

Article

Classification of Muscular Dystrophies from MR Images Improves Using the Swin Transformer Deep Learning Model

Alfonso Mastropietro ¹, Nicola Casali ^{1,2}, Maria Giovanna Taccogna ³, Maria Grazia D'Angelo ⁴,
Giovanna Rizzo ^{1,*} and Denis Peruzzo ⁵

- ¹ Istituto di Sistemi e Tecnologie Industriali Intelligenti per il Manifatturiero Avanzato, Consiglio Nazionale delle Ricerche, 20133 Milan, Italy; alfonso.mastropietro@cnr.it (A.M.); nicola.casali@polimi.it (N.C.)
² Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy
³ Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, 20054 Segrate, Milan, Italy; mariagiovanna.taccogna@itb.cnr.it
⁴ Unit of Rehabilitation of Rare Diseases of the Central and Peripheral Nervous System, Scientific Institute IRCCS Eugenio Medea, 23842 Bosisio Parini, Lecco, Italy; grazia.dangelo@lanostrafamiglia.it
⁵ Neuroimaging Unit, Scientific Institute IRCCS Eugenio Medea, 23842 Bosisio Parini, Lecco, Italy
* Correspondence: giovanna.rizzo@cnr.it

Abstract: Muscular dystrophies present diagnostic challenges, requiring accurate classification for effective diagnosis and treatment. This study investigates the efficacy of deep learning methodologies in classifying these disorders using skeletal muscle MRI scans. Specifically, we assess the performance of the Swin Transformer (SwinT) architecture against traditional convolutional neural networks (CNNs) in distinguishing between healthy individuals, Becker muscular dystrophy (BMD), and limb–girdle muscular Dystrophy type 2 (LGMD2) patients. Moreover, 3T MRI scans from a retrospective dataset of 75 scans (from 54 subjects) were utilized, with multiparametric protocols capturing various MRI contrasts, including T1-weighted and Dixon sequences. The dataset included 17 scans from healthy volunteers, 27 from BMD patients, and 31 from LGMD2 patients. SwinT and CNNs were trained and validated using a subset of the dataset, with the performance evaluated based on accuracy and F-score. Results indicate the superior accuracy of SwinT (0.96), particularly when employing fat fraction (FF) images as input; it served as a valuable parameter for enhancing classification accuracy. Despite limitations, including a modest cohort size, this study provides valuable insights into the application of AI-driven approaches for precise neuromuscular disorder classification, with potential implications for improving patient care.

Keywords: deep learning; classification; MRI; Swin Transformer; neuromuscular diseases; skeletal muscle



Citation: Mastropietro, A.; Casali, N.; Taccogna, M.G.; D'Angelo, M.G.; Rizzo, G.; Peruzzo, D. Classification of Muscular Dystrophies from MR Images Improves Using the Swin Transformer Deep Learning Model. *Bioengineering* **2024**, *11*, 580. <https://doi.org/10.3390/bioengineering11060580>

Academic Editors: Chunfeng Zhao and Ravinder Reddy Regatte

Received: 22 April 2024
Revised: 28 May 2024
Accepted: 5 June 2024
Published: 7 June 2024



Copyright: © 2024 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

Muscular dystrophies (MDs) are widespread muscular disorders that pose relevant clinical and diagnostic challenges. They are all characterized by progressive muscle tissue atrophy, with resulting muscle weakness significantly contributing to cardiopulmonary failure and heightened mortality risks [1,2]. Among the diverse MD subclasses, including limb–girdle muscular dystrophies (LGMDs) and dystrophinopathies such as Becker muscular dystrophy (BMD), some of them exhibit overlapping muscle involvement patterns [3], as observed in BMD and LGMD type 2 (LGMD2) [4,5].

Currently, the prevailing approaches for MD diagnosis and tissue characterization are muscle biopsy and genetic testing. Genetic assessment is the gold-standard method for a final diagnosis and it is usually based on genetic panels or next-generation sequencing (NGS) [6–8]. However, NGS, renowned for its capability to identify more than 90% of dystrophinopathy cases [9], is hindered by its high costs and complex operational demands, limiting its widespread application. Furthermore, NGS's accuracy diminishes when the candidate gene fails to align with the phenotype, when disease-causing variants are

elucidated by more than one gene, or when identified variants have unknown effects [10,11]. As for muscle biopsy, it can provide a quantitative characterization of the disease's impact on the muscular tissues; however, it is an invasive procedure and offers only a local assessment of the sampling sites, making its accuracy dependent on the amount of collected samples [12].

In this context, magnetic resonance imaging (MRI) can have a promising role, being a non-invasive multiparametric tool for investigating skeletal muscle morphology and structure [13]. Widely integrated into clinical practice, MRI plays a pivotal role in shaping diagnostic hypotheses by unveiling distinct muscle involvement patterns, which is crucial for informed clinical decision-making [14]. Its versatility extends to the assessment of fatty infiltration and abnormal increases in muscle water content, making it indispensable for diagnosing myopathies. Multiparametric muscle MRI proves invaluable beyond mere diagnosis. It not only aids in pinpointing muscle groups with inflammatory changes, facilitating cost-effective biopsies, but also extends its utility to provide crucial insights into disease progression and treatment responses within clinical trials and practical medical settings [3,15–17]. This multifaceted approach becomes particularly advantageous when clinical presentations suggest the involvement of multiple myopathy groups when a diagnosis is evident but confirmatory tests remain inconclusive, or when genetic analyses reveal variants of uncertain significance [14]. Finally, MRI serves as a confirmatory tool for clinical evaluations and directs patients toward more complex diagnostic investigations, enhancing the overall diagnostic process.

Clinical reports based on MRI images rely heavily on the skills of individual experts, mainly radiologists and neurologists which make them partly subjective and potentially inaccurate. Recently, artificial intelligence (AI) methods have emerged as promising tools to support MR image analysis and improve the objectivity of the diagnosis. In particular, deep learning (DL) approaches can extract relevant features from medical images to classify specific patterns, enabling precise and unbiased diagnostic evaluations [18,19]. DL methods, and particularly convolutional neural networks (CNNs), have demonstrated effectiveness in clinical applications for various medical tasks [19]. Vision Transformer (ViT), a recent innovation in this field, can overcome traditional CNN performances [20]. ViT excels in capturing long-range dependencies through self-attention mechanisms, facilitating the analysis of relationships among distant elements in an image. Notably, Swin Transformer (SwinT) offers scalability, allowing it to process images of different sizes and adapt to diverse datasets and resolutions while maintaining competitive performance, reducing computational demands [21]. Impressively, SwinT approaches exhibit robust generalization capabilities across a spectrum of computer vision tasks with minimal task-specific adjustments, and their interpretability and mitigation of overfitting contribute to their appeal [22,23].

