

Article

Not peer-reviewed version

Predicting Multiple Sclerosis Disease Activity from Longitudinal MRI and EDSS Data Using Multimodal Deep Learning: A Pilot Study

[Marina Barulina](#)^{*}, Oleg Samsonov, Ivan Chechkov, [Alexander Zakharov](#), Elena Khivintseva

Posted Date: 22 April 2026

doi: 10.20944/preprints202604.1490.v1

Keywords: multiple sclerosis; disease activity prediction; deep learning; transfer learning; multimodal data; ResNet; recurrent neural network; clinical decision support



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

Predicting Multiple Sclerosis Disease Activity from Longitudinal MRI and EDSS Data Using Multimodal Deep Learning: A Pilot Study

Marina Barulina ^{1,2,*}, Oleg Samsonov ², Ivan Chechkov ², Alexander Zakharov ³ and Elena Khivintseva ³

¹ Saint Petersburg State University, Saint-Petersburg, Russia

² Institute of Physical and Mathematical Sciences, Perm State University, Bukireva Street, 15, 614990 Perm, Russia

³ Samara State Medical University, Samara, Russia, 89 Chapaevskaya St., Samara, Russia

* Correspondence: m.a.barulina@spbu.ru

Abstract

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system characterized by high heterogeneity of clinical progression and limited accuracy of traditional prognostic approaches. Predicting the transition to an active disease phase is critical for timely adjustment of disease-modifying therapy and reduction of disability risk. Three deep learning architectures were developed (ResNet3D, PretrainedResNet2D, and ResNeXt) for predicting MS activity within a one-year horizon using longitudinal multimodal data comprising T1-weighted MRI volumes (50 axial slices, 128×128 px) and Expanded Disability Status Scale (EDSS) functional subscores. The dataset included 28 patients (67 annual observations) with a confirmed MS diagnosis and a minimum two-year follow-up. Class imbalance (≈10% active-phase cases) was addressed through weighted cross-entropy loss (1:3.5) and Gaussian noise augmentation. The best performance was achieved by PretrainedResNet2D, which combines a 3D-to-2D MRI embedding with a pretrained ResNet18 backbone and a recurrent classifier, yielding F1 = 0.80. ResNeXt with self-normalizing layers and ReZero connections reached F1 = 0.73, while fully trainable ResNet3D achieved F1 = 0.62, constrained by the limited dataset size. Transfer learning effectively compensates for data scarcity in rare clinical settings. The proposed pipeline demonstrates feasibility as a decision-support tool in neurological practice and establishes a foundation for prospective multicenter validation.

Keywords: multiple sclerosis; disease activity prediction; deep learning; transfer learning; multimodal data; ResNet; recurrent neural network; clinical decision support

1. Introduction

Multiple sclerosis (MS) is a chronic autoimmune inflammatory disease of the central nervous system (CNS) that affects both the brain and spinal cord, leading to progressive demyelination and neurodegeneration. The condition exhibits substantial heterogeneity in its clinical manifestations, including visual impairment, sensory disturbances, motor dysfunction, pelvic organ disorders, and cognitive decline [1]. The growing prevalence of MS and its considerable impact on patients' quality of life, particularly among young and economically active individuals, underscore its clinical and social significance.

According to international epidemiological studies, the global prevalence of multiple sclerosis (MS) is approximately 35.9 cases per 100,000 individuals; however, these rates vary considerably depending on geographic region, ethnic background, and climatic conditions [2]. The highest prevalence has been reported in countries with temperate and northern climates, including Canada, Northern European nations, and Russia [3]. In Europe, the average prevalence of MS exceeds 100

cases per 100,000 population, while in various regions of Russia it ranges from 30 to 80 cases per 100,000 individuals [4].

The onset of multiple sclerosis (MS) typically occurs between the ages of 20 and 40 years, with women being affected approximately twice as often as men [5]. Recent evidence indicates a rising incidence of MS over the past decades, which may be partially attributed to improvements in diagnostic capabilities, as well as to environmental influences, lifestyle changes, and epigenetic mechanisms [6]. MS remains one of the leading causes of neurological disability among young adults, posing a substantial burden on both healthcare systems and society as a whole [7].

These characteristics make multiple sclerosis (MS) not only a medical but also a socio-economic challenge. In the absence of a radical cure, the focus shifts toward early detection, prognosis, and personalized treatment approaches, which necessitate the integration of advanced technologies, including those based on artificial intelligence methods.

Given the heterogeneity of MS clinical progression and the complexity of its pathogenesis, the task of predicting disease development remains of critical importance. The ability to forecast the transition to an active phase or disease progression enables timely adjustment of therapeutic strategies, reduction of disability risks, and improvement of patients' quality of life.

Currently, several approaches are employed to predict the course of multiple sclerosis (MS):

1. Clinical prognostic models based on initial disease manifestations, age-related characteristics, intervals between relapses, and the degree of recovery after each episode. Although these parameters are routinely considered in clinical practice, their predictive accuracy remains limited [8].
2. Disability scales, such as the Expanded Disability Status Scale (EDSS), are used to assess a patient's current functional state. While EDSS dynamics over time can serve as a reference for evaluating disease progression, the scale itself is not designed to predict future relapses or the rate of deterioration [9].
3. MRI markers, including the number and volume of new lesions, the presence of active contrast enhancement, and the degree of brain atrophy, offer greater sensitivity to pathomorphological changes and are widely applied for assessing disease activity. However, even with advanced MRI protocols, predictive value remains constrained due to substantial interindividual variability [10].
4. Blood and cerebrospinal fluid biomarkers, such as neurofilament light chain (NfL), glymphatic markers, and cytokines, demonstrate correlations with inflammatory activity and neurodegeneration. Nonetheless, their clinical implementation is still limited, primarily due to the lack of validated threshold values [11].

Machine learning and artificial intelligence (AI) methods enable the development of multimodal models that integrate clinical, MRI, and laboratory data. Such models have the potential to uncover nonlinear relationships and predict outcomes with higher accuracy, particularly when large datasets are available [12].

The implementation of prognostic models in clinical practice is essential for advancing personalized therapy. A predicted high risk of relapse may serve as grounds for switching to second line disease-modifying therapies (DMTs), such as natalizumab or alemtuzumab, whereas a favorable prognosis may justify continued observation or a less intensive treatment regimen [13].

Furthermore, prognostication plays a pivotal role in planning non-pharmacological interventions, including neurorehabilitation, cognitive support, and lifestyle adjustments. The earlier information about potential deterioration is obtained, the more effectively preventive strategies can be applied.

Among all available diagnostic modalities, magnetic resonance imaging (MRI) demonstrates the highest sensitivity to structural changes associated with multiple sclerosis. It provides non-invasive visualization of both old and new demyelinating lesions, contrast-enhancing, inflammation, axonal loss, and brain atrophy, thus covering both early and late pathophysiological stages of the disease [14].

In the context of predicting the course of multiple sclerosis (MS), MRI provides important quantitative and qualitative biomarkers, including the number and volume of T2-hyperintense lesions, the appearance of new T1-hypointense “black holes”, the presence of gadolinium-enhancing (Gd+) lesions, and the rate of progressive gray and white matter atrophy. These parameters show robust correlations with clinical disease activity, the rate of progression, and the degree of disability.

Moreover, in the monitoring of DMT efficacy, MRI remains the gold standard for assessing subclinical disease activity, particularly in cases where overt clinical relapses are absent. Many contemporary clinical trials and patient management protocols employ the concept of “no evidence of disease activity” (NEDA), which encompasses not only the absence of clinical relapses and disability progression but also the absence of new or enlarging MRI lesions.

Thus, MRI combines high diagnostic value, prognostic relevance, and versatility as a marker of treatment effectiveness, making it an indispensable component not only of diagnosis but also of longitudinal monitoring in patients with MS. At the same time, the application of artificial intelligence methods to MRI data opens new perspectives for identifying patterns that may escape visual assessment and for developing more accurate individualized prognostic models.

Given the high variability of MS progression and the limited accuracy of traditional prognostic approaches, there is a growing need to introduce novel methods based on the analysis of multimodal data, particularly MRI and clinical disability scales. Modern artificial intelligence techniques, especially deep learning, make it possible to process such data with a high degree of granularity and to detect patterns that are not accessible to conventional visual analysis.

The aim of this study was to develop and validate a deep learning-based model for predicting the probability of transition of multiple sclerosis to an active disease phase within a one-year horizon time horizon, using longitudinal MRI data and temporal dynamics of EDSS scores.

The main results of this study are:

A multimodal deep learning pipeline combining 3D MRI and EDSS subscales is proposed;

5. Three architectures (ResNet3D, PretrainedResNet2D, and ResNeXt) are compared on a rare clinical dataset;
6. Transfer learning is demonstrated to compensate for data limitations in predicting multiple sclerosis.

2. Materials and Methods

2.1. Dataset

The study included data from 28 patients with a clinically confirmed diagnosis of multiple sclerosis (according to the McDonald criteria), who had been under observation for a minimum of two years. For each patient, the following data were available: a series of MRI examinations (covering one to four years of follow-up) and Expanded Disability Status Scale (EDSS) scores. The target variable, reflecting the presence of an active disease phase in the subsequent year, was defined as disease activity. Accordingly, the study cohort was stratified by a binary label: activity and remission.

Each observation comprised 3D MRI images (T1-weighted sequence, DICOM format) and tabular data collected in the year of the corresponding examination.

The target variable (class label) for each examination was assigned based on the following screening procedure. The target variable took one of two values: 1 if the patient’s condition deteriorated at the subsequent examination, and 0 if the condition improved or remained stable.

The Expanded Disability Status Scale (EDSS) is a widely used instrument for assessing the severity of disease progression in patients with multiple sclerosis. Patient status is evaluated across seven functional system subscales, each scored either on a range from 0 to 5 or from 0 to 6, along with an eighth subscale scored from 0 to 12. The EDSS encompasses the following functional domains: visual function, brainstem function, pyramidal function, cerebellar function, sensory function, bladder and bowel function, cerebral (cognitive) function, and ambulation. Based on these eight subscales, a composite EDSS score is derived: 0 indicates a neurologically intact individual, while

scores ranging from 1.0 to 10.0 in increments of 0.5 reflect varying degrees of neurological impairment.

2.2. Metrics

The following metrics were employed for model evaluation.

Confusion matrix. The confusion matrix provides a structured summary of the model's predictive performance, characterizing the number of correct and incorrect predictions across all classes (Table 1). All subsequent evaluation metrics are derived from the confusion matrix.

Table 1. Confusion matrix.

| | Predicted: Positive (Activity) | Predicted: Negative (Remission) |
|-------------------------------------|---------------------------------------|--|
| Actual: Positive (Activity) | TP = True Positive | FN = False Negative |
| Actual: Negative (Remission) | FP = False Positive | TN = True Negative |

Accuracy. Accuracy is a standard metric for classification problems, computed as the ratio of the number of correct predictions to the total number of predictions. However, in the present study its interpretability is limited due to class imbalance: for instance, a trivial model that predicts only the negative class (constant zero) would still achieve an accuracy of approximately 0.9:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Precision. Precision is defined as the proportion of instances predicted as positive that are in fact positive:

$$Precision = \frac{TP}{TP + FP}$$

Recall. Recall is defined as the proportion of truly positive instances that are correctly identified by the classifier as positive:

$$Recall = \frac{TP}{TP + FN}$$

F1-score. The F1-score is the harmonic mean of precision and recall and thus provides a single summary metric that balances both components:

$$F_1 = 2 \frac{Recall * Precision}{Recall + Precision}$$

F_β-score. The F_β-score is a weighted harmonic mean of precision and recall. It quantifies classifier performance when recall is considered β times more important than precision. It reduces to the F1-score for β = 1.

$$F_\beta = \frac{(\beta^2 + 1)(Recall * Precision)}{Recall + \beta^2 Precision}$$

2.3. Data Preprocessing

The MRI data exhibited several characteristics that necessitated careful preprocessing:

Color encoding. Unlike the standard RGB color model, in which pixel intensity values are bounded within the range of 0 to 255, MRI data may contain substantially higher pixel intensity values, with the maximum varying across different scans (exceeding 3000 in some cases). Consequently, pixel intensity values were normalized to the range of 0 to 255 for each MRI scan individually.

T1- and T2-weighted images. T1- and T2-weighted images represent the two primary acquisition types in magnetic resonance imaging, differing in the mechanism of contrast generation due to

variations in the physical properties of tissues. As a result, different tissue types may appear with distinct signal characteristics (see Table 2). Each MRI examination contains multiple images of the same brain acquired at the same time point but under different pulse sequences: some correspond to T1-weighted variants, while others correspond to T2-weighted variants. Given that T1-weighted images are more numerous across the dataset and that T2-weighted sequences are not consistently available for all scans — whereas T1-weighted sequences are present in all cases — T1-weighted images were selected as the primary input modality. Accordingly, only the T1-weighted slices from each scan were passed to the model.

Table 2. Signal intensity characteristics of different tissue types on T1- and T2-weighted MRI sequences.





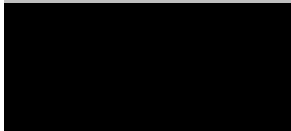
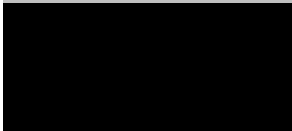




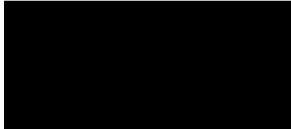

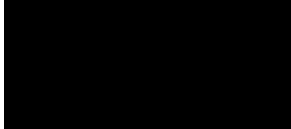
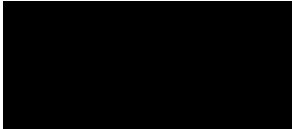


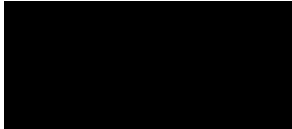
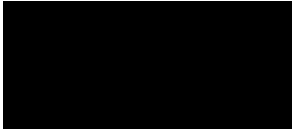
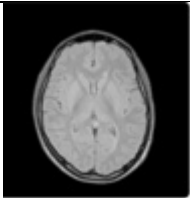
| | T1 | T2 |
|------------------------|---|---|
| Bone marrow: |  |  |
| Muscles: |  |  |
| Blood: |  |  |
| White matter: |  |  |
| Gray matter: |  |  |
| Fluids : |  |  |
| Bones (cortical bone): |  |  |
| Fat: |  |  |
| Air: |  |  |

Image dimensions. Each MRI examination can be conceptualized as a stack of sequentially arranged two-dimensional axial slices. Across the dataset, the number of slices per examination varied, with the total range spanning from 30 to 60 layers. This variability necessitated standardization to a uniform depth. A fixed depth of 50 axial slices was adopted, as this value minimized the number of examinations affected by information loss due to slice reduction. For MRI volumes containing fewer than 50 slices, zero-padded images were appended symmetrically at both the superior and inferior ends to reach the target depth. In addition, each individual slice was spatially resampled to a standardized in-plane resolution of 128×128 pixels.

As a result, the data for each observation of each patient consisted of a sequence of MRI examinations, EDSS scores, and the corresponding disease status (Table 3). Data preprocessing was performed for both MRI and EDSS data.

For the MRI data, 50 axial slices were extracted from each T1-weighted sequence per patient, uniformly distributed along the anatomical axis. When the number of available slices was insufficient, zero-filled (empty) slices were symmetrically added; when the number exceeded 50, slices were subsampled at regular intervals. All images were resampled to a spatial resolution of 128×128 pixels, and voxel intensities were normalized to the range 0–255. In this way, 3D tensors of size $[50 \times 128 \times 128]$ were constructed and then aggregated across follow-up years into 4D sequences.

Table 3. Data types included in each observation.

| Raw data description | Observation | | |
|----------------------|---|----------------------------|------------------------|
| | MRI | Status (Active/Not active) | EDSS |
| Data type | Sequence of MRI scans | Binary values (0 or 1) | Vector of 8 integers |
| Example value |  | 0 | 2, 1, 0, 2, 1, 0, 1, 0 |

EDSS data were preprocessed as follows. Instead of using the composite EDSS score, we employed its components, namely the seven functional system subscores, each of which reflects the severity of a specific functional domain (e.g., visual function, sensory deficits, bladder and bowel function). This approach helped to avoid excessive collinearity between features and to capture more differentiated patterns of disability.

The subscores were normalized and represented as a temporal sequence aligned with each year of follow-up.

The target variable for each observation was defined as the Status (Active/Not active) at the subsequent visit, as the model was designed to predict future disease activity. The resulting dataset was imbalanced: cases corresponding to an active disease phase accounted for approximately 10% of all observations. To mitigate this issue, we applied filtering of observations with an excessive number of negative labels, performed data augmentation by adding Gaussian noise to 20% of MRI volumes corresponding to active MS, and excluded incomplete or corrupted images. As a result, a total of 67 training samples (annual observations) were obtained, each comprising a 3D MRI volume, the EDSS subscore vector, and the activity label.

3. Model Architectures

A series of multimodal deep neural network architectures were trained and validated, from which three models with the best performance metrics were selected: ResNet3D, PretrainedResNet2D, and ResNeXt. Each of these models employs a distinct strategy for processing imaging data and tabular features. The primary criterion for model selection was their ability to predict the onset of an active phase of MS within a one-year horizon based on MRI examinations and EDSS data.

3.1. ResNet3D Architecture

The ResNet3D model is based on the standard Residual Network (ResNet) architecture introduced by He et al. in 2015 [15]. The central idea of the original architecture is the use of skip connections between layers, which facilitate the optimization of deep networks and improve training stability.

