-3-030-70713-2_4

Henderi, H. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *IJIIS: International Journal of Informatics and Information Systems*, 4(1). https://doi.org/10.47738/ijiis.v4i1.73

Ho, S.-B., Haque, R., Chai, I., Teoh, C.-W., Abdullah, A., Tan, C.-H., & Dollmat, K. S. (2020). Integrating Mobile Devices with Cohort Analysis into Personalised Weather-Based Healthcare. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),12496 LNAI*. https://doi.org/10.1007/978-3-030-63007-2_47

Hu, T., Wang, W., Lin, C., & Cheng, G. (2020). *Regularization Matters: A Nonparametric Perspective on Overparametrized Neural Network*. 130. http://arxiv.org/abs/2007.02486

Li, X., Chen, S., Hu, X., & Yang, J. (2019). Understanding the disharmony between dropout and batch normalization by variance shift. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June*. https://doi.org/10.1109/CVPR.2019.00279

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. 7th International Conference *on Learning Representations, ICLR 2019*.

Lyu, Z., Yu, Y., Samali, B., Rashidi, M., Mohammadi, M., Nguyen, T. N., & Nguyen, A. (2022). Back-Propagation Neural Network Optimized by K-Fold Cross-Validation for Prediction of Torsional Strength of Reinforced Concrete Beam. *Materials*, 15(4). https://doi.org/10.3390/ma15041477

Mukherjee, H., Ghosh, S., Dhar, A., Obaidullah, S. M., Santosh, K. C., & Roy, K. (2021). Deep neural network to detect COVID-19: one architecture for both CT Scans and Chest X-rays. *Applied Intelligence*, 51(5). https://doi.org/10.1007/s10489-020-01943-6

Pandey, S. K., & Janghel, R. R. (2019). Recent Deep Learning Techniques, Challenges and Its Applications for Medical Healthcare System: A Review. *Neural Processing Letters*, 50(2). https://doi.org/10.1007/s11063-018-09976-2

Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*. https://doi.org/10.1109/ICSSIT48917.2020.9214160

Reddel, H. K., Bacharier, L. B., Bateman, E. D., Brightling, C. E., Brusselle, G. G., Buhl, R., Cruz, A. A., Duijts, L., Drazen, J. M., FitzGerald, J. M., Fleming, L. J., Inoue, H., Ko, F. W., Krishnan, J. A., Levy, M. L., Lin, J., Mortimer, K., Pitrez, P. M., Sheikh, A., … Boulet, L. P. (2022). Global Initiative for Asthma Strategy 2021 Executive Summary and Rationale for Key Changes. *American Journal of Respiratory and Critical Care Medicine*, 205(1). https://doi.org/10.1164/rccm.202109-2205PP

Roburin, S., de Mont-Marin, Y., Bursuc, A., Marlet, R., Pérez, P., & Aubry, M. (2020). Spherical perspective on learning with batch norm. In *arXiv*.

Saud, A. S., & Shakya, S. (2021). Analysis of L2 Regularization Hyper Parameter for Stock Price Prediction. *Journal of Institute of Science and Technology*, 26(1). https://doi.org/10.3126/jist.v26i1.37830

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*. https://doi.org/10.1016/j.asoc.2019.105524

Song, S., Thakkar, O., & Thakurta, A. (2020). Characterizing Private Clipped Gradient Descent on Convex Generalized Linear Problems. In *arXiv*.

Soumare, H., Benkahla, A., & Gmati, N. (2021). Deep learning regularization techniques to genomics data. *Array*, 11. https://doi.org/10.1016/j.array.2021.100068

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15.

Sun, S., Cao, Z., Zhu, H., & Zhao, J. (2020). A Survey of Optimization Methods from a Machine Learning Perspective. *IEEE Transactions on Cybernetics*, 50(8). https://doi.org/10.1109/TCYB.2019.2950779

Wright, L. G., Onodera, T., Stein, M. M., Wang, T., Schachter, D. T., Hu, Z., & McMahon, P. L. (2022). Deep physical neural networks trained with backpropagation. *Nature*, 601(7894). https://doi.org/10.1038/s41586-021-04223-6

Xu, D., Zhang, S., Zhang, H., & Mandic, D. P. (2021). Convergence of the RMSProp deep learning method with penalty for nonconvex optimization. *Neural Networks*, 139. https://doi.org/10.1016/j.neunet.2021.02.011

Yi, D., Ahn, J., & Ji, S. (2020). An effective optimization method for machine learning based on ADAM. *Applied Sciences (Switzerland)*, 10(3). https://doi.org/10.3390/app10031073

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into Deep Learning Release 0.17.5. In *Journal of the American College of Radiology*, 17(5).

Zhou, P., Feng, J., Ma, C., Xiong, C., Hoi, S., & Weinan, E. (2020). Towards theoretically understanding why SGD generalizes better than ADAM in deep learning. *Advances in Neural Information Processing Systems*, *2020-December*.

Zhou, X., Zhang, W., Chen, Z., Diao, S., & Zhang, T. (2021). Efficient Neural Network Training via Forward and Backward Propagation Sparsification. *Advances in Neural Information Processing Systems*, 19(NeurIPS), 15216–15229.