AI research for classifying neuromuscular disorders using MR images shows promise [24]. Previous studies used machine learning (ML) [25], DL [26,27], or both [28,29]. Some investigations distinguished MD from non-MD cases [26], others differentiated DMD from BMD [28,29], and one study classified DMD, congenital MD, and healthy controls [27]. To date, the most comprehensive model, based on the Mercuri score and data from 10 MDs, lacks automated MRI feature extraction [25].

Considering what has been described above, it is evident that the existing literature lacks research papers dedicated to the automatic detection and differential diagnosis among BMD, LGMD2, and healthy controls using a multi-class classification approach. This gap is especially significant due to the overlapping muscle involvement patterns observed in both BMD and LGMD2 cases. Additionally, there remains a lack of literature focusing on the application of SwinT in this specific domain.

To address the above-identified research gaps, this study aims to develop and test a novel three-class classification approach utilizing the Swin Transformer model to identify healthy individuals and patients affected by BMD and LGMD2. Furthermore, considering the intrinsic multiparametric nature of MRI protocols, our study aims to evaluate the best

MRI contrast for classification performance. By undertaking this research, we anticipate contributing valuable insights that can enhance the accuracy and efficiency of muscular dystrophy diagnoses, ultimately improving patient care and outcomes.

2. Materials and Methods

2.1. Subjects

A total of 75 MRI scans (from 54 subjects) were included in this retrospective study. Among them, 36 scans come from 15 subjects who were acquired multiple times (up to 3 times) in a longitudinal study. More specifically, the whole dataset included 17 scans from healthy volunteers, 27 from BMD patients (17 subjects), and 31 from LGMD2 patients (10 LGMD2A and 10 LGMD2B patients). Table 1 lists the main demographic and clinical variables describing the subjects.

The diagnosis for all patients included in this study was determined through an evaluation that included clinical phenotyping, genetic testing, and muscle biopsies (specifically for those with LGMD2). Gene sequencing revealed a mutation in the dystrophin gene in patients with Becker muscular dystrophy. In cases of LGMD2A, two mutations were identified in the CAPN3 (calpain3) gene, along with a reduction in muscular calpain 3 as indicated by western blot analysis. For patients diagnosed with LGMD2B, two mutations were found in the DYSF (dysferlin) gene.

Patients with beta-sarcoglycan-related LGMD (n = 3), NDD LGMD (n = 1), ANO5-related LGMD (n = 1), gamma-sarcoglycan-related LGMD (n = 1), caveolin-3 LGMD (n = 1), Fukutin-related-protein LGMD (n = 2), and congenital myopathy (n = 2) were excluded due to the low number of occurrences. Furthermore, nine patients diagnosed in the early stages were also excluded, as they did not exhibit any clinical signs detectable in MRI scans.

This study adhered to ethical standards, with written informed consent being obtained from all participating individuals, thereby affirming their voluntary agreement to take part in the research. Furthermore, the local ethical committee (IRCCS Medea) provided its formal approval, endorsing the project's adherence to ethical guidelines and ensuring the protection of participants' rights and welfare.

Table 1. List of demographic characteristics and clinical variables of the enrolled volunteers. MFM = Motor function measure; n.a. = not applicable.

	Total	Healthy Subjects	BMD Patients	LGMD2 Patients
N	75	17	27	31
Age (mean ± SD)	42.4 ± 11.8 yo	39.1 ± 11.5 yo	39.3 ± 10.3 yo	47.0 ± 11.9 yo
Gender (M/F)	50/25	10/7	27/0	13/18
Disease duration	25.7 ± 9.4 y (pat. only)	n.a.	23.3 ± 8.3 yo	28.0 ± 10.1 yo
D1 (MFM) (mean ± SD) (pat. only)	17.3 ± 18.6 (pat. only)	n.a.	22.7 ± 18.6	12.9 ± 17.9
Ambulation (Y/N) (pat. only)	29/29 (pat. only)	n.a.	19/8	10/21

2.2. MRI Protocol

The MRI protocol included two sequences, a T1W gradient echo sequence and a DIXON sequence. The T1W sequence provides 2D structural images that can be used to identify and characterize the size of the different muscle bundles. DIXON sequences [30] are used to characterize the fat infiltration in the muscle tissue and provide three difference contrast maps: water (W) based contrast, fat (F) based contrast, and the fat fraction (FF) map.

MR images were acquired using a Philips Achieva dStream 3T MR scanner (Best, The Netherlands) and the sequence setup is presented in Table 2. All sequences were acquired on one of the subject thighs (dx or sx depending on the clinical/research query) and the field of view (FOV) upper limit was set to the middle of the femoral head.

Table 2. Acquisition parameters for the MRI scans.

	T1W	DIXON
N° of slices	30/36/50 *	30/36/50 *
Slice thickness	6 mm	6 mm
In-plane voxel size	1 × 1 mm	1.6 × 1.6 mm
Acquired matrix	256 × 256	160 × 160
N° of echoes	1	12
Echo Time	2.025 ms	1.48–14.68 ms
Echo spacing	-	1.2 ms
Repetition Time	647 ms	16.11 ms
Flip Angle	90 deg	3 deg
Percentage sampling	100%	78%
Averages	2	2

* Two healthy subjects with 50 slices and one BMD patient with 36 images.

2.3. Dataset

The dataset included a comprehensive collection of 75 MRI scans, each containing 30 slices, except for two scans of healthy control subjects that included 50 slices each, and one BMD scan consisting of 36 images. A subject-centric selection process was carried out to optimize the model training. For the classification experiments, a subset of 46 subjects, accounting for 67 scans and 2036 images, was selected for model training (60/61 scans) and validation (7/6 scans). To evaluate the effect of patients' selection for model training on classification performances, 10 independent experiments were conducted, each time randomly shuffling the 67 scans ensuring that each of them was at least once included in the validation set. The test set included the remaining 8 subjects (8 scans; 260 images), which were randomly selected from the non-longitudinal dataset. An example of MR images included in the dataset is displayed in Figure 1.