The modification of ResNet3D proposed in this study is tailored to the following assumptions. First, the model must operate directly on three-dimensional image data. Second, it should apply convolutions along the slice axis without aggressively reducing its resolution, thereby preserving the spatial context across slices.

The proposed 3D-ResNet architecture is depicted in Figure 2. Each MRI examination is first fed into a 3D convolutional layer with a kernel size of $7 \times 7 \times 7$, followed by normalization, a ReLU activation function, and a max-pooling operation. The resulting feature maps are then passed through a stack of 3D residual blocks (ResidualBlock3D), followed by average pooling. The pooled features are flattened into a one-dimensional vector, concatenated with the EDSS subscores, and subsequently passed to a recurrent neural network, which produces the final prediction outputs.

The architecture is composed of ResidualBlock3D units with explicit skip connections (Figure 3). When the convolution stride is equal to 1 and the number of channels remains unchanged, the data are processed using the second configuration (Way 2); otherwise, the first configuration (Way 1) is applied. In Way 1, the main branch first passes through a 3D convolution with a kernel size of $3 \times 3 \times 3$ and stride $(1, s, s)$, which maps m input channels to n output channels (where s , m , and n are block hyperparameters), followed by normalization, a ReLU activation, and a second 3D convolution with a $3 \times 3 \times 3$ kernel, unit stride $(1, 1, 1)$, and unchanged channel dimensionality, again followed by normalization. In the skip branch, the input undergoes a $1 \times 1 \times 1$ convolution with stride $(1, s, s)$ to project from m to n channels, followed by normalization. The outputs of the main and skip branches are then added element-wise and passed through a final ReLU activation. In Way 2, the skip branch remains an identity mapping, and the input is forwarded unchanged along the residual connection.

After the 3D-ResNet component, which processes only the MRI volumes, the resulting feature representation is concatenated with the EDSS scores and fed into a recurrent neural network (Figure 2). It should be noted that, in all models considered in this work, a standard recurrent neural network (RNN) architecture is employed. This choice is motivated by the fact that more complex recurrent architectures, such as LSTMs or GRUs, are primarily advantageous for modeling long sequences, whereas in the present study all temporal sequences are limited to a maximum length of five time points.

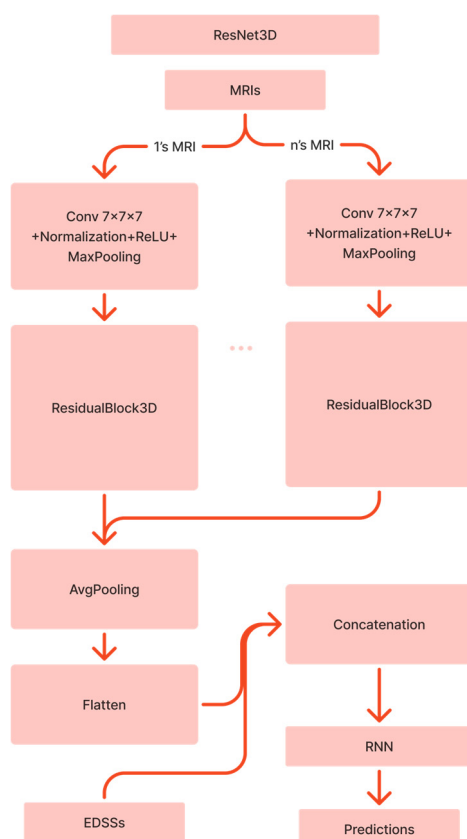


Figure 2. ResNet3D architecture: three-dimensional residual neural network for joint processing of MRI volumes and EDSS features.

3.2. PretrainedResNet2D Architecture

The second model in this study leverages transfer learning techniques. To enable the use of pretrained two-dimensional architectures, we employ an embedding procedure that maps three-dimensional MRI volumes into a two-dimensional representation. In this approach, a small convolutional network (Figure 4) first reshapes the 3D data into a 2D layout and then applies a 2D convolution with a kernel size of 1×1 and stride (1, 1), transforming 60 input channels into 3 output channels.

The core of the model is a pretrained ResNet18 network, chosen as a backbone due to its favorable trade-off between efficiency and model complexity. The model processes each MRI examination through the embedding module and then feeds the resulting 2D representation into ResNet18. The extracted features are subsequently concatenated with the EDSS scores and passed to a recurrent neural network, which produces the final predictions (Figure 5).

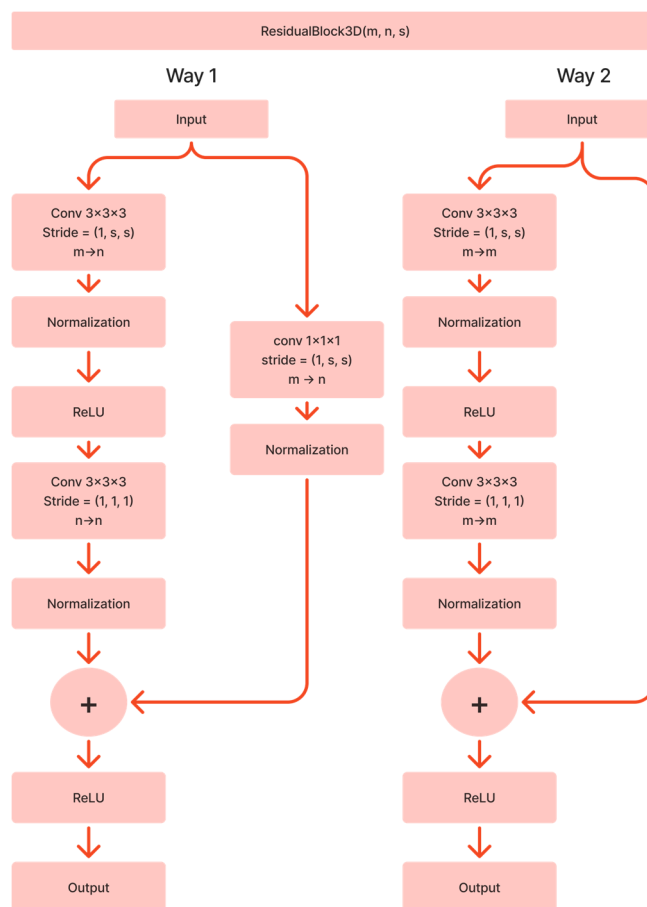


Figure 3. Architecture of the ResidualBlock3D unit with two configurations of the skip connection (Way 1 with projection shortcut and Way 2 with identity shortcut).

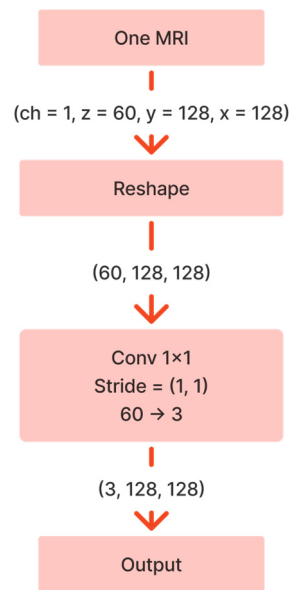


Figure 4. Embedding pipeline used in the PretrainedResNet2D model: transformation of 3D MRI volumes into 2D feature maps suitable for processing by the pretrained ResNet18 backbone.

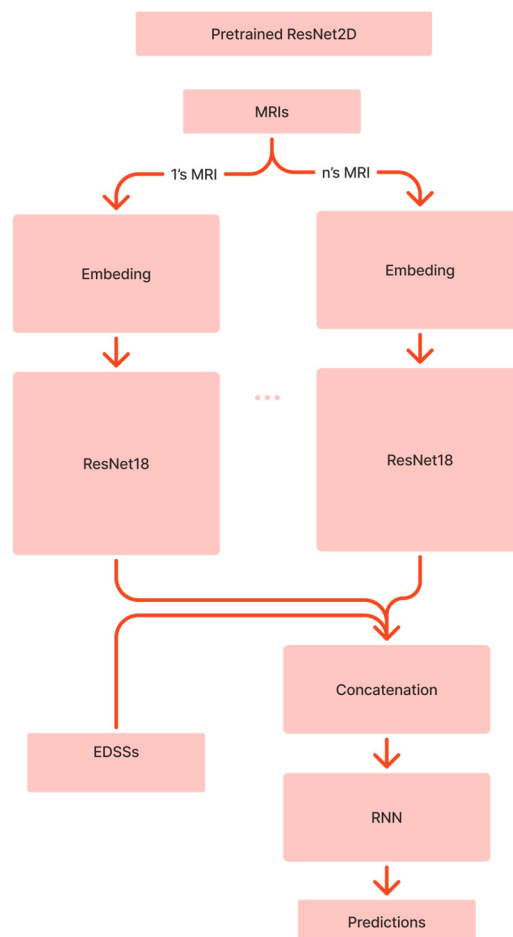


Figure 5. PretrainedResNet2D architecture: multimodal model combining 3D-to-2D MRI embeddings processed by a pretrained ResNet18 backbone with EDSS features, followed by a recurrent neural network classifier. .

3.3. ResNeXt Architecture

To address the prediction task, we employed a ResNeXt-based architecture [17], illustrated in Figure 6. In this architecture, the input to the convolutional block, consisting of C_{in} channels, is processed by b parallel convolutional branches with 1×1 kernels, each mapping the input to an intermediate representation with C_{mid} channels. Each of these intermediate feature maps is then passed through an additional convolution that preserves the channel dimensionality. The resulting b feature maps are concatenated along the channel dimension to form a single tensor with $b \cdot C_{mid}$ channels, which is subsequently transformed back to the original channel dimensionality by a 1×1 convolution and added to the block input via a residual connection.

The standard ResNeXt implementation was further modified by incorporating ReZero-based skip connections [18], applying an attention mechanism to the output of the convolutional component [19], and implementing the model in accordance with SNN design principles [20]:

- The inputs to the model are normalized to have zero mean and unit variance.
- The scaled exponential linear unit (SELU) activation function is employed.
- Batch normalization is not used.
- Network weights are initialized from a normal distribution with zero mean and variance equal to $1/N$, where N denotes the number of incoming connections.
- Instead of standard dropout, alpha-dropout is applied; rather than setting units to zero at random, it randomly sets them to the limiting value of $\text{SELU}(x)$ as $x \rightarrow -\infty$ which equals $-\lambda\alpha$.

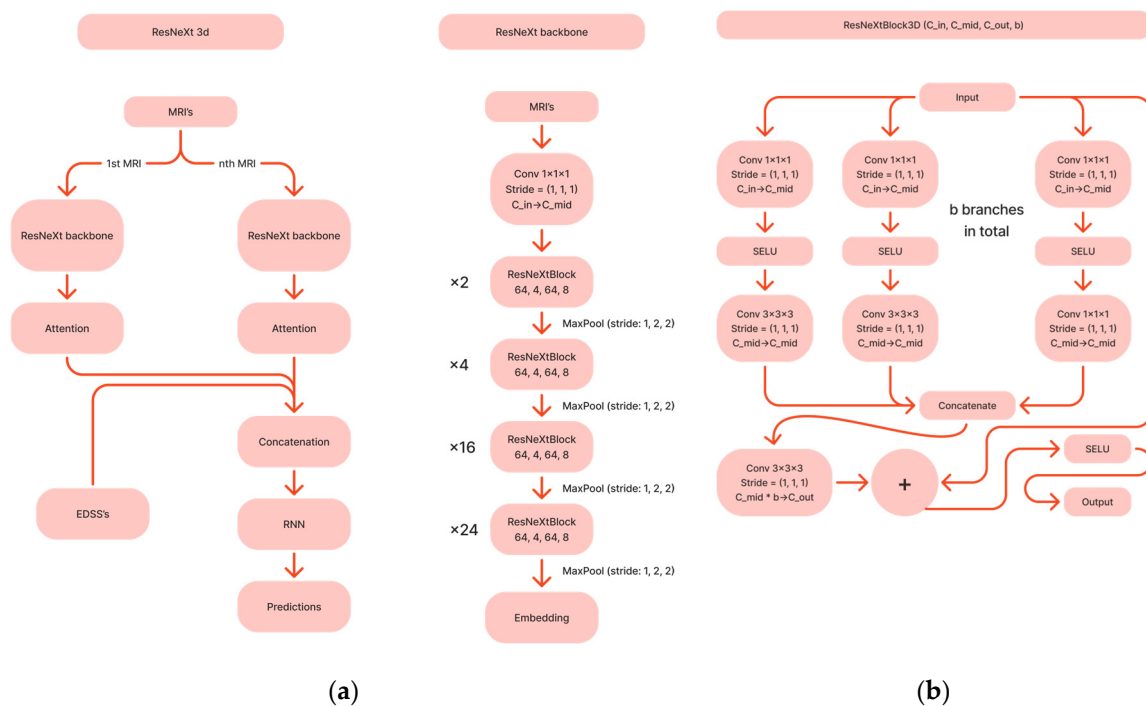


Figure 6. (a) ResNeXt3D architecture for multimodal processing of MRI volumes and EDSS features; (b) structure of the ResNeXtBlock3D units implementing aggregated residual transformations.

4. Model Training and Evaluation Metrics

All three proposed models were trained using the following computational resources: an Intel Core i9 CPU and an NVIDIA GeForce RTX 4060 Ti GPU with 8 GB of memory. The target value for the F1-score metric was set to 0.7; once this threshold was exceeded, the model was considered to perform sufficiently well in terms of prediction quality.

For training each of the models, the CrossEntropyLoss function was employed in combination with the Softmax activation function. The latter is used to transform the vector of raw prediction logits into a probability distribution over the possible classes, whereas the former quantifies the error of these probabilistic predictions.

The definition of the Softmax activation function is:

$$\text{Softmax}(y)_i = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

The CrossEntropyLoss function:

$$\text{Loss} = - \sum_i y_i * \log \hat{y}_i$$

The 3D-ResNet model was trained for 120 epochs using the Adam optimizer with a learning rate of $1 \cdot 10^{-4}$. The loss function was defined as a weighted CrossEntropyLoss with class weights of 1:3.5 to compensate for the dataset imbalance. This choice of class weights was empirically determined to yield the most effective model performance during experimentation.

The model's precision exceeded 0.70; however, the F1-score remained below the predefined target threshold of 0.7 (Table 4). Error analysis indicated a tendency toward overfitting, particularly in the presence of non-informative or augmented sequences. This behavior can be attributed to the limited size of the training set and the model's sensitivity to class balance.

Table 4. Evaluation results for the ResNet3D model.

| Metrics | Values |
|-----------------------|---|
| Accuracy | 0.714 |
| Precision | 0.600 |
| Recall | 0.600 |
| F1-score | 0.620 |
| Fb-score (beta = 0.5) | 0.600 |
| Confusion matrix | TP: 0.21 FP: 0.14 FN: 0.14 TN: 0.5 |

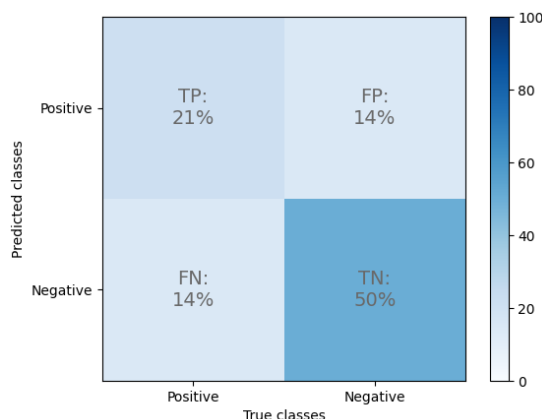


Figure 7. Confusion matrix for the ResNet3D model on the test set.

The PretrainedResNet2D model was trained for 120 epochs using the same hyperparameters as the 3D-ResNet model and demonstrated stable convergence and strong generalization performance. The achieved F1-score exceeded the target threshold by 0.1 (Table 5), and the balance between precision and recall indicates high model reliability. This approach effectively leverages the advantages of transfer learning, reducing the need for large training datasets in this setting.

Table 5. Evaluation results for the PretrainedResNet2D model.

| Metrics | Values |
|-----------------------|----------------------|
| Accuracy | 0.857 |
| Precision | 0.800 |
| Recall | 0.800 |
| F1-score | 0.800 |
| Fb-score (beta = 0.5) | 0.800 |
| Confusion matrix | TP: 0.29 FP: 0.07 |

FN: 0.07 TN: 0.57

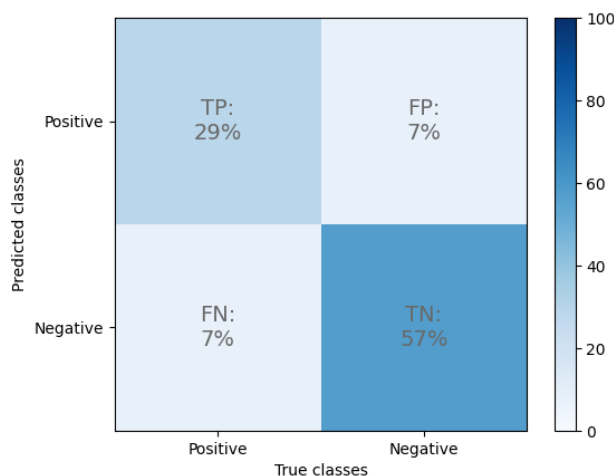


Figure 8. Confusion matrix for the PretrainedResNet2D model on the test set.

The ResNeXt model was trained for 23 epochs using the Nadam optimizer, which extends the standard Adam algorithm by incorporating Nesterov momentum, with the following hyperparameters: a learning rate of $1 \cdot 10^{-3}$ and a weight decay of 0.05. As in the previous experiments, the loss function was defined as a weighted CrossEntropyLoss with class weights of 1:3.5.

