International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Diagnosis of Diabetic Peripheral Neuropathy Based on Muscle Ultrasound with Hybrid Deep Transfer and Machine Learning

Kadhim Al-Barazanchi[1*] , Ali H. Al-Timemy[2] , Zahid Kadhim[3]

[1] Biomedical Engineering Department, College of Engineering, Al-Nahrain University, Baghdad 10070, Iraq
[2] Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad 10070, Iraq
[3] College of Medicine, University of Babylon, Babylon 51001, Iraq

Corresponding Author Email: st.kadhim.k.hasan@ced.nahrainuniv.edu.iq

ABSTRACT

One of the common neuromuscular disorders in diabetic patients is diabetic peripheral neuropathy (DPN), which leads to a range of complications, from tingling sensations to limb loss. Quantitative assessment through muscle ultrasound has become a valuable tool for diagnosing DPNs. This study creates a hybrid model that employs deep transfer learning via pre-trained convolutional neural networks (CNN) to extract features and machine learning algorithms to classify ultrasound images. The collection consists of 6200 ultrasound images of the tibialis anterior (TA) muscle obtained from 53 individuals. The effectiveness of VGG19, Shuffle Net, and ResNet101 is assessed by visualizing the features extracted through gradient-weighted class activation mapping (Grad-CAM). Moreover, t-distributed stochastic neighbor embedding (t-SNE) is employed to investigate the clustering of labeled features and formulate hypotheses regarding their interconnections. The support vector machine (SVM) and logistic regression classifier are assessed using metrics consisting of confusion matrix, accuracy, the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and F1-score. The ResNet101 combined with the SVM model achieves 99.5% accuracy for training and validation and 75.8% for testing. Integrating deep transfer learning with machine learning in a hybrid classification framework greatly enhances the precision and dependability of diagnostic systems. This innovative approach provides comprehensive performance analysis, distinguishing healthy and DPN.

1. INTRODUCTION

Diabetic peripheral neuropathy (DPN) is a neuromuscular condition that affects people who have diabetes mellitus. In the United States, it's estimated that around 28% of adults with diabetes experience this issue [1]. The primary risk factors include neuropathic pain, diminished sensation, ulcers in the lower limbs, and amputations, all of which negatively affect quality of life and restrict daily activities [2]. Early identification of DPN helps prevent severe consequences and significantly improves prognosis [3]. The diagnosis of DPN relies on quantitative evaluations as well as clinical findings. Nerve conduction tests assess the performance of peripheral nerves and determine the onset of DPN, but they mainly concentrate on large nerve fibers and require specialized expertise [4]. One of the important tools for diagnosing neuromuscular disorders is muscle ultrasound, due to its capability to identify muscle atrophy, intramuscular fibrosis, and fatty infiltration [5]. Improvements in ultrasound technology have enhanced muscle tissue images, making neuromuscular ultrasonography valuable for assessing nerve and muscular disorders [6]. Muscle ultrasound complements clinical and electrophysiological tests by providing information about abnormalities in neuromuscular diseases [7, 8]. Grayscale analysis indicates that diseased muscles appear more echogenic on ultrasounds than healthy ones [9]. Additionally, quantitative analysis of muscle echo intensity (EI) offers a promising new method for evaluating muscle quality and improving diagnostic accuracy [10, 11].

Advancements in computer vision and artificial intelligence (AI) have significantly impacted computer-aided diagnosis (CAD) systems, particularly in the field of medical imaging research [12]. Scientists are employing AI-driven techniques to examine intricate patterns and interpret imaging data for possible diagnostic applications. In particular, CNNs have shown significant advancements in recognizing and analyzing images [13]. Methods utilizing AI in muscle ultrasound can address challenges associated with evaluations that depend on physicians, including inconsistencies in capturing images and diagnosing muscle disorders. Current research in AI-based muscle ultrasonography focuses on potentially applying machine learning techniques to enhance everyday clinical practice performance [14, 15]. Deep learning, especially transfer learning with CNNs, is increasingly used for muscle classification and analysis in ultrasound imaging [16, 17]. It offers practical approaches for handling limited data while

conserving time and resources. Transfer learning techniques involve utilizing a pre-trained network as a feature extractor or adjusting a pre-trained network specifically for medical data [18, 19]. High-dimensional features greatly influence the performance of machine learning classification, and deep learning models continue to be viewed as opaque systems. Multiple metrics for evaluating the performance of deep learning models can be utilized together to provide a thorough evaluation. This method entails placing a blank window over an image before using the model to assess how each area affects the model's prediction [20, 21].

Muscular ultrasound has not been extensively studied, particularly regarding the quantitative aspects that could enhance physicians' diagnostic capabilities for DPN. The quantitative evaluation of muscle ultrasound is considered more dependable and sensitive than visual assessment, which frequently requires the input of a skilled physician. Further research is required to explore the potential for automating this quantification process. This study aims to establish a muscle ultrasound-based diagnostic system that employs hybrid deep transfer learning and machine learning methodologies as a novel and precise diagnostic approach.

In this regard, the research study contributes explicitly to the following ways:

1) State-of-the-art techniques for assessing muscle ultrasound are considered an additional diagnostic resource for DPN and are built upon the CAD system.
2) Introduces an innovative hybrid model combining the deep transfer learning technique and machine learning algorithms to construct a robust prediction model.
3) This study comprehensively evaluates the muscle ultrasound classification system. It employs gradient-weighted class activation mapping and *t*-distributed stochastic neighbor embedding to visualize the extracted features, enhancing physicians' understanding of CAD prediction.

2. RELATED WORK

This part examines contemporary studies on using muscle ultrasound to identify DPN and neuromuscular disorders. It emphasizes the use of deep neural networks and machine learning, underscoring techniques such as feature extraction, and the use of pre-trained models, along with approaches for assessing performance in image classification. Scientists have employed various methods for feature extraction; König et al. [22] extracted first-order statistics, wavelet-based, and Haralick's features from ultrasound images. They utilized feature selection and reduction techniques to determine a consistent set of features and investigated two linear classifiers: Fisher's classifier, support vector machine (SVM), and nonlinear k-nearest neighbour for myositis detection. This system utilized principal component analysis for feature reduction and linear SVM for discriminating between healthy and pathological muscle tissue, achieving a classification accuracy of 87%. Nodera et al. [23] examining lower leg ultrasound images by extracting texture features using histogram analysis, grey-level co-occurrence matrix, neighborhood grey-level difference matrix, grey-level run length matrix, and grey-level zone length matrix. Classification was performed with logistic regression, SVM, and random forest, achieving an accuracy of 78.4% with random forest.