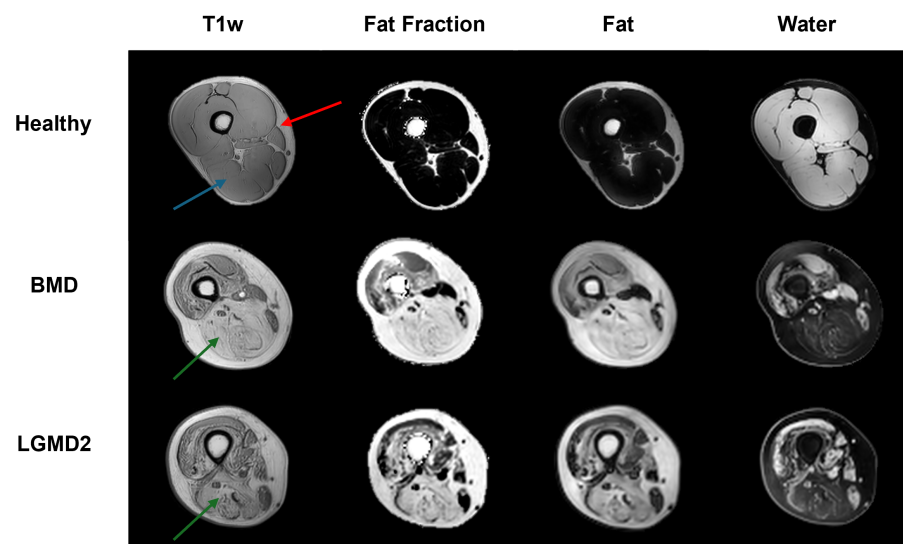


Figure 1. The image exemplifies the three distinct classes (healthy, BMD, and LGMD2), each delineated by unique image contrasts (T1w, FF, F, and W). The predicted subjects, chosen at random, show a typical cross-section of the thigh. The red arrow highlights the area of subcutaneous fat; the blue arrow (healthy subject) points to an example of intact skeletal muscle tissue, while the green arrows (BMD and GMD2 patients) show examples of heavily fat infiltrated muscles, a hallmark of neuromuscular disorders.

2.4. Implementation of DL Architectures for the Classification Task

As a first step, experiments involving various DL architectures were conducted with the specific aim of comparing CNNs and the SwinT. The considered architectures included (see Figure 2 for a graphical representation): (i) a 2D SwinT tiny [31] with a number of total trainable parameters = 27,914,621; (ii) a ResNet50 [32] with a number of total trainable parameters = 24,585,219; (iii) a VGG19 [33]: having a number of total trainable parameters = 20,288,579.

Considering the SwinT implementation, each patch was treated as a token with a size of 4, and its feature was formed by concatenating the raw pixel RGB values. Within each window, 7 patches were selected. In the tiny variant of the model, used in this work, the stages consist of 2, 2, 6, and 2 layers, respectively. The initial stage of the tiny variant has hidden layers with 96 channels. Each architecture employed was pre-trained on the ImageNet-1k [34] dataset and fine-tuned for three-class classification.

In all architectures, a global average pooling layer is placed at the top, followed by a dense layer with 512 neurons, a dropout layer with a dropout rate of 40% to mitigate overfitting, and a final dense layer with 3 neurons employing a softmax activation function for classification purposes. All images were resized to 224 × 224 pixels and background pixels were put to zero. To utilize the ImageNet-1k pre-trained architectures, the MRI single channel of the images was replicated across all three RGB channels of the ImageNet samples.

A data augmentation strategy was implemented to improve the generalizability of the proposed approaches. It included random (i) vertical flip; (ii) horizontal and vertical shifts, within a range of −10% to +10% of the total number of pixels in both directions; (iii) resizing of the image within 90% and 110% of the original size; (iv) rotations between −10 and +10 degrees. All experiments were performed using the Adam optimizer with a learning rate of 10^{-5} , a batch size of 16, and a total of 40 epochs utilizing the Keras framework with a TensorFlow v. 2.10.1 backend on an NVIDIA GeForce RTX 3070 Ti GPU [35]. Independent training, validation, and testing of the models were conducted using the four available MRI contrasts: T1-w, FF, F, and W images.

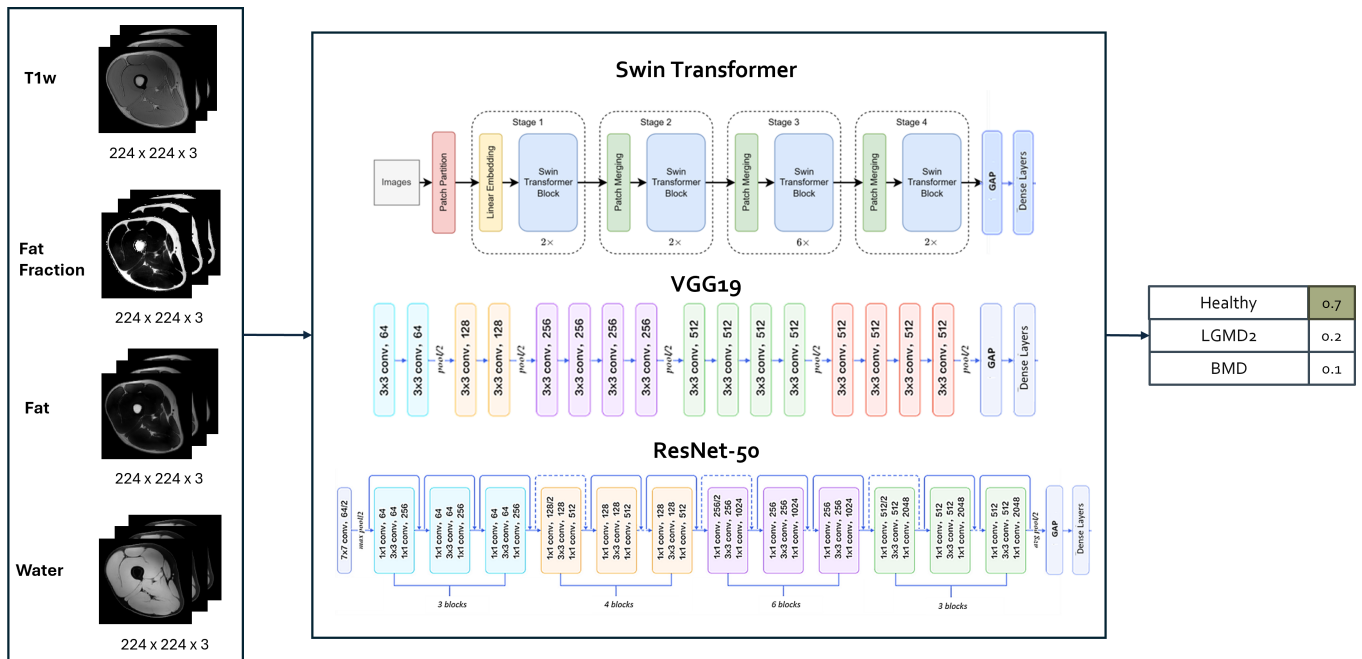


Figure 2. Architecture model selection. Figures adapted from [36–38].

2.5. Evaluation of the Classification Performance

The global performances of the classification task were assessed by means of accuracy (Acc), which represents the proportion of correctly classified slices out of the total slices in the dataset. Furthermore, to assess the class-specific performances, the F-Score, which represents the harmonic mean of precision and recall, was also calculated on the test dataset.