The ResNeXt model also achieved an F1-score exceeding the predefined target value (Table 6). Although its performance metrics were slightly inferior to those of the PretrainedResNet2D model, it should be noted that no transfer learning was applied in this architecture, indicating that its potential for further performance improvement has not yet been fully exploited.

Table 6. Evaluation results for the ResNeXt.

| Metrics | Values |
|-----------------------|---|
| Accuracy | 0.786 |
| Precision | 0.667 |
| Recall | 0.800 |
| F1-score | 0.727 |
| Fb-score (beta = 0.5) | 0.690 |
| Confusion matrix | TP: 0.29 FP: 0.14 FN: 0.07 TN: 0.5 |

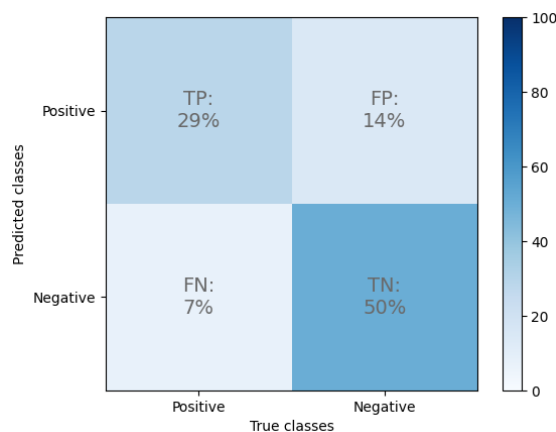


Figure 9. Confusion matrix for the ResNeXt model on the test set.

5. Discussion

The results of this study confirm the promise of deep learning methods for predicting multiple sclerosis (MS) activity over a one-year horizon using multimodal data. Within the proposed framework, we evaluated three neural network architectures, including 3D convolutional models and architectures combining embeddings with pretrained backbones. The performance metrics of these models are summarized in Table 7.

The best performance was achieved by the PretrainedResNet2D model, which is based on a pretrained 2D ResNet receiving 2D embeddings derived from 3D MRI data. This finding supports the hypothesis that transfer learning and pretrained convolutional layers can effectively compensate for limited training data in medical imaging tasks. The obtained F1-score of 0.80 exceeds the predefined target threshold and indicates both high sensitivity and high precision of the model.

Table 7. Comparative performance metrics of the proposed models.

| Architecture | Accuracy | Precision | Recall | F1-score | Fb-score |
|--------------------|----------|-----------|--------|----------|----------|
| 3D-ResNet | 0.714 | 0.600 | 0.600 | 0.600 | 0.600 |
| PretrainedResnet2D | 0.857 | 0.800 | 0.800 | 0.800 | 0.800 |
| ResNeXt | 0.786 | 0.667 | 0.800 | 0.727 | 0.690 |

The 3D-ResNet model demonstrated acceptable, though less compelling, performance. This is likely due to the fact that a fully trainable 3D network typically requires a substantially larger amount of data than was available in the present study. In turn, the ResNetX -based model with self-normalizing layers did not achieve adequate sensitivity, which makes it of limited utility in a clinical context where failing to detect an active disease phase is particularly undesirable.

It is important to note that this study has several limitations.

Small sample size. The training cohort comprised 28 patients, which constrains the generalizability of the models.

Restricted data modalities. Only T1-weighted MRI sequences were used, without additional modalities such as FLAIR, T2-weighted, or contrast-enhanced images.

Target definition. The binary annual activity label may not fully capture the complexity and heterogeneity of the clinical course of MS.

Lack of external validation. The models were not evaluated on independent external cohorts, which limits the robustness and real-world applicability of the findings.

6. Conclusion

This pilot study demonstrates the feasibility of multimodal deep learning for predicting MS disease activity. Among the three evaluated architectures, PretrainedResNet2D achieved the highest performance (F1 = 0.80), outperforming fully trainable ResNet3D (F1 = 0.62), confirming that transfer learning effectively compensates for limited data availability in rare clinical settings.

The development of algorithms for predicting MS activity has important practical implications. A model capable of estimating the probability of transition to an active disease phase in advance can:

- Enable timely revision of therapy, including escalation to more aggressive DMTs.
- Improve the efficiency of scheduling MRI follow-up and other diagnostic assessments.
- Serve as a tool for patient stratification in clinical trials.
- Function as a component of clinical decision support systems in neurological practice.

Thus, the proposed approach may be valuable not only for personalized medicine, but also as an integral element of a digital clinic infrastructure. The study demonstrates that neural network models trained on MRI data and EDSS scores can be successfully used to predict multiple sclerosis activity with high accuracy. The proposed methodology shows strong potential for clinical

integration and provides a foundation for the development of digital assistants to support the management of patients with MS.

Building on the findings of this study, future work should prioritize expanding the patient cohort and validating the proposed models on independent external datasets to enhance robustness and generalizability. Another key direction is the integration of additional MRI modalities (such as FLAIR, T2-weighted, and contrast-enhanced sequences) and complementary biomarkers, which may allow a more comprehensive characterization of MS pathophysiology and disease activity. From a methodological perspective, further improvements could be achieved by investigating advanced multimodal fusion strategies and uncertainty-aware prediction frameworks to increase the reliability of model outputs in clinically critical scenarios. Finally, prospective clinical studies are required to assess the real-world utility of the proposed models when embedded into clinical decision support systems and digital patient pathways for MS management.

Author Contributions: Conceptualization, M.B. and A.Z.; methodology, M.B., O.S. and I.C.; software, O.S. and I.C.; validation, M.B., O.S. and I.C.; formal analysis, O.S. and I.C.; investigation, A.Z. and E.K.; resources, A.Z. and E.K.; data curation, A.Z. and E.K.; writing — original draft preparation, M.B., O.S. and I.C.; writing — review and editing, M.B. and A.Z.; visualization, O.S. and I.C.; supervision, M.B.; project administration, M.B.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Economic Development of the Russian Federation (Agreement No. 139-15-2025-007, dated April 16, 2025; ID: 000000C313925P3O0002).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Samara State Medical University (protocol code 227, approval date May 26, 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are not publicly available due to privacy and ethical restrictions governing patient clinical data. Anonymized data may be made available upon reasonable request from the corresponding author, subject to institutional approval.