Conversely, numerous researchers have employed deep-learning techniques to categorize muscle ultrasound images, like Ahmed et al. [24], a revised lightweight YOLOv5 architecture was suggested, incorporating a convolutional block attention module, spatial pyramid pooling-fast plus, and an exponential linear unit activation function to enable automatic detection and classification of inflammatory myopathies. This model achieves an accuracy of 98% for binary and multiclass classification. Additionally, Uçar [25], Liao et al. [26] and Zhou et al. [27] developed a CAD system using deep learning methods combining VGG16, VGG19, multi-scale fusion and attention mechanisms to segment and diagnose musculoskeletal ultrasound images automatically. Burlina et al. [28] compared the classification accuracy of an automated deep learning method with a semi-automated machine learning for diagnosing neuromuscular ultrasound. The deep learning method achieved an accuracy of 86.6%, compared to the machine learning method achieved 84.3%. From previous studies, no hybrid methodology combining deep transfer learning and machine learning was proposed and utilized to classify muscle ultrasound in diagnosing DPN. Moreover, this study highlights the application of performance analysis through innovative techniques such as gradient-weighted class activation mapping (Grad-CAM) and t-distributed stochastic neighbor embedding (t-SNE). These methods facilitate a comparative examination of various pre-trained models and their effects on classification accuracy and diagnostic efficacy, thereby addressing gaps identified in prior research.

3. MATERIALS AND METHODS

This research seeks to develop and evaluate a combined deep transfer feature extraction and machine learning model for the detection of DPN in muscle ultrasound images. The suggested method, illustrated in Figure 1, leverages transfer and machine learning techniques to improve the accuracy and dependability of the DPN diagnostic system. This section outlines our sequential process for creating and assessing the hybrid model.

3.1 Dataset

This study was conducted with the neurophysiology clinical center at Ghazi Al Hariri Surgical Hospital in Baghdad, Iraq. A case-control, retrospective investigation included 26 DPN patients diagnosed with type 2 diabetes mellitus and 27 healthy controls (CTR) from August 2022 to May 2023. The demographic information of the 53 participants is presented in Table 1. The research adhered to the ethical principles outlined in the Declaration of Helsinki and received approval from the local health ethics committee. The electrodiagnostic evaluation confirmed DPN, and an HbA1c test was part of the exclusion criteria, ruling out other neuromuscular disorders, kidney failure, and cancer in both groups. An ultrasonographic assessment used a Philips iU22 ultrasound with a 5-12MHz linear probe, following musculoskeletal presets. Participants were positioned supine and relaxed while a qualified physician performed the ultrasound using specific imaging presets. A skilled physician performed the ultrasound imaging of the muscle, considering its anatomical positioning and citing a previously published article that serves as a resource for researchers conducting ultrasound assessments of the TA [29].

The presets set up on the ultrasound machine were utilized for recording, with the musculoskeletal general preset configured to a gain of 50%, a compression level of 62, medium pressure applied, and a depth of 3 cm.

Ultrasound images were obtained from the muscle belly, and the best image showcasing muscle fibers in a transverse view, with the bone indicated in the background, was saved. The images were kept in DICOM format on the ultrasound machine and labeled with the patient code number and muscle designation for future reference and analysis. The appearance of muscle tissue on ultrasound varies according to the muscle type, examination view, and age. Healthy muscle appears to have low echogenicity, with black visualization, and is easily distinguished from surrounding tissue, whereas diseased muscles may exhibit increased echogenicity. Ultrasound is a user-dependent imaging modality with variations in screening direction based on the probe orientation and anatomical positioning. Therefore, multiple images are acquired per subject to overcome this limitation and solve the model's bias issue. Deep learning and machine learning models also require substantial quantities of data for both training and validation. To guarantee accurate outcomes, thousands of images from the subjects have been gathered to enhance the model's performance.
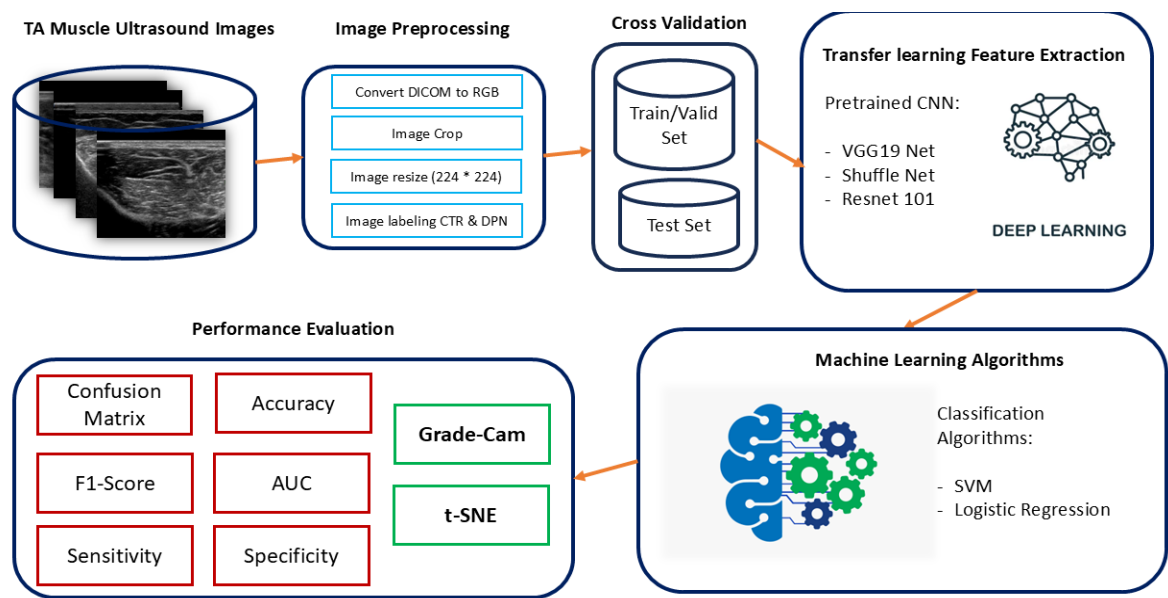

