A complementary Diagnostic Tool for Diabetic Peripheral Neuropathy Through Muscle Ultrasound and Machine Learning Algorithms

Kadhim Kamal¹, Ali Hussein Al-Timemy², Zahid M. Kadhim³, Kosai Raoof⁴

Abstract

Diabetic peripheral neuropathy represents one of the common long-term complications that affect about fifty percentage of diabetes patients. The habitual diagnosis tool based on nerve conduction study that examine the nerve damage and classify the patient status into normal and diabetic peripheral neuropathy with degree of severity without considering the effect on skeletal muscle and take on patient data. A complementary diagnostic tool proposed, in this study integrates the patient’s data including body mass index, age and duration of diabetic, average blood glucose levels, nerve conduction study that involves amplitude and latency of peroneal and tibial nerves and muscle ultrasound alongside the machine learning algorithms to facilitate the clinicians for a precise diagnosis. A group of healthy and diabetic patients utilized to gather the data with calculating the muscle thickness and statistical properties from the gray-level ultrasound images of six skeletal muscles. Support vector machine, naïve bayes, ensemble of bagged tree and artificial neural network supervised machine learning algorithms categorize each class with a high classification accuracy, 98.1% for tibialis anterior with naïve bayes algorithm. The outcomes of this study show a promising complementary diagnostic tool that will help the clinicians to perform an exact diagnosis and disclose the side effect on both nerves and muscles of diabetic patients.

Keywords: Diabetic Peripheral Neuropathy, Muscle Ultrasound, Machine Learning Algorithm.
1. Introduction

Diabetic peripheral neuropathy (DPN) impacts around 50% of the adult patient with diabetes through their life with associated morbidity pain, foot ulcers, and some cases lower limb amputation [1] [2]. Furthermore, DPN is the prevalent and precocious complication which cause cumulative injury of nerve fibers resulting on inferior life quality [3]. The diagnosis of DPN is commonly done by electrophysiological nerve conduction studies (NCS) that can find the changes resulting from nerve damage and assess the clinical features of DPN patients with different symptoms [4]. Since the routine diagnosis tool with NCS considering only the effect of diabetes on nerve fiber, muscle ultrasound (MUS) represents the compulsory quantitative tool that evaluates the muscle condition and gives a scoring diagnosis of the side effect on muscles fiber [5]. MUS is an advantageous and easy to use tool for measuring muscle echo intensity on a quantitative way by grayscale level analysis [6]. Its reliable and affordable for diagnosis alongside the NCS with high image quality [7], [8] [9] [10].

A survey on the previous work carried out to find the best tool that has been used for the diagnosis of DPN and investigates the new techniques that can be combined on an integration system for the precise diagnosis system: Zahid et al in 2021 [11] showed that the muscle ultrasound as a valuable and complementary tool for the diagnosis of DPN, where the thickness and the echogenicity of lower limbs muscle ultrasound in patients with DPN and case-control study had been conducted alongside the NCS, physical examination and history characteristic of those groups. The outcome result was a statistically significant increase in the echogenicity and decrease in muscle thickness of tested patients compared with control group. Alon et al in 2019 [12] explored the assessment of muscle thickness by ultrasound in correlation with electrophysiological study. The measured MUS thickness showed a useful diagnostic tool of various neuromuscular ultrasound and correlated with the electrophysiological study and clinical findings. Andrea et al in 2020 [13] investigated the quality of skeletal muscle by measuring the ultrasound echogenicity alongside the age and body mass index (BMI). The result shows a positive correlation between high muscle echogenicity and age, where there was a high muscle echogenicity related to the overweight and obesity aged participants. Tim et al in 2015 [14] developed and tested computer-aided diagnosis (CAD) for myositis detection in biceps brachii ultrasound images. First-order statistics, Haralick’s and wavelet-based features were extracted from image regions of interest (ROI) then feed to fisher’s and support vector machine (SVM) classifier were the result shows that the SVM was best algorithm for discriminating pathological and healthy muscle tissue, achieving 87% classification accuracy. Xiaofeng et al in 2022 [15] proposed a quantitative feature classification algorithm for breast ultrasound images using improved Naïve Bayes (NB). A group of histogram, texture and grayscale co-generation features extracted from the ultrasound image and used for improved NB algorithm to achieve high classification accuracy.