Formal definitions of Acc and F-score are, respectively, as follows:

$$Acc = \frac{\sum_{n=1}^N TP + \sum_{n=1}^N TN}{\sum_{n=1}^N TP + \sum_{n=1}^N TN + \sum_{n=1}^N FP + \sum_{n=1}^N FN} \quad (1)$$

$$F - score = \frac{2TP}{2TP + FP + FN} \quad (2)$$

where 'N' represents the total number of classes. 'TP' stands for True Positives, which are instances correctly identified as positive by the model. 'FP' denotes false positives, referring to instances incorrectly labeled as positive. Conversely, 'TN' refers to true negatives, which are instances correctly identified as negative, and 'FN' stands for false negatives, indicating instances that were incorrectly labeled as negative. The Acc and F-Score were computed as the average values obtained over 10 separate runs.

2.6. Statistical Analysis

The significance of the effects of the different DL architectures and contrasts on the classification task performances was assessed using a two-way ANOVA on rank-transformed data. Then, pairwise Wilcoxon tests with multiple testing corrections (false detection rate) were performed to compare groups of variables. First, the different DL architectures were compared. Then, the contrasts were compared among the implemented networks. p -values < 0.05 were considered significant.

3. Results

3.1. Global Classification Performances

The selection of DL architectures and the type of image contrast play a crucial role in classification tasks, as evidenced by the results of the two-way ANOVA analysis. Notably, both factors—DL architecture and image contrast—had a significant influence on the Acc, with p -values registering at 1.97×10^{-5} and 2.04×10^{-5} , respectively. When evaluating DL architectures against each other, grouped by image contrast types, SwinT provides the best median Acc scores. In detail, SwinT's median Acc spanned from 0.885 for F images to 0.956 for FF maps, averaging 0.919. In comparison, VGG19's median Acc varied from 0.848 with W images to 0.91 with T1w, averaging at 0.879; meanwhile, ResNet50's median Acc fluctuated from 0.844 for F images to 0.921 for FF maps, with an overall average of 0.886. In general, SwinT performs significantly better or, at least, not worse than the other DL architecture in any contrast map. In particular, Swin-T has significantly better performances than both VGG19 and ResNet50 in the case of FF ($p = 0.006$ and $p = 0.016$, respectively) and W ($p = 0.018$ and $p = 0.021$, respectively) images, whereas SwinT is significantly better than ResNet50 in the case of F images ($p = 0.048$). When considering the T1w contrast, no significant differences among architectures were found. Please, refer to Figure 3 for further details.

In the evaluation of image contrasts across different deep learning architectures, SwinT stands out with FF image achieving a median Acc of 0.956, significantly outperforming its own performance with F and W contrasts, as evidenced by p -values of 0.027. Conversely, for VGG19 and ResNet50 architectures, T1-w images lead in performance with median Accs of 0.91 and 0.92 respectively. However, these differences did not reach statistical significance when compared to other contrasts.

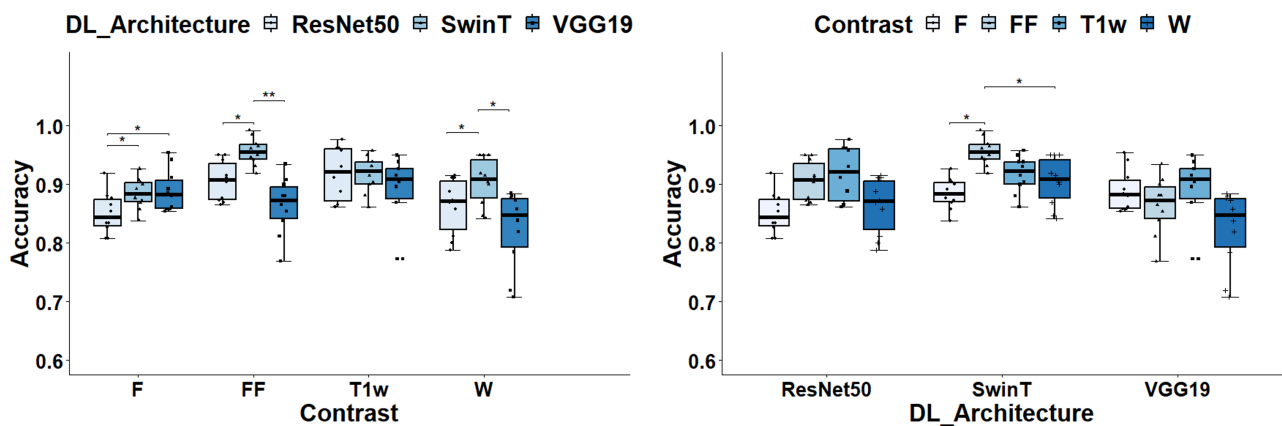


Figure 3. Global accuracy assessment for different DL architectures and image contrast. In the left panel, the box plots display ACC grouped by Image contrasts whereas in the right panel data are grouped by DL architecture. Significance markers indicate *p*-values, with * denoting $p < 0.05$ and ** signifying $p < 0.01$.

3.2. Class-Specific Classification Performances

The analysis of class-specific classification performances confirms the critical influence of both the choice of DL architecture and image contrast on F-Scores. This is supported by the results of a two-way ANOVA analysis.

In particular, for the classification of healthy subjects, the DL architecture, and image contrast were found to have a significant impact, with *p*-values of 0.01 and 0.04, respectively. When classifying LGMD2 subjects, the *p*-values were lower, at 4.35×10^{-6} and 2.51×10^{-6} , respectively. Similarly, for BMD classification, the *p*-values were 3.70×10^{-5} and 8.62×10^{-6} . As to the classification of healthy subjects, as depicted in Figure 4, all DL models and contrast levels yielded high performance, with F-Scores ranging between 0.987 and 1. ResNet50 stood out with a median F-Score of 1. For VGG19, peak performance was achieved using F contrast, reaching a median F-Score of 0.994. Similarly, SwinT achieved its highest performance with F contrast, marked by a median F-Score of 0.997.

As depicted in Figure 5, ResNet50 exhibited remarkable accuracy, misclassifying a mere single image out of 800. In comparison, VGG19 had a slightly higher misclassification rate with 10 images out of 800—considering the cumulative results from 10 independent runs—specifically in the case of F. Even SwinT demonstrated high performance, incorrectly classifying 5 out of 800 images, even in the case of F. It is important to note that most of the images that were misclassified were predominantly identified as belonging to BMD.

In the classification of LGMD2 individuals, as shown in Figure 4, the median F-Scores ranged from 0.71 to 0.95 indicating the complexity of this task compared to classifying healthy volunteers. In particular, the SwinT model exhibited superior efficacy, achieving a median F-Score of 0.95. In contrast, VGG19’s optimal performance was observed with T1w images, attaining an F-Score of 0.88. Similarly, ResNet50 achieved the best median F-Score of 0.89 when utilizing T1w images.

Figure 5 can help elucidate how critical was the classification of LGMD2 patients. Notably, a significant number of misclassifications occurred during this task. Specifically, using T1w images, VGG19’s optimal setup incorrectly classified 240 out of 900 images, resulting in a 27% error rate. Similarly, ResNet50 had a 14% error rate, misclassifying 122 out of 900 T1w images. SwinT, utilizing FF images in its most effective configuration, had the lowest error rate at 7%, with 63 misclassifications out of 900 images. It is important to highlight that the majority of these errors involved mislabeling LGMD2 slices as BMD.

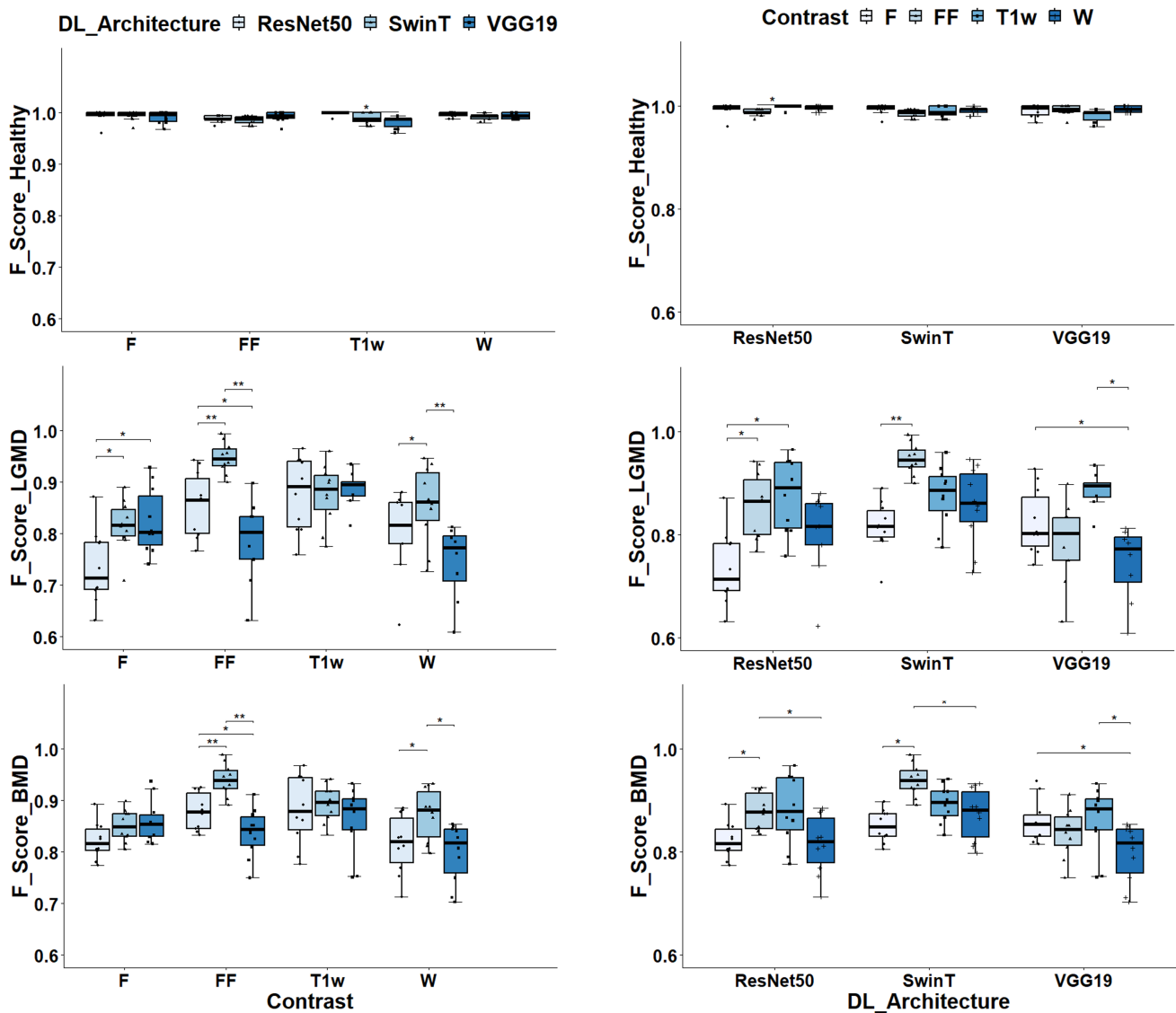


Figure 4. Class-specific F-Scores assessment for different dl architectures and image contrasts. In the left panels, the box plots display F-Scores grouped by Image contrasts and DL architectures in the case of healthy volunteers classification. In the middle panels, the same is displayed for LGMD classification whereas in the right panels, data are displayed for BMD classification task. Significance markers indicate *p*-values, with * denoting $p < 0.05$ and ** indicating $p < 0.01$.

Finally, considering the classification of BMD patients, as shown in Figure 4, median F-Scores range from 0.82 to 0.94. Specifically, SwinT obtained the best performance using FF images with a median F-Score of 0.94. Conversely, both VGG19 and ResNet50 models reached their optimal performance with T1w images, each obtaining a median F-Score of 0.88. As illustrated in Figure 5, confusion matrices reveal that SwinT’s optimal setup resulted in 16 misclassifications out of 900 BMD images, equating to a 2% error rate. In comparison, VGG19 had 9 misclassifications, while ResNet50 had 15. It is important to note that the majority of these inaccuracies occurred when BMD images were incorrectly labeled as LGMD2.