Figure 1. The proposed hybrid DPN diagnosis model architecture

Table 1. Demographic of the study group dataset

Variable	DPN (26)	Control (27)
Age (years)	51.5±9.3	42.5±9.6
Gender Male/Female Ratio	20/6	20/7

3.2 Image preprocessing

Image preprocessing plays a vital role in our DPN diagnosis system, as it sets the stage for deeper analysis of input images. This involves various techniques applied to raw images to improve quality and highlight essential information. A thorough approach to image preprocessing can significantly enhance the precision and effectiveness of later stages of analysis. The first step involves converting images exported from the ultrasound machine in DICOM format to a JPG file format image. This format preserves image resolution and details, making it easy to analyze for diagnosis purposes based on the quantification of EI. The ultrasound images contain information about the ultrasound machine, examination, patient details, annotations, and remarks. This research is primarily centered on ultrasound images of muscles, and we have trimmed the images to display solely the muscle region, leaving out any extraneous details. To maintain uniform dimensions of input images and decrease computational load, the images were adjusted to a standard size of 224×224 pixels. This procedure is vital when using pre-trained deep learning models, which typically have defined size standards. Adjusting the images to correspond with the model's input size enables us to leverage the model's learned features effectively. The image dataset contained object labels, including subject type, number, and image sequence. To simplify the analysis, the object labels were modified to include only the subject type and sequence of all images in the group dataset. This change makes data management for further analysis more straightforward and systematic.

3.3 Transfer learning feature extractor

Deep feature extraction involves using CNNs to extract informative features from raw data. This process is known as transfer learning and has been successfully applied using a dataset of millions of photos with 1000 types of nature imagery. These features capture high-level representations and are extracted unsupervised, meaning no external guidance influences the information obtained from the image's pixels [30, 31]. VGG19, Shuffle, and ResNet101 architectures were conducted as deep feature extraction models for our DPN diagnostic system. These three models were selected due to their excellent performance with medical image datasets [32]. The VGG19 model, uses multiple sequential 3×3 convolutional kernels instead of larger ones. The convolution stride is fixed to 1 pixel; the padding is 1 pixel for 3×3 convolution layers. Spatial pooling is carried out by five max-pooling layers. Max-pooling is performed over a 2×2-pixel window, with stride 2. The configuration of the fully connected layers is the same in all networks. All hidden layers are equipped with the rectification non-linearity function [33].

Shuffle Net includes 172 layers of depth-wise convolution, channel shuffle, and pointwise group convolution. The design begins with a bottleneck residual block that uses a 3×3 depth-wise convolution for efficiency. The initial 1×1 layer is replaced with pointwise group convolution, and a channel shuffle is included. A second pointwise group convolution restores the channel dimension to match the shortcut paths, and skipping an additional shuffle has minimal effect. Instead of rectified linear units after depth-wise convolution, batch normalization and nonlinearity are applied [34]. ResNet101 utilizes residual connections between layers to mitigate the vanishing gradient problem while effectively reducing the additional parameters within a structure comprising 101 layers. The network performs down-sampling directly through convolutional layers incorporating a stride of 2 and employing 3×3 filters. Conclusively, it features a global average pooling layer, resulting in 34 weighted layers. The integration of shortcut connections transforms the network into its corresponding residual version [35].

3.4 Machine learning classification algorithms

The DPN diagnosis investigation utilized hybrid classical machine learning techniques and deep feature extraction. This approach minimized the number of dimensions, tackled the issue of class imbalance, enhanced the detection of out-of-distribution samples, performed exploratory data analysis, and utilized model compression methods, along with various other benefits. The study employed two machine learning algorithms to accomplish its goals. The SVM algorithm found the best plane to separate classes in a sample space. It aims to create the widest margin between classes efficiently, using derivatives to calculate the widest margin among potential solutions [36]. Conversely, logistic regression is a binary classification method that estimates the probability of an event happening, with values ranging from 0 to 1, utilizing the logistic or sigmoid function. Its goal is to determine the best weights that reduce the error by maximizing the likelihood function [37]. To train and assess the model, 6,200 ultrasound images of the TA muscle from 53 individuals were used. The dataset is split into two parts: one for training/validation and the other for testing. Table 2 presents the specifics of the dataset concerning subjects and images. Our proposed model is evaluated using a completely independent test set. During its training and validation, five-fold cross-validation was implemented.

Table 2. Data set division for training and evaluation

Class	Train/Valid	Test
DPN	20 Subjects / 2400 Images	7 Subjects / 700 Images
CTR	20 Subjects / 2400 Images	6 Subjects / 700 Images

The MATLAB platform offered essential deep learning and machine learning capabilities for our research. Its framework facilitated the design, implementation, validation, and testing of algorithms, enabling efficient trials and effective analysis of our proposed method for diagnosing (DPN). Our model was trained using different feature extraction and classification parameters in this project. To extract deep features from the pre-trained networks, the images were input to each pre-trained network and directly extracted the feature vectors at the fully connected layer without retraining the network [38]. Each model can be used as a feature extractor by excluding the

fully connected layers at the network's end. Then, the deep features are fed into machine learning algorithms; SVM and logistic regression, separately in the hybrid model.

SVM classifiers were utilized with a Gaussian Kernel function and a Kernel scale parameter set to 32. Additionally, for the logistic regression trained model object, binary Gaussian kernel classification was employed using random feature expansion with an automatic number of dimensions in the expanded space and a Kernel scale parameter.

3.5 Performance evaluation

We utilized various strategies to evaluate the performance of our proposed method and confirm its effectiveness. These benchmarks aim to assess different facets of the method's efficiency, and they are detailed as follows:

Grad-CAM is a technique used in CNN to visualize significant features of an input image. It generates heat maps to indicate the relevance of particular pixels or areas related to the extracted features, which assists in recognizing and classifying objects. By leveraging gradient information from the last convolutional layer, Grad-CAM assigns importance scores to each neuron, which helps to emphasize class-specific characteristics for making decisions. Guided Grad-CAM enhances the precision of category visualization and aids in identifying distinguishing attributes [39].

The t-SNE method illustrates high-dimensional data on a two-dimensional map while preserving its intrinsic organization. It is beneficial for visualizing data with numerous dimensions, such as feature vectors obtained from a deep transfer learning model [40].