From the previous studies, no one of the researchers used all the patient data, NCS and MUS as an input data for a complementary diagnosis system alongside the machine learning (ML) algorithms which will be introduce in this study by following sections.

In this paper, a complementary diagnosis system will be proposed by using muscle ultrasound beside the electrophysiology NCS with an integration of patients’ data and the hemoglobin A1c (HbA1c) test for making a precise and accurate decision that help the clinicians in making better diagnosis. While considering all side effects of diabetic on patient health condition form the body mass index, cumulative blood sugar, nerve fiber damage and the fibrosis on muscle tissue as a whole picture that play an important role in the treatment and monitoring of peripheral neuropathy.

2. Materials and Methods

2.1 Data Collection

This study is based on a collaboration with the neurophysiology unit, Al-Shaheed Ghazi Al-Hariri hospital in medical city, Baghdad where the data collected from twenty-six patients (20 males and 6 females) with type 2 diabetes mellitus (DM) as the DPN group and twenty-seven healthy subjects (20 males and 7 females) as control group (CTR) were included. The ethics committee of hospital and university approved this study, also the database registered in medical city archiving department. Furthermore, this study done with considering the ethical standards of the Declaration of Helsinki for human experimentation. The age of DPN and CTR group ranged from 29 to 67 years and 25 to 61 years respectively. A full set of data collected from each participate including:

- Age, sex, body mass index (BMI) calculated, biochemical investigations Include HbA1c were obtained,
- Electrodiagnostic study include the amplitude and latency of peroneal and tibial nerves and
- Ultrasound examination of a group of muscle was recorded.

The ultrasound examination performed for six muscles, three on the upper limb including biceps brachii (BB), brachioradialis (BR), abductor digiti mini (ADM) and another three on the lower limb including rectus femoris (RF), tibialis anterior (TA) and abductor hallucis (AHL). Those muscles have been selected based on their position from proximal to distal for each limb “upper and lower”, where the study based on the investigation of the effect on each limb muscles. Moreover, DPN is length dependent disease and effect distal muscle more than proximal. Furth more, those muscle usually examine by physician as a reference for electromyography test for neuromuscular disorders. Figure 1 shows the workflow of this study based on the three types of data that used to feed into the ML algorithms in order to find the best diagnosis result.

Phillips ultrasound machine, iU22 used for muscle ultrasound examination with linear probe adjusted with frequency 5-12 MHz and special preset for musculoskeletal examination used for examine all the
muscles. Data collection procedure involved first physical examination of all participants by an experienced physician to confirm their suitability for the study, where the exclusion criteria of all groups consider all kind of other neuromuscular disorders and other diseases like kidney failure and cancer, and make the groups contain only healthy and DPN patients. Then, collecting the age, duration of DM with height and weight variables were measured for all patients, and the BMI was also calculated. The second step is NCS examination of all participant by an experienced neurophysiologist, measuring the amplitude and latency of peroneal and tibial nerves, where the patient report gathers to the third and final step of examination which is the muscle ultrasound.

The muscle ultrasound examination by the same experienced neurophysiologist based on the anatomical structure of tested muscle with carrying on the previous research for finding the best examination position for ADM [16], BB [17], BR [18], AHB [19], TA [20], and RF [21] muscles, where the ultrasound images take on the muscle belly, the ultrasound probe position perpendicular on the muscle takes a screen after observing the bone behind the belly and saved on the device. The muscle thickness was calculated on the ultrasound machine from the saved images for all the participants and documented beside the patient data and NCS on one sheet for each muscle.

2.2 Image segmentation and feature extraction

The MUS images exported from the machine by hard disk in digital imaging and communications in medicine (DICOM) format, then, on a laptop the format change to jpg images by using MicroDicom software. Image segmenter application from MATLAB 2023a was used to select the region of interest (ROI) for each muscle based on the anatomical structure and the reference mention above on the data collection section with the supervision of the authors who is clinical neurophysiology. Seven features extracted from each ROI including standard deviation (STD), variance (VAR), entropy which is a statistical measure of randomness that can be used to characterize the texture of the input image [22], and four grey-level co-occurrence matrix (GLCM), contrast, correlation, energy and homogeneity which is a commonly applied features that extracted statistical measurements about the coexistence of different from grey-scale images [23], [24], this procedure illustrate in Figure 2. The seven features were combined together with the muscle thickness, four readings of NCS, age, BMI, DM duration and HbA1c blood glucose level in one sheet for each muscle including the two groups. Those patient data and muscle ultrasound features collected on one sheet per each muscle where the rows represent the subjects and the columns represent the patient’s data and features extracted from each muscle totally sixteen columns for 53 subjects, and six sheet for all muscles.