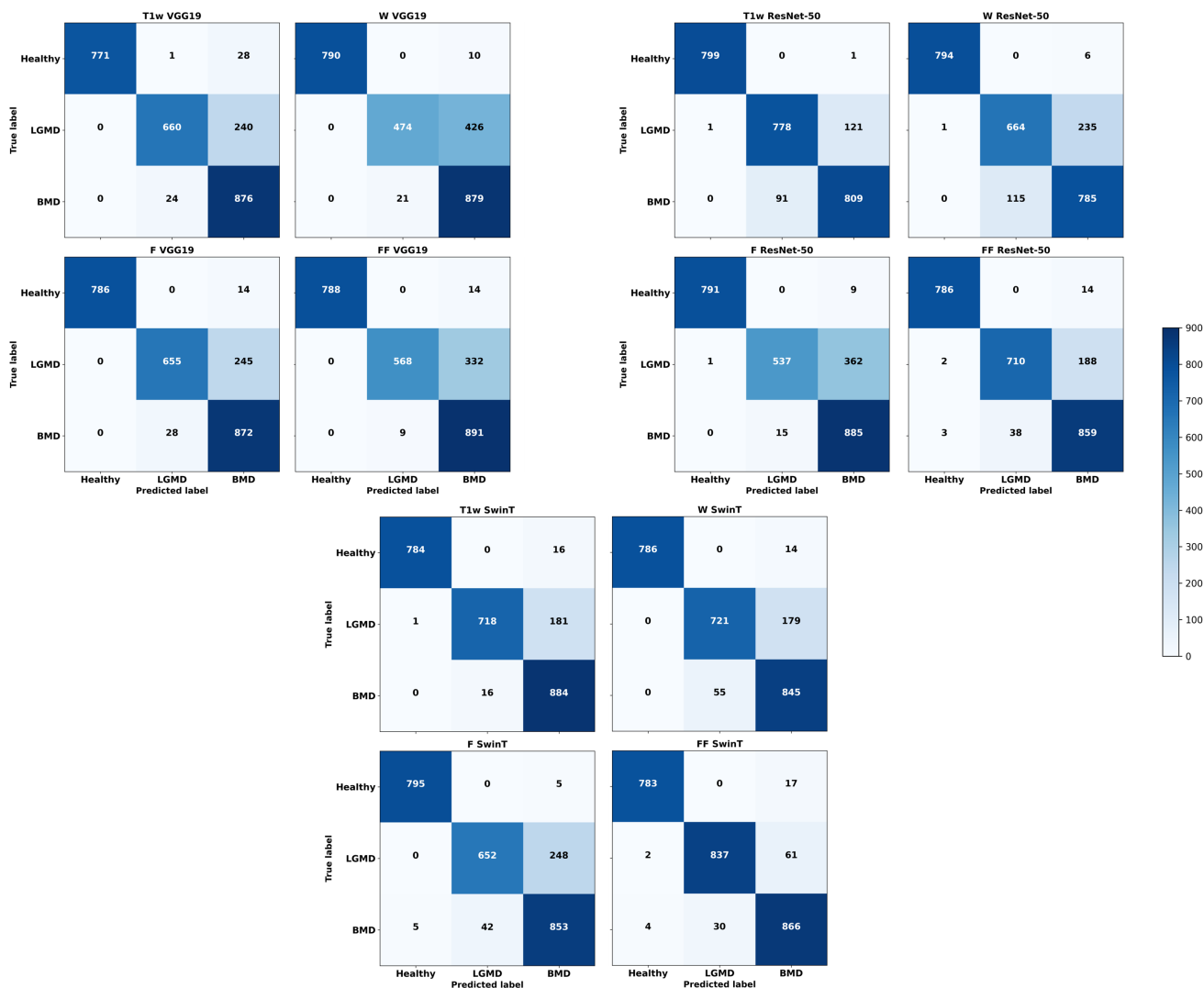


Figure 5. Cumulative confusion matrices reflecting various deep learning architectures and image contrast levels, derived from ten independent runs.

3.3. Computational Times

Table 3 indicates that training times range from 11 to 22 min, with ResNet50 taking approximately 11 min, VGG19 taking about 22 min and SwinT taking approximately 17 min per run. Regarding inference times, they range from about 0.423 to 1.087 s on average.

Table 3. Computational times.

Architecture	Training Time (min)	Classification Time (s)
ResNet50	11 ± 0.36	0.423 ± 0.014
SwinT	17 ± 0.08	1.087 ± 0.026
VGG19	22 ± 3.85	0.606 ± 0.006

Performance obtained using an NVIDIA GeForce RTX 3070 Ti GPU with 1.77 GHz Boost Clock and 8 GB RAM.

4. Discussion

This study evaluated the SwinT model’s performance in tripartite classification tasks that differentiated between healthy individuals, BDM, and LGMD2 patients using muscle MRI scans. Additionally, the research aimed to determine the most effective MRI contrast for enhancing classification accuracy, considering the various protocols typical of clinical MRI settings. The use of DL methods for classifying neuromuscular diseases from MR

images holds significant promise for enhancing clinical decision-making. This is particularly true in instances where there is substantial overlap in muscle involvement patterns, such as with BMD and LGMD2. In these conditions, muscles like the biceps femoris and semimembranosus are commonly and severely affected. Conversely, the sartorius and gracilis muscles are affected less often and with less severity [4,5]. Leveraging DL can aid in distinguishing these subtle variations, providing a more nuanced understanding that can inform treatment strategies.

In this paper, the SwinT method was evaluated against traditional convolutional neural networks (CNNs), namely VGG19 and ResNet50. These two architectures are extensively recognized for their effectiveness in classification tasks, as documented in the literature [39,40]. The results demonstrated that the SwinT, a cutting-edge deep learning model, delivered superior performance when compared to other convolutional neural network approaches when FF images are used as input. Notably, even when employing VGG19 and ResNet50 models, the overall accuracy (Acc) exceeded 0.90 in optimal configurations. This underlines the robust performance of CNNs, despite the superior results achieved with SwinT for this application. SwinT outperforms state-of-the-art techniques by effectively capturing long-range dependencies. This is achieved through its innovative self-attention mechanisms, which not only enhance the model's scalability and generalization capabilities but also decrease computational requirements if compared to ViT [21].

The best contrast for use was found to be the FF maps, particularly when combined with the SwinT approach. FF maps allow the quantitative evaluation of fat infiltration in muscle tissue by combining F and W images, making them more informative than the originating images and even T1w images. FF maps serve as a critical tool for MD diagnosis by quantifying the percentage of fat infiltration in muscles, thereby enhancing diagnostic sensitivity and supporting the goal of early MD detection through MRI [41,42]. In addition to the insights gained from FF maps, it is noteworthy that T1w images have exhibited significant performance metrics. Specifically, they have achieved high Acc and F-scores, indicating superior contrast quality when utilized with both VGG19 and ResNet50 architectures. This suggests that T1w images may provide an optimal contrast for enhancing the performance of these CNN models.

When evaluating computational efficiency, the SwinT model necessitates about 17 min for each training session. This initial step, essentially performed only once, leads to a subsequent inference time of approximately 1.087 s, on average, per batch (i.e., 260 images). DL methodologies thus provide substantial time-saving benefits in their application (i.e., in the classification of new data). In contrast, visual image classification by a radiologist could take considerably longer than this automated process.

Previous studies have already explored the use of AI to classify different neuromuscular disorders from MR images showing interesting results. For example, in a study by Al-Wesabi and colleagues [28], the main goal was a binary classification to discriminate between BMD and non-BMD and between Duchenne muscular dystrophy (DMD) and non-DMD. Their approach integrated a DL capsule network (CapsNet) for feature extraction and employed the extreme learning machine (ELM) technique for classification, resulting in a diagnostic accuracy of 96.45%. Similarly, Gopalakrishnan and colleagues [29] introduced an innovative framework, based on the synergic DL method, for the automatic two-class classification between DMD vs. non-DMD as well as BMD vs. non-BMD, achieving classification performance rates of 96.18% for DMD and 94.25% for BMD, respectively. In another recent study by Yang et al. [26], an entirely deep learning-based method was developed, with a focus on enhancing binary classification accuracy across various types and subtypes of MD. The proposed approach based on a pre-trained ResNet50 model to distinguish dystrophinopathies from non-dystrophinopathies, yielded an accuracy of 91%. In Verdú-Díaz et al. [25], researchers employed a fatty replacement mercury score and a random forest-supervised machine learning technique to develop a model capable of accurately identifying correct diagnoses, achieving an impressive accuracy rate of 95.7%. Finally, Cai and colleagues [27] developed a three-class classification problem,

distinguishing DMD, congenital muscular dystrophy (CMD), and normal subjects. They applied the ResNet18 for precise MD image classification and introduced an effective visualization method to highlight important image textures specific to DMD and CMD, visually different from each other, resulting in a classification accuracy of 91.7%.

Considering what was previously described, this research represents the first investigation into the application of the Swin Transformer (SwinT) architecture for a tripartite classification challenge using skeletal muscle MRI scans. When examining the quantitative outcomes, the SwinT, in conjunction with fat fraction (FF) images, achieved a peak accuracy rate of 95.6%. This benchmark places the methodology at the forefront of classification strategies for neuromuscular disorders, demonstrating its potential as a powerful diagnostic tool. This study acknowledges certain limitations. Primarily, the scope is narrowed by the modest cohort size and the finite selection of neuromuscular disorders examined. Additionally, the methodological approach of consolidating LGMD2A and LGMD2B into a single cohort, while potentially contentious, was employed as a means to strengthen the group's statistical significance. These limitations should be taken into account when interpreting the findings and warrant consideration in future research activities.

5. Conclusions

In conclusion, this study marks a significant step forward in the application of DL methodologies for the classification of neuromuscular disorders using skeletal muscle MRI scans. Through the evaluation of the SwinT architecture alongside traditional CNNs, we have demonstrated the superior performance of SwinT in achieving high accuracy rates, particularly when utilizing FF images as input. The results underscore the potential of SwinT as a powerful diagnostic tool for enhancing clinical decision-making in the context of neuromuscular disorders, offering nuanced insights into muscle involvement patterns that can inform treatment strategies. Furthermore, our findings highlight the importance of selecting optimal MRI contrasts, with FF maps proving to be invaluable image-based biomarkers for enhancing classification accuracy. Despite certain limitations, this study lays the groundwork for future research studies aimed at further validating and expanding upon these findings. Overall, the outcomes of this investigation contribute to advancing our understanding of utilizing AI-driven approaches for the accurate classification of neuromuscular disorders, ultimately facilitating improved patient care and management.

Author Contributions: Conceptualization, A.M., M.G.D., D.P. and G.R.; methodology, A.M., M.G.T. and N.C.; formal analysis, M.G.T., N.C. and A.M.; data curation, D.P., M.G.D. and A.M.; writing—original draft preparation, M.G.T. and A.M.; writing—review and editing, D.P., G.R., N.C. and M.G.D.; supervision, A.M. and D.P.; funding acquisition, A.M., G.R. and D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by a grant from “Italian Ministry of Health” to Denis Peruzzo (Ricerca Corrente, Scientific Institute IRCCS Medea, RC2024) and by Fondazione Cariplo and Regione Lombardia within the project “Active3 - Everyone, Everywhere, Everyday”, Ref. (2021-0612).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and approved by the Local Ethics Committee of IRCCS E. Medea, Associazione La Nostra Famiglia.

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

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

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

References

1. Emery, A.E. The muscular dystrophies. *Lancet* **2002**, *359*, 687–695. [[CrossRef](#)]
2. Lovering, R.M.; Porter, N.C.; Bloch, R.J. The muscular dystrophies: From genes to therapies. *Phys. Ther.* **2005**, *85*, 1372–1388. [[CrossRef](#)]