A confusion matrix is an organized table used to assess an algorithm's performance. It includes false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN), where P and N denote positive and negative samples from the original dataset. Accuracy is an important measure for evaluating a classification model. It shows how many predictions the model got right compared to the total number of examples. In simple terms, it indicates how well the model correctly classifies data into the right categories.

Sensitivity, often called the true positive rate, assesses the model's effectiveness in recognizing positive cases. It is determined by dividing the number of accurately identified positive results (TP) by the total number of genuine positive instances, including true positives and false negatives. The emphasis of sensitivity is on the model's capability to detect the existence of a condition. Specificity assesses how effectively the model can recognize negative cases. It evaluates the model's capacity to accurately identify negative instances, or true negatives, out of all cases lacking the condition. This category encompasses both true negatives and false positives. A high level of specificity indicates that the model effectively minimizes false positives.

The F1-score combines sensitivity and specificity into a single metric. It balances precision (which refers to correct positive predictions) and recall (sensitivity) by considering both false positives and false negatives. The receiver operating characteristic curve (ROC) is a valuable method for illustrating a classifier's performance across various thresholds. The area beneath the ROC curve (AUC) indicates the model's ability to differentiate between classes. A greater AUC value signifies improved accuracy in the model's predictions, whereas a lower AUC implies that the model might struggle to produce dependable predictions [41].

4. RESULT ANALYSIS

This part outlines the performance analysis results for our suggested hybrid model, which utilizes three deep transfer learning feature extractors and two machine learning algorithms to diagnose DPN. This results analysis addresses the performance evaluation in two parts, discussing the results obtained and comparing them based on different approaches.
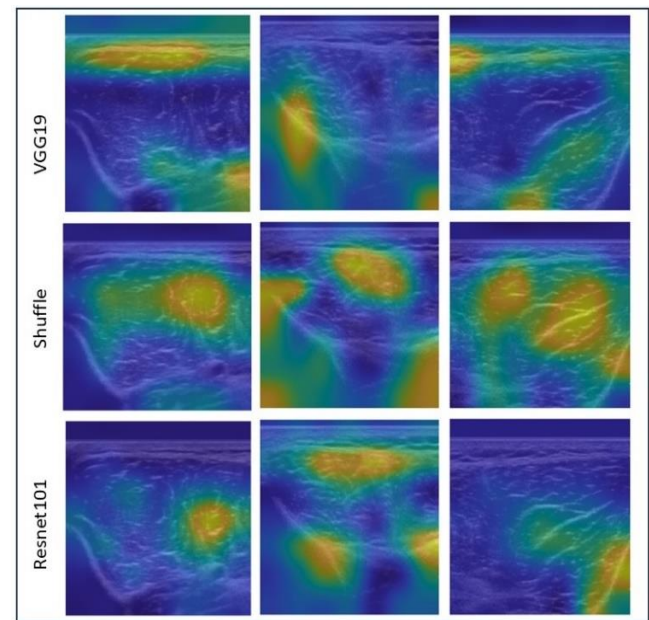

Figure 2. Health ultrasound images of three subjects are in columns with Grad-Cam heat map visualization among VGG19, shuffle, and ResNet101 in rows
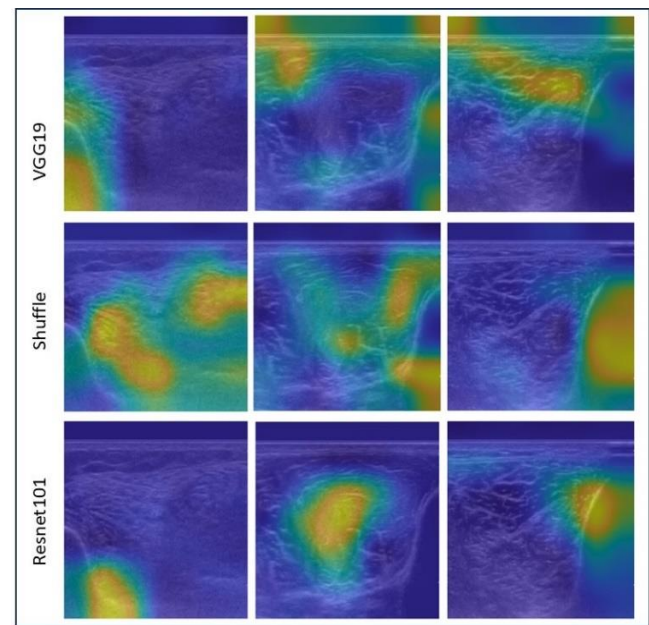

Figure 3. DPN ultrasound images of three subjects are in columns with Grad-Cam heat map visualization among VGG19, shuffle, and ResNet101 in rows

4.1 Visualization evaluation

An investigation and explanation of the deep transfer learning feature extractor using the Grad-CAM mapping for three CNNs is shown in Figure 2 for healthy and Figure 3 for DPN ultrasound images. To better understand the regions of the extracted features, heat maps with variable intensity highlighted specific areas in the input image, providing a clear visualization of the deep learning model's operation. The ResNet101 and shuffle CNNs exhibit high intensity on the muscle belly, while VGG19 focuses on the muscle borders. The lines on Grad-CAM images delineate the areas with features significantly impacting the classification outcomes. In CAD systems, this representation could enhance physicians' trust in AI conclusions and lower the incidence of incorrect diagnoses in qualitative evaluations of muscle ultrasound.

Additionally, Figures 4 and 5 depict the t-SNE visualization of high-dimensional feature datasets extracted from ultrasound images of training/validation and testing sets. This visualization assigns each sample a location on a 2D map, providing a discriminative visualization for the class of interest and aiding in the intuitive explanation of predictions. It is evident that the train/validation and test feature datasets are non-linearly separated, resulting in overlapping data clusters that guide the selection of classification algorithms. Through qualitative analysis of the feature dataset, the performance evaluation revealed that t-SNE enhanced classification accuracy for both the training/validation and testing datasets.

4.2 Classification performance

The classification assessment outcomes are based on the performance metrics presented in Table 3 for the feature datasets utilized in both training and validation. These datasets implemented 5-fold cross-validation to assess the model's effectiveness on an independent dataset and accurately identify issues such as overfitting or selection bias for meaningful insights. This evaluation employed SVM alongside logistic regression using the extracted deep features. The findings demonstrate that the hybrid ResNet101 combined with SVM attained the peak classification accuracy of 99.5% for training and validation, whereas the sensitivity, specificity, F1-score, and AUC were recorded at 99.4%, 99.5%, 99.4%, and 99.9%, respectively.