2.3 Classification

The classification was performed with supervised machine learning algorithms by using MATLAB 2023a software, student License. Four supervised algorithms including support vector machine (SVM), naïve bayes (NB), ensemble of bagged tree and artificial neural network (ANN) had been used to classify the data from each muscle into CTR & DPN class. The selection of these algorithms based on general survey to found the highest accuracy, where these four achieved best performances to be selected in this work.

The SVM algorithm idea based on statistical learning theory that tries to search for the hyper-plane that maximizes the margin between two classes where its widely used on medical image classification [25]. The NB algorithm based on the assumption of the inputs are conditionally independent in each class, which provide fast results better than the sophisticated methods. NB algorithm not considers correlation between features and lowering the variance meanwhile increasing the bias to avoiding over fitting on the training set [26]. Ensemble of bagged tree algorithm based on the idea of ensemble of classifiers built on bootstrap replicates that generated randomly from the training set, which is divided into subsets via bootstrap resampling used to construct training data of each decision trees [27]. ANN is considered as common machine learning algorithm with wide applications of health care like clinical diagnosis, it’s like statistical techniques including generalized linear models. ANN has multiple layers, consist of processing units (nodes or neurons) that are interconnected by a set of adjustable weights which let signals to move on the network in parallel and consecutive way. In generally ANN split into three layers of neurons: input (receives information), hidden (responsible for extracting patterns, perform most of internal processing), and output (produces and presents final network outputs) [28], [29].

The data divided into five parts were the five-fold cross validation has been used to test each algorithm performance and calculate the classification accuracy, 80% of data used for the training the classifier and the rest 20% used for testing the classification accuracy for each trail. The test portion of data changes for next training and testing trail and this process repeated five times where the accuracy saved per each trail. The overall classification accuracy is the average of five trails to investigate the classifier performance.
3. Results and Discussion

The data sheet of each muscle which includes sixteen features from patient data, NCS and MUS are feed to four machine learning algorithms: SVM, NB, ensemble of bagged tree and ANN by using the classification leaner application on MATLAB 2023a, where the classification accuracy for each algorithm among the muscle groups has been calculated as in Table 1. It's observed that there is a variation on the classification accuracy in each muscle group between the algorithms due to the anatomical variation, size and fiber orientation of each muscle. Also, the physiological behavior like the load capacity between the upper and lower limbs muscle effect on the extracted feature from ultrasound images of each muscle. Based on the above result, best algorithm selected for each muscle to investigate the classification performance which includes the accuracy (ACC), Area under the receiver operating characteristic curve (AUC) and provides an aggregate measure of performance across all possible classification thresholds, sensitivity, specificity and F-Score that is the harmonic mean of a system's precision and recall values as in Table 2.

![Figure (1): The workflow of this study](image1)

![Figure (2): Procedure of MUS images saving, converting, selecting ROI and feature extracting.](image2)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>ADM</th>
<th>AHB</th>
<th>BB</th>
<th>BR</th>
<th>RF</th>
<th>TA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machine</td>
<td>90.6%</td>
<td>88.7%</td>
<td>90.6%</td>
<td>92.5%</td>
<td>96.2%</td>
<td>86.8%</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>92.5%</td>
<td>94.3%</td>
<td>92.5%</td>
<td>92.5%</td>
<td>94.3%</td>
<td>98.1%</td>
</tr>
<tr>
<td>Ensemble bagged tree</td>
<td>92.5%</td>
<td>92.5%</td>
<td>94.3%</td>
<td>94.3%</td>
<td>94.3%</td>
<td>94.3%</td>
</tr>
</tbody>
</table>
Table (2): Performance evaluation of the best algorithm for each muscle group.