3. Nicolau, S.; Naddaf, E. Muscle MRI for Neuromuscular Disorders. *Pract. Neurol.* **2020**, July/August, 27–32.
4. Wattjes, M.P.; Kley, R.A.; Fischer, D. Neuromuscular imaging in inherited muscle diseases. *Eur. Radiol.* **2010**, *20*, 2447–2460. [[CrossRef](#)]
5. Tasca, G.; Iannaccone, E.; Monforte, M.; Masciullo, M.; Bianco, F.; Laschena, F.; Ottaviani, P.; Pelliccioni, M.; Pane, M.; Mercuri, E.; et al. Muscle MRI in Becker muscular dystrophy. *Neuromuscul. Disord.* **2012**, *22*, S100–S106. [[CrossRef](#)]
6. Fanin, M.; Angelini, C. Progress and challenges in diagnosis of dysferlinopathy. *Muscle Nerve* **2016**, *54*, 821–835. [[CrossRef](#)]
7. Manzur, A.Y.; Muntoni, F. Diagnosis and new treatments in muscular dystrophies. *Postgrad. Med. J.* **2009**, *85*, 622–630. [[CrossRef](#)]
8. Angelini, C. LGMD. Identification, description and classification. *Acta Myol.* **2020**, *39*, 207.
9. Okubo, M.; Minami, N.; Goto, K.; Goto, Y.; Noguchi, S.; Mitsuhashi, S.; Nishino, I. Genetic diagnosis of Duchenne/Becker muscular dystrophy using next-generation sequencing: Validation analysis of DMD mutations. *J. Hum. Genet.* **2016**, *61*, 483–489. [[CrossRef](#)]
10. Nigro, V.; Savarese, M. Next-generation sequencing approaches for the diagnosis of skeletal muscle disorders. *Curr. Opin. Neurol.* **2016**, *29*, 621–627. [[CrossRef](#)]
11. Ghaoui, R.; Cooper, S.T.; Lek, M.; Jones, K.; Corbett, A.; Reddel, S.W.; Needham, M.; Liang, C.; Waddell, L.B.; Nicholson, G.; et al. Use of whole-exome sequencing for diagnosis of limb-girdle muscular dystrophy: Outcomes and lessons learned. *JAMA Neurol.* **2015**, *72*, 1424–1432. [[CrossRef](#)]
12. Joyce, N.C.; Oskarsson, B.; Jin, L.W. Muscle biopsy evaluation in neuromuscular disorders. *Phys. Med. Rehabil. Clin.* **2012**, *23*, 609–631. [[CrossRef](#)]
13. Li, K.; Dortch, R.D.; Welch, E.B.; Bryant, N.D.; Buck, A.K.; Towse, T.F.; Gochberg, D.F.; Does, M.D.; Damon, B.M.; Park, J.H. Multi-parametric MRI characterization of healthy human thigh muscles at 3.0 T—relaxation, magnetization transfer, fat/water, and diffusion tensor imaging. *NMR Biomed.* **2014**, *27*, 1070–1084. [[CrossRef](#)]
14. Venturelli, N.; Tordjman, M.; Ammar, A.; Chetrit, A.; Renault, V.; Carlier, R.Y. Contribution of muscle MRI for diagnosis of myopathy. *Rev. Neurol.* **2023**, *179*, 61–80. [[CrossRef](#)]
15. Malartre, S.; Bachasson, D.; Mercy, G.; Sarkis, E.; Anquetil, C.; Benveniste, O.; Allenbach, Y. MRI and muscle imaging for idiopathic inflammatory myopathies. *Brain Pathol.* **2021**, *31*, e12954. [[CrossRef](#)]
16. Garibaldi, M.; Nicoletti, T.; Bucci, E.; Fionda, L.; Leonardi, L.; Morino, S.; Tufano, L.; Alfieri, G.; Lauletta, A.; Merlonghi, G.; et al. Muscle magnetic resonance imaging in myotonic dystrophy type 1 (DM1): Refining muscle involvement and implications for clinical trials. *Eur. J. Neurol.* **2022**, *29*, 843–854. [[CrossRef](#)]
17. Forbes, S.C.; Willcocks, R.J.; Rooney, W.D.; Walter, G.A.; Vandenborne, K. MRI quantifies neuromuscular disease progression. *Lancet Neurol.* **2016**, *15*, 26–28. [[CrossRef](#)]
18. Tobore, I.; Li, J.; Yuhang, L.; Al-Handarish, Y.; Kandwal, A.; Nie, Z.; Wang, L. Deep learning intervention for health care challenges: Some biomedical domain considerations. *JMIR MHealth UHealth* **2019**, *7*, e11966. [[CrossRef](#)]
19. Kim, M.; Yun, J.; Cho, Y.; Shin, K.; Jang, R.; Bae, H.j.; Kim, N. Deep learning in medical imaging. *Neurospine* **2019**, *16*, 657. [[CrossRef](#)]
20. He, K.; Gan, C.; Li, Z.; Rekek, I.; Yin, Z.; Ji, W.; Gao, Y.; Wang, Q.; Zhang, J.; Shen, D. Transformers in medical image analysis. *Intell. Med.* **2023**, *3*, 59–78. [[CrossRef](#)]
21. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
22. Wei, C.; Ren, S.; Guo, K.; Hu, H.; Liang, J. High-resolution Swin transformer for automatic medical image segmentation. *Sensors* **2023**, *23*, 3420. [[CrossRef](#)]
23. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12009–12019.
24. Piñeros-Fernández, M.C. Artificial intelligence applications in the diagnosis of neuromuscular diseases: A narrative review. *Cureus* **2023**, *15*, e48458. [[CrossRef](#)]
25. Verdú-Díaz, J.; Alonso-Pérez, J.; Nuñez-Peralta, C.; Tasca, G.; Vissing, J.; Straub, V.; Fernández-Torrón, R.; Llauger, J.; Illa, I.; Díaz-Manera, J. Accuracy of a machine learning muscle MRI-based tool for the diagnosis of muscular dystrophies. *Neurology* **2020**, *94*, e1094–e1102. [[CrossRef](#)]
26. Yang, M.; Zheng, Y.; Xie, Z.; Wang, Z.; Xiao, J.; Zhang, J.; Yuan, Y. A deep learning model for diagnosing dystrophinopathies on thigh muscle MRI images. *BMC Neurol.* **2021**, *21*, 13. [[CrossRef](#)]
27. Cai, J.; Xing, F.; Batra, A.; Liu, F.; Walter, G.A.; Vandenborne, K.; Yang, L. Texture analysis for muscular dystrophy classification in MRI with improved class activation mapping. *Pattern Recognit.* **2019**, *86*, 368–375. [[CrossRef](#)]
28. Al-Wesabi, F.N.; Obayya, M.; Hilal, A.M.; Castillo, O.; Gupta, D.; Khanna, A. Multi-objective quantum tunicate swarm optimization with deep learning model for intelligent dystrophinopathies diagnosis. *Soft Comput.* **2022**, *27*, 13077–13092. [[CrossRef](#)]
29. Gopalakrishnan, T.; Sudhakaran, P.; Ramya, K.; Sathesh Kumar, K.; Al-Wesabi, F.N.; Alohal, M.A.; Hilal, A.M. An Automated Deep Learning Based Muscular Dystrophy Detection and Classification Model. *Comput. Mater. Contin.* **2022**, *71*, 305–320.

30. Eggers, H.; Börnert, P. Chemical shift encoding-based water–fat separation methods. *J. Magn. Reson. Imaging* **2014**, *40*, 251–268. [[CrossRef](#)]
31. Sayak, P. Implementation of Swin Transformers in TensorFlow along with Converted Pre-Trained Models, Code for off-the-Shelf Classification and Fine-Tuning. 2022. Available online: <https://github.com/sayakpaul/swin-transformers-tf> (accessed on 21 April 2024).
32. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
33. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
34. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
35. Chollet, F. Keras. 2023. Available online: <https://keras.io/> (accessed on 4 April 2024).
36. Tang, Y. YanTang’s Blog. 2023. Available online: <https://ehehe.cn/2023/08/09/Swin/> (accessed on 5 April 2024).
37. Rastogi, A. ResNet50. 2022. Available online: <https://blog.devgenius.io/resnet50-6b42934db431> (accessed on 5 April 2024).
38. Varshney, P. VGG19. 2020. Available online: <https://www.kaggle.com/code/blurredmachine/vggnet-16-architecture-a-complete-guide> (accessed on 5 April 2024).
39. Krishnapriya, S.; Karuna, Y. Pre-trained deep learning models for brain MRI image classification. *Front. Hum. Neurosci.* **2023**, *17*, 1150120. [[CrossRef](#)]
40. Swati, Z.N.K.; Zhao, Q.; Kabir, M.; Ali, F.; Ali, Z.; Ahmed, S.; Lu, J. Brain tumor classification for MR images using transfer learning and fine-tuning. *Comput. Med Imaging Graph.* **2019**, *75*, 34–46. [[CrossRef](#)]
41. Gaeta, M.; Messina, S.; Mileto, A.; Vita, G.L.; Ascenti, G.; Vinci, S.; Bottari, A.; Vita, G.; Settineri, N.; Bruschetta, D.; et al. Muscle fat-fraction and mapping in Duchenne muscular dystrophy: Evaluation of disease distribution and correlation with clinical assessments: preliminary experience. *Skelet. Radiol.* **2012**, *41*, 955–961. [[CrossRef](#)]
42. Morrow, J.M.; Sinclair, C.D.; Fischmann, A.; Machado, P.M.; Reilly, M.M.; Yousry, T.A.; Thornton, J.S.; Hanna, M.G. MRI biomarker assessment of neuromuscular disease progression: A prospective observational cohort study. *Lancet Neurol.* **2016**, *15*, 65–77. [[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.