Acknowledgments: Not applicable.

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

References

1. Dobson, R.; Giovannoni, G. Multiple sclerosis – a review. *Eur. J. Neurol.* **2019**, *26*, 27–40. <https://doi.org/10.1111/ene.13819>
2. Walton, C.; King, R.; Rechtman, L.; Kaye, W.; Leray, E.; Marrie, R.A.; Robertson, N.; La Rocca, N.; Uitdehaag, B.; Van Der Mei, I.; et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult. Scler. J.* **2020**, *26*, 1816–1821. <https://doi.org/10.1177/1352458520970841>
3. Kingwell, E.; Marriott, J.J.; Jetté, N.; Pringsheim, T.; Makhani, N.; Morrow, S.A.; Fisk, J.D.; Evans, C.; Béland, S.G.; Kulaga, S.; et al. Incidence and prevalence of multiple sclerosis in Europe: A systematic review. *BMC Neurol.* **2013**, *13*, 128. <https://doi.org/10.1186/1471-2377-13-128>
4. Gusev, E.I.; Zavalishin, I.A.; Boiko, A.N.; Khoroshilova, N.L.; Iakovlev, A.P. Epidemiological characteristics of multiple sclerosis in Russia. *Zh. Nevrol. Psikhiatr. Im. S.S. Korsakova* **2002**, Suppl., 3–6.
5. Orton, S.-M.; Herrera, B.M.; Yee, I.M.; Valdar, W.; Ramagopalan, S.V.; Sadovnick, A.D.; Ebers, G.C. Sex ratio of multiple sclerosis in Canada: A longitudinal study. *Lancet Neurol.* **2006**, *5*, 932–936. [https://doi.org/10.1016/S1474-4422\(06\)70581-6](https://doi.org/10.1016/S1474-4422(06)70581-6)
6. Ascherio, A. Environmental factors in multiple sclerosis. *Expert Rev. Neurother.* **2013**, *13* (Suppl. 2), 3–9. <https://doi.org/10.1586/14737175.2013.865866>

7. Ernstsson, O.; Gyllensten, H.; Alexanderson, K.; Tinghög, P.; Friberg, E.; Norlund, A. Cost of illness of multiple sclerosis — A systematic review. *PLoS ONE* **2016**, *11*, e0159129. <https://doi.org/10.1371/journal.pone.0159129>
8. Kalincik, T.; Havrdova, E.; Horakova, D.; Izquierdo, G.; Prat, A.; Girard, M.; Duquette, P.; Grammond, P.; Lugaresi, A.; Grand'Maison, F.; et al. Comparison of fingolimod, dimethyl fumarate and teriflunomide for multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* **2019**, *90*, 458–468. <https://doi.org/10.1136/jnnp-2018-319831>
9. Meyer-Moock, S.; Feng, Y.-S.; Maeurer, M.; Dippel, F.-W.; Kohlmann, T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurol.* **2014**, *14*, 58. <https://doi.org/10.1186/1471-2377-14-58>
10. Sormani, M.P.; Bruzzi, P. MRI lesions as a surrogate for relapses in multiple sclerosis: A meta-analysis of randomised trials. *Lancet Neurol.* **2013**, *12*, 669–676. [https://doi.org/10.1016/S1474-4422\(13\)70103-0](https://doi.org/10.1016/S1474-4422(13)70103-0)
11. Disanto, G.; Barro, C.; Benkert, P.; Naegelin, Y.; Schädelin, S.; Giardiello, A.; Zecca, C.; Blennow, K.; Zetterberg, H.; Leppert, D.; et al. Serum neurofilament light: A biomarker of neuronal damage in multiple sclerosis. *Ann. Neurol.* **2017**, *81*, 857–870. <https://doi.org/10.1002/ana.24954>
12. Eshaghi, A.; Young, A.L.; Wijeratne, P.A.; Prados, F.; Arnold, D.L.; Narayanan, S.; Guttmann, C.R.G.; Barkhof, F.; Alexander, D.C.; Thompson, A.J.; et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat. Commun.* **2021**, *12*, 2078. <https://doi.org/10.1038/s41467-021-22265-2>
13. Montalban, X.; Gold, R.; Thompson, A.J.; Otero-Romero, S.; Amato, M.P.; Chandraratna, D.; Clanet, M.; Comi, G.; Derfuss, T.; Fazekas, F.; et al.ECTRIMS/EAN guideline on the pharmacological treatment of people with multiple sclerosis. *Eur. J. Neurol.* **2018**, *25*, 215–237. <https://doi.org/10.1111/ene.13536>
14. Bakshi, R.; Thompson, A.J.; Rocca, M.A.; Pelletier, D.; Dousset, V.; Barkhof, F.; Inglese, M.; Guttmann, C.R.; Horsfield, M.A.; Filippi, M. MRI in multiple sclerosis: Current status and future prospects. *Lancet Neurol.* **2008**, *7*, 615–625. [https://doi.org/10.1016/S1474-4422\(08\)70137-6](https://doi.org/10.1016/S1474-4422(08)70137-6)
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
16. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2021**, *109*, 43–76. <https://doi.org/10.1109/JPROC.2020.3004555>
17. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500. <https://doi.org/10.1109/CVPR.2017.634>
18. Bachlechner, T.; Majumder, B.P.; Mao, H.; Cottrell, G.; McAuley, J. ReZero is all you need: Fast convergence at large depth. In *Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021)*; PMLR: Cambridge, MA, USA, 2021; Volume 161, pp. 1352–1361.
19. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>
20. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-normalizing neural networks. In *Advances in Neural Information Processing Systems (NIPS 2017)*; Curran Associates: Red Hook, NY, USA, 2017; Volume 30, pp. 972–981.

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.