The other hybrid model shows optimal accuracy and other metric performance, ResNet101 with logistic regression archive accuracy of 96.2%, and VGG19 shows accuracies of 97.9% with SVM and 95% with logistic regression. Shuffle Net indicates a classification accuracy of 99.3% with SVM and 95.8% with logistic regression. Meanwhile, the sensitivity, specificity, F1-score, and AUC are variable with a reasonable range among different models based on variation of different classification algorithms.

Additionally, Table 4 indicates the results of the testing feature datasets. The performance metrics analysis of various algorithms for the DPN diagnosis system provides a detailed overview of the best result obtained from Resnet, with the SVM model achieving an accuracy of 75.8%. Moreover, the sensitivity, specificity, F1-score, and AUC were 81.1%, 71.9%, 73.5%, and 83.7% respectively. The VGG19 shows a classification accuracy of 75.1% with SVM and 73.4% with logistic regression.

On the other hand, Shuffle Net achieved classification accuracy of 73.2% with SVM and 70.4% with logistic regression. The rest of the proposed model results show that Resnet with logistic regression indicates the lowest classification accuracy of 70.1% and undesired performance evaluation results. Also, there are variations in the sensitivity,

specificity, F1-score, and AUC between the hybrid models, indicating the efficiency of the classification algorithms in identifying the healthy and DPN ultrasound images from different subjects based on the EI quantification.

Figure 6 displays the confusion matrix for the binary classifier, which utilizes a hybrid model combining the ResNet101 deep feature extractor with the SVM machine learning algorithm. The values in the matrix indicate the quantity of images corresponding to each class. Our model demonstrated higher accuracy during testing in predicting DPN images than healthy/control images.
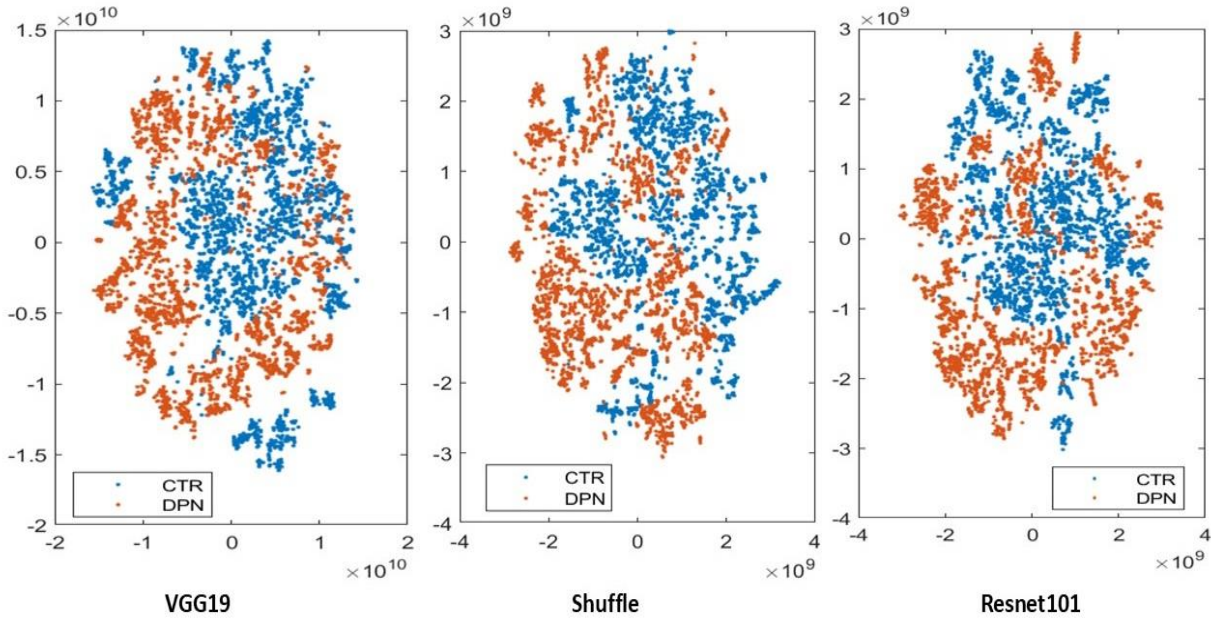

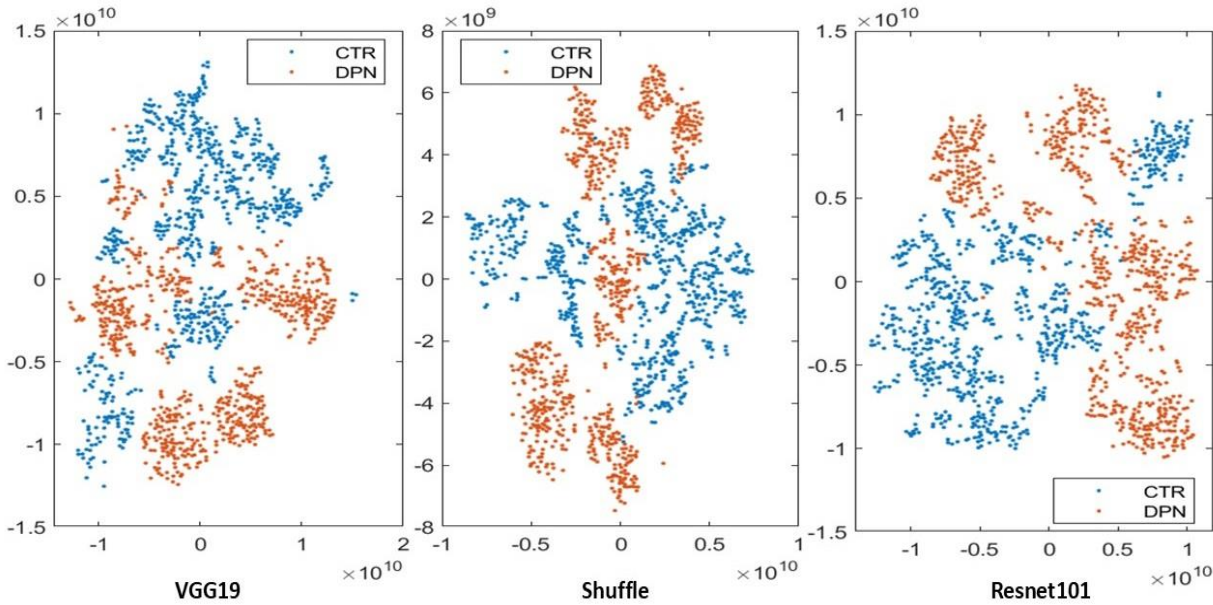**Figure 4.** Train/validate feature datasets from three transfer learning networks t-SNE visualization

Figure 5. Test feature datasets from three transfer learning networks t-SNE visualization

Table 3. Metric analysis for train/valid dataset

Hybrid Model	Accuracy	Sensitivity	Specificity	F1-Score	AUC
ResNet101+SVM	99.5%	99.4%	99.5%	99.4%	99.9%
VGG19+SVM	97.9%	97.3%	98.5%	97.9%	99.0%
Shuffle+SVM	99.3%	99.3%	99.3%	99.3%	99.9%
ResNet101+Logistic Regression	96.2%	94.8%	97.7%	96.2%	99.3%
VGG19+Logistic Regression	95.0%	93.8%	96.1%	95.0%	99.0%
Shuffle+Logistic Regression	95.8%	95.0%	96.7%	95.9%	99.3%

Table 4. Metric analysis for train/valid dataset

Hybrid Model	Accuracy	Sensitivity	Specificity	F1-Score	AUC
ResNet101+SVM	75.8%	81.1%	71.9%	73.5%	83.7%
VGG19+SVM	75.1%	87.3%	69.0%	70.3%	84.6%
Shuffle+SVM	73.2%	82.0%	68.2%	68.9%	82.5%
ResNet101+Logistic Regression	70.1%	69.6%	71.7%	71.3%	79.1%
VGG19+Logistic Regression	73.4%	87.8%	67.0%	67.0%	83.2%
Shuffle+Logistic Regression	70.4%	83.1%	64.7%	63.4%	81.0%

Figure 6. Confusion matrix for ResNet101+SVM hybrid model for train/valid and test features dataset

5. DISCUSSION

Table 5 presents a performance evaluation of various deep learning, transfer learning, and machine learning algorithms as described in the scientific literature for diagnosing neuromuscular disorders using muscle ultrasound. The comparison considers parameters such as model type, number of images utilized, and accuracy performance metrics.

Our research tackles the limitations found in earlier attempts, significantly enhancing the DPN diagnostic technique. Our hybrid model offers a reliable and accurate approach to identifying DPN by combining machine learning and transfer learning strategies. Transfer learning enables the model to identify and derive valuable features from complex image datasets, while machine learning techniques contribute to both interpretability and generalization. Transfer learning addresses three main issues with classical machine learning: distribution mismatch, computational power limitations, and scarce labeled data. This technique makes deep learning models more efficient by reducing training time, utilizing existing data more effectively, improving the model's ability to generalize, managing the complexity of deep models, and reducing overfitting. Because of these benefits, transfer learning is a highly effective strategy for utilizing current knowledge and producing better results even with constrained resources. The quantitative evaluation of muscle ultrasound utilizing a transfer learning approach addresses challenges associated with texture features and effectively delineates the underlying muscular tissue. This method enhances the capacity of machine learning algorithms to classify images with increased accuracy. The performance metrics for the testing data set underscore the robust capabilities of the proposed hybrid model, particularly in classifying independent unlabeled images from different subjects that were not included during the training and validation phases. This advancement in the CAD system facilitates the integration of these techniques into ultrasound machines, thereby assisting physicians in making enhanced diagnostic decisions and streamlining their daily workflows.

Table 5. Comparison of cutting-edge studies in muscle ultrasound classification

Study	Algorithm	Model	No. of Subjects / Images	Performance Evaluation	Accuracy
König et al. [22]	Machine learning	2D-DWT+PCA+SVM	18 / 60	Metric	87%
Nodera et al. [23]	Machine learning	Texture features+(simple logistic, SVM and random forest)	51 / 51	Metric	78.4%
Ahmed et al. [24]	Deep learning	YOLO-CSE+SPPF+ELU	80 / 3214	Metric	98%
Uçar [25]	Deep learning	VGG16+VGG19	80 / 3214	Metric	96.1%
Liao et al. [26]	Deep learning & Transfer learning	LeNet, AlexNet, VGG-16, VGG-16TL, VGG-19, and VGG-19TL	85 / 1700	Grad-CAM+Metric	94.2%
Zhou et al. [27]	Deep learning	MMA-Net	NA / 1827	Segmentation+Metric	95.6%
Burlina et al. [28]	Deep learning & Machine learning	DL-DCNNs & ML-RF	80 / 3214	Metric	86.6% & 84.3%
Our approach	Hybrid, Transfer & machine learning	ResNet101, VGG19, Shuffle Net+SVM, logistic regression	53 / 6200	Grad-CAM, t-SNE+Metric	For ResNet101 +SVM, 99.5% training, 75.8 testing

In summary, our hybrid models, ResNet101 with SVM algorithm, have excellent results in terms of accuracy, 99.5% for Training/validation and 75.8% for testing, demonstrating their strong ability to distinguish between positive and negative cases. This suggests a higher likelihood of assigning elevated predicted probabilities to positive instances. A dependability assessment was conducted to evaluate the models' trustworthiness and consistency, examining

performance stability across various iterations and datasets. This analysis provided insights into the models' ability to generalize, highlighting potential constraints or biases. The assessment allowed for informed decisions regarding the models' effectiveness in real-world scenarios.

This study addresses the limitations of quantitative muscle ultrasound assessments, which depend on specific imaging protocols and focus on a single muscle type. Future research should include a wider variety of muscles, particularly distal ones in the upper and lower limbs, since diabetic peripheral neuropathy primarily affects these early on. Additionally, employing advanced ultrasound technologies could lead to a generalized quantification tool for diverse machine types, aiding in the development of a computer-aided diagnosis system for muscle ultrasound.

6. CONCLUSIONS

In this study, a hybrid model has been developed and evaluated, integrating deep transfer learning with machine learning algorithms for diagnosing diabetic peripheral neuropathy utilizing muscle ultrasound imaging. This model employs three convolutional neural networks: VGG19, Shuffle Net, and ResNet101, to extract deep features from ultrasound images. Subsequently, the resulting feature dataset is utilized in conjunction with binary classifiers, specifically SVM and logistic regression, to classify muscle ultrasound images. The quantitative analysis of muscle ultrasound within this diagnostic system demonstrates a valuable and promising approach rooted in our hybrid model. Grad-CAM is considered one of the most effective methods for creating visual heat maps. These maps highlight areas of the input image with greater intensity values, which are associated with critical features and information utilized by the CNN network to make its predictions. It allows the physicians to correctly identify the visualized category, helping understand the model working based on specific regions and measuring whether the proposed classification algorithm can distinguish between classes. According to the evaluation metrics, the ResNet101 hybrid model with the SVM algorithm is the preferred option for the DPN diagnosis system, providing superior accuracy of 99.5%, sensitivity of 99.4%, specificity of 99.5%, F1 score of 99.4%, and AUC score of 99.9% for the training/validation data set. The findings indicate that the development of a computer-aided diagnosis system utilizing muscle ultrasound, in conjunction with a hybrid model, will significantly enhance the functionality of ultrasound machines and improve diagnostic tools, thereby facilitating more informed decision-making in clinical practice.

REFERENCES

[1] Hicks, C.W., Selvin, E. (2019). Epidemiology of peripheral neuropathy and lower extremity disease in diabetes. Current Diabetes Reports, 19: 1-8. https://doi.org/10.1007/s11892-019-1212-8

[2] Galiero, R., Caturano, A., Vetrano, E., Beccia, D., Brin, C., Alfano, M., Salvo, J.D., Epifani, R., Piacevole, A., Tagliaferri, G., Rocco, M., Iadicicco, I., Docimo, G., Rinaldi, L., Sardu, C., Salvatore, T., Marfella, R., Sasso, F.C. (2023). Peripheral neuropathy in diabetes mellitus: Pathogenetic mechanisms and diagnostic options. International Journal of Molecular Sciences, 24(4): 3554. https://doi.org/10.3390/ijms24043554

[3] Fadel, A.W., Nawar, A.E., Elahwal, L.M., Ghali, A.A., Ragab, O.A. (2024). Early detection of peripheral neuropathy in patients with diabetes mellitus type 2. The Egyptian Journal of Neurology, Psychiatry and Neurosurgery, 60(1): 4. https://doi.org/10.1186/s41983-23-00782-9

[4] Yu, Y. (2021). Gold standard for diagnosis of DPN. Frontiers in Endocrinology, 12: 719356. https://doi.org/10.3389/fendo.2021.719356

[5] Pillen, S., Arts, I.M., Zwarts, M.J. (2008). Muscle ultrasound in neuromuscular disorders. Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine, 37(6): 679-693. https://doi.org/10.1002/mus.21015

[6] Hobson-Webb, L.D. (2020). Emerging technologies in neuromuscular ultrasound. Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine, 61(6): 719-725. https://doi.org/10.1002/mus.26819

[7] Pillen, S., van Dijk, J.P., Weijers, G., Raijmann, W., de Korte, C.L., Zwarts, M.J. (2009). Quantitative gray-Scale analysis in skeletal muscle ultrasound: A comparison study of two ultrasound devices. Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine, 39(6): 781-786. https://doi.org/10.1002/mus.21285

[8] Paris, M.T., Mourtzakis, M. (2021). Muscle composition analysis of ultrasound images: A narrative review of texture analysis. Ultrasound in Medicine & Biology, 47(4): 880-895. https://doi.org/10.1016/j.ultrasmedbio.2020.12.012

[9] van Alfen, N., Gijsbertse, K., de Korte, C.L. (2018). How useful is muscle ultrasound in the diagnostic workup of neuromuscular diseases? Current Opinion in Neurology, 31(5): 568-574. https://doi.org/10.1097/WCO.0000000000000589

[10] Stock, M.S., Thompson, B.J. (2021). Echo intensity as an indicator of skeletal muscle quality: Applications, methodology, and future directions. European Journal of Applied Physiology, Springer Science and Business Media Deutschland GmbH, 121: 369-380. https://doi.org/10.1007/s00421-020-04556-6

[11] Chiou, H.J., Yeh, C.K., Hwang, H.E., Liao, Y.Y. (2019). Efficacy of quantitative muscle ultrasound using texture-Feature parametric imaging in detecting Pompe disease in children. Entropy, 21(7): 714. https://doi.org/10.3390/e21070714

[12] Hemalatha, R.J., Vijaybaskar, V., Thamizhvani, T.R. (2019). Automatic localization of anatomical regions in medical ultrasound images of rheumatoid arthritis using deep learning. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 233(6): 657-667. https://doi.org/10.1177/0954411919845747

[13] Dinescu, S.C., Stoica, D., Bita, C.E., Nicoara, A.I., Cirstei, M., Staiculesc, M.A., Vreju, F. (2023). Applications of artificial intelligence in musculoskeletal ultrasound: Narrative review. Frontiers in Medicine, 10: 1286085. https://doi.org/10.3389/fmed.2023.1286085

[14] Brattain, L.J., Telfer, B.A., Dhyani, M., Grajo, J.R., Samir, A.E. (2018). Machine learning for medical ultrasound: Status, methods, and future opportunities.

Abdominal Radiology, 43(4): 786-799. https://doi.org/10.1007/s00261-018-1517-0

[15] Shahid, N., Rappon, T., Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. PloS One, Public Library of Science, 14(2): e0212356. https://doi.org/10.1371/journal.pone.0212356

[16] Ardhianto, P., Tsai, J.Y., Lin, C.Y., Liau, B.Y., Jan, Y.K., Akbari, V.B.H., Lung, C.W. (2021). A review of the challenges in deep learning for skeletal and smooth muscle ultrasound images. Applied Sciences, 11(9): 4021. https://doi.org/10.3390/app11094021

[17] Huang, S.C., Pareek, A., Jensen, M., Lungren, M.P., Yeung, S., Chaudhari, A.S. (2023). Self-Supervised learning for medical image classification: A systematic review and implementation guidelines. NPJ Digital Medicine, Nature Research, 6(1): 74. https://doi.org/10.1038/s41746-023-00811-0

[18] Kim, H.E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M.E., Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. BMC Medical Imaging, BioMed Central Ltd, 22(1): 69. https://doi.org/10.1186/s12880-022-00793-7

[19] Lai, Z., Deng, H. (2018). Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. Computational Intelligence and Neuroscience, 2018(1): 2061516. https://doi.org/10.1155/2018/2061516

[20] Christopher, M., Belghith, A., Bowd, C., Proudfoot, J.A., Goldbaum, M.H., Weinreb, R.N., Girkin, C.A., Liebmann, J.M., Zangwill, L.M. (2018). Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Scientific Reports, 8(1): 16685. https://doi.org/10.1038/s41598-018-35044-9

[21] Zhang, J., Qiu, Y., Peng, L., Zhou, Q., Wang, Z., Qi, M. (2022). A comprehensive review of methods based on deep learning for diabetes-Related foot ulcers. Frontiers in Endocrinology, 13: 945020. https://doi.org/10.3389/fendo.2022.945020

[22] König, T., Steffen, J., Rak, M., Neumann, G., von Rohden, L., Tönnies, K.D. (2015). Ultrasound texture-based CAD system for detecting neuromuscular diseases. International Journal of Computer Assisted Radiology and Surgery, 10: 1493-1503. https://doi.org/10.1007/s11548-014-1133-6

[23] Nodera, H., Sogawa, K., Takamatsu, N., Hashiguchi, S., Saito, M., Mori, A., Osaki, Y., Izumi, Y., Kaji, R. (2019). Texture analysis of sonographic muscle images can distinguish myopathic conditions. The Journal of Medical Investigation, 66(3.4): 237-247. https://doi.org/10.2152/jmi.66.237

[24] Ahmed, A.H., Youssef, S.M., Ghatwary, N., Ahmed, M.A. (2023). Myositis detection from muscle ultrasound images using a proposed yolo-CSE model. IEEE Access, 11: 107533-107547. https://doi.org/10.1109/ACCESS.2023.3320798

[25] Uçar, E. (2022). Classification of myositis from muscle ultrasound images using deep learning. Biomedical Signal Processing and Control, 71: 103277. https://doi.org/10.1016/j.bspc.2021.103277

[26] Liao, A.H., Chen, J.R., Liu, S.H., Lu, C.H., Lin, C.W., Shieh, J.Y., Weng, W.C., Tsui, P.H. (2021). Deep learning of ultrasound imaging for evaluating ambulatory function of individuals with Duchenne muscular dystrophy. Diagnostics, 11(6): 963. https://doi.org/10.3390/diagnostics11060963

[27] Zhou, L., Liu, S., Zheng, W. (2023). Automatic analysis of transverse musculoskeletal ultrasound images based on the multi-task learning model. Entropy, 25(4): 662. https://doi.org/10.3390/e25040662

[28] Burlina, P., Billings, S., Joshi, N., Albayda, J. (2017). Automated diagnosis of myositis from muscle ultrasound: Exploring the use of machine learning and deep learning methods. PloS One, 12(8): e0184059. https://doi.org/10.1371/journal.pone.0184059

[29] Varghese, A., Bianchi, S. (2014). Ultrasound of tibialis anterior muscle and tendon: Anatomy, technique of examination, normal and pathologic appearance. Journal of Ultrasound, 17(2): 113-123. https://doi.org/10.1007/s40477-013-0060-7

[30] Lavric, A., Anchidin, L., Popa, V., Al-Timemy, A.H., Alyasseri, Z., Takahashi, H., Yousefi, S., Hazarbassanov, R.M. (2021). Keratoconus severity detection from elevation, topography and pachymetry raw data using a machine learning approach. IEEE Access, 9: 84344-84355. https://doi.org/10.1109/ACCESS.2021.3086021

[31] Al-Timemy, A.H., Alzubaidi, L., Mosa, Z.M., Abdelmotaal, H., Ghaeb, N.H., Lavric, A., Hazarbassanov, R.M., Takahashi, H., Gu, Y., Yousefi, S. (2023). A deep feature fusion of improved suspected keratoconus detection with deep learning. Diagnostics, 13(10): 1689. https://doi.org/10.3390/diagnostics13101689

[32] Salehi, A.W., Khan, S., Gupta, G., Alabduallah, B.I., Almjally, A., Alsolai, H., Siddiqui, T., Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability, 15(7): 5930. https://doi.org/10.3390/su15075930

[33] Gu, F., Deng, M., Chen, X., An, L., Zhao, Z. (2022). Research on classification method of medical ultrasound image processing based on neural network. Computational Intelligence and Neuroscience, 2022(1): 8912566. https://doi.org/10.1155/2022/8912566

[34] Marzoog, Z.S., Nawir, M.H., Al Zegair, F. (2022). Detecting Covid-19 and other pneumonia diseases using shufflent CNN. Webology, 19(3). http://www.webology.org.

[35] Kandel, I., Castelli, M. (2020). Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review. Applied Sciences, 10(6): 2021. https://doi.org/10.3390/app10062021

[36] Shukla, P., Verma, A., Abhishek, Verma, S., Kumar, M. (2020). Interpreting SVM for medical images using Quadtree. Multimedia Tools and Applications, 79: 29353-29373. https://doi.org/10.1007/s11042-020-09431-2

[37] Awad, F.H., Hamad, M.M., Alzubaidi, L. (2023). Robust classification and detection of big medical data using advanced parallel K-means clustering, YOLOv4, and logistic regression. Life, 13(3): 691. https://doi.org/10.3390/life13030691

[38] Al-Timemy, A.H., Khushaba, R.N., Mosa, Z.M., Escudero, J. (2021). An efficient mixture of deep and machine learning models for covid-19 and tuberculosis detection using x-ray images in resource limited settings. Artificial Intelligence for COVID-19, 77-100. https://doi.org/10.1007/978-3-030-69744-0_6

[39] Hussain, T., Shouno, H. (2023). Explainable deep learning approach for multi-class brain magnetic resonance imaging tumor classification and localization using gradient-weighted class activation mapping. Information, 14(12): 642. https://doi.org/10.3390/info14120642

[40] Hajibabaee, P., Pourkamali-Anaraki, F., Hariri-Ardebili, M.A. (2021). An empirical evaluation of the t-SNE algorithm for data visualization in structural engineering. In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Pasadena, CA, USA, pp. 1674-1680. https://doi.org/10.1109/ICMLA52953.2021.00267

[41] Al-Barazanchi, K.K., Al-Timemy, A.H., Kadhim, Z.M. (2024). Bag of feature-based ensemble subspace KNN classifier in muscle ultrasound diagnosis of diabetic peripheral neuropathy. Mathematical and Computational Applications, 29(5): 95. https://doi.org/10.3390/mca29050095