<table>
<thead>
<tr>
<th>Muscle</th>
<th>Classifier</th>
<th>ACC</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADM</td>
<td>ANN</td>
<td>94.3%</td>
<td>98.01%</td>
<td>96.15%</td>
<td>92.59%</td>
<td>94.34%</td>
</tr>
<tr>
<td>BB</td>
<td>Bagged</td>
<td>94.3%</td>
<td>96.3%</td>
<td>96.15%</td>
<td>92.59%</td>
<td>94.34%</td>
</tr>
<tr>
<td>BR</td>
<td>Bagged</td>
<td>94.3%</td>
<td>96.65%</td>
<td>96.15%</td>
<td>92.95%</td>
<td>94.34%</td>
</tr>
<tr>
<td>AHB</td>
<td>NB</td>
<td>94.3%</td>
<td>99.43%</td>
<td>90%</td>
<td>100%</td>
<td>94.74%</td>
</tr>
<tr>
<td>TA</td>
<td>NB</td>
<td>98.1%</td>
<td>99.72%</td>
<td>96.43%</td>
<td>100%</td>
<td>98.18%</td>
</tr>
<tr>
<td>RF</td>
<td>SVM</td>
<td>96.2%</td>
<td>97.86%</td>
<td>100%</td>
<td>92.86%</td>
<td>96.15%</td>
</tr>
</tbody>
</table>

From above table, it is observed that BB and BR muscle tested high classification accuracy by the same algorithm which is the ensemble of bagged tree due to the functional and physiological similarity of those muscles in the arm-forearm region of the upper limb while both of them work together for doing a movement. The same thing with AHB and TA muscles in lower limb that sharing the high accuracy with naïve bayes algorithm.

Figure (3) shows the receiver operating characteristic (ROC) that is a measurement of the ability of an algorithm to discriminate whether a specific condition is present or not present of the best algorithm for each muscle.

Figure (4) shows the confusion matrix for each machine learning algorithms per muscle. It demonstrates the classification accuracy between the true and predict class and gives us an idea about which class the algorithm has better accuracy comparing with the others. It's obvious that the CTR muscle group has the highest classification accuracy in general comparing with the DPN muscle group among all machine learning algorithms, this is due to the many changes in DPN muscle that comes from the fibrosis that result from the complication of DM on patients as a long-term disease.

Figure (5) shows the t-distributed stochastic neighbor embedding (t-SNE) which is a single map used for visualizing high-dimensional data. It's embedded by mapping the data points to a two or finite-dimensional space. The t-SNE reveals the intrinsic structures in a high-dimensional dataset, including trends, patterns, and outliers, through a nonlinear dimension reduction technique [30].

It visualizes the boundary between features of the two classes and give an idea where the features overlapped for each group. Clearly it can be seemed and recognized that the features of the two groups, CTR & DPN are well separated and this is supporting the highest classification accuracy that obtained among different algorithms. The boundary line is good observed and differentiated the groups where very few data point overlapped with others group. This figure help researchers to understand how the features are distributed into two-dimensional scale and how to select the best classification algorithm based on this distribution.

![Figure (3): The ROC curve of best algorithm for each muscle.](image-url)
According to the data on Table 3. It can be seen that the complementary data for TA muscle with NB machine learning algorithm, the ACC accuracy archive 98.1 % which is higher than Ultrasound texture-based CAD [14] 87 % and the quantitative feature classification [15] 89.7 % by 11.1 % and 8.4 % respectively.

**Table (3):** Comparison of ACC value results of different algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>98.1 %</td>
</tr>
<tr>
<td>CAD [14]</td>
<td>87.1 %</td>
</tr>
<tr>
<td>Quantitative [15]</td>
<td>89.7 %</td>
</tr>
</tbody>
</table>

4. Conclusions:

The integration between the patient data, electrophysiological study and MUS features in the computer aided diagnosis shows a promising result in this study, by classifying the patient with DPN through MI. algorithms. The outcome of this study will make the clinical diagnosis more precise, accurate and considering all the information to make a complementary diagnosis system. Meanwhile improving the workflow of the neurophysiologist and providing such a complete diagnosis system that will help them for getting a better diagnosis. This work provides a promising technique among others studies, with a contribution of the integration system for a significant outcome that achieved best result of complementary diagnostic system.

6